ANALYSIS OF NON-CANONICAL INTRONS IN THE HUMAN TRANSCRIPTOME

Parada G., Munita R., Gysling K.
Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile.

Pre-mRNAs of higher eukaryotes are often interrupted by non-coding sequences, called introns, that are accurately excised through two sequential trans-esterification reactions catalyzed by five small ribonucleoproteins (spliceosome). This process is known as splicing.\(^1\)

The introns have common sequences at their boundaries that allow them to be recognized by the splicing machinery, the most conserved are the initial and terminal dinucleotides. There are two types of introns, the largest part of those are U2-type and have 5'GT and 3'AG dinucleotide ends, while the U12-type may also have AT-AC dinucleotides.\(^2\) The dinucleotides ends are crucial for the splicing process; several mutations at these nucleotides have been shown in mis-splicing diseases.\(^3-7\)

There are introns that are recognized even though they have non-canonical dinucleotides. Early analysis of non-canonical splice junctions found U2-type introns with non-canonical 5'GC ends.\(^8\) These introns have strong compensatory consensus signals at the splicing acceptor sites and they have been implicated in alternative splicing.\(^9-10\) More recent, genome-wide analysis have shown the existence of another U2-type variation, the GT-TG introns, that have also been involved in alternative splicing.\(^11\) Other particular cases of non-canonical splice junctions have been reported, suggesting that we still do not know the entire diversity of non-canonical splice junctions.

In the present study, we have done a compressive analysis of non-canonical splice junctions in the human transcriptome. For this purpose, about 1.2 billion of RNA-seq reads from Illumina bodymap 2.0 (ERP000546) were mapped with two different softwares (MapSplice and GSNAp\(^19\)). Introns longer than 40 bp were extracted and post filtered in order to get an initial confident list of non-canonical splice junctions.

Putative SNPs at the dinucleotides can lead to the misidentification of non-canonical splice junctions. To avoid this source of error, we also mapped RNA-seq data from the GM17828 cell line to their own genome and applied the same methodology. The non-canonical splice junctions detected by this way confirmed a fraction of the cases previously found.

Several non-canonical splice junctions found are also supported by mapped mRNA and EST from UCSC genome database\(^20\) and GENECODE annotation models.\(^21\) We successfully validated some of these non-canonical splice junctions by RT-PCR.

Our future work will focus in alternative splicing and conservational analysis of these non-canonical splice junctions. We will also use these introns as a model to study the splicing itself, because non-canonical introns might have compensatory sequences or epigenetic modifications that allow their recognition and processing by the spliceosome.

**Supported by grants Fondecyt N° 1110392 and MSI N° P10-063-F**