Annotation and analysis of the dynamics of splice acceptor sites in *Trypanosoma cruzi* under gamma radiation stress

Reis, A. L. M.[1]; Bitar, M.[2]; Vieira, H. G. S.[3]; Kaczorowski, D.[3], Macedo, A. M.[1]; Machado, C. R.[1]; Franco, G. R.[1];

1.Departamento de Bioquímica e Imunologia, UFMG, BH/MG, Brazil; 2.QIMR Berghofer, Brisbane, Australia; 3.Garvan Institute of Medical Research, Sydney, Australia

Unlike most eukaryotes, in Trypanosomatids, genes are not interrupted by introns and transcription is polycistronic. Maturation of individual mRNAs is accomplished by coupling spliced leader trans-splicing (SLTS) and polyadenylation. The attachment of a 39-nucleotide sequence, the spliced leader, to the 5′ end of individual cistrons upon recognition of splice acceptor sites resolves the 5′UTR. Alternative SLTS has not been characterized in *Trypanosoma cruzi*, although different regulatory roles have been speculated for this mechanism. A deeper investigation is necessary to clarify how specific splice acceptor sites are selected under diverse environmental conditions. This parasite is highly resistant to different sources of stress, sustaining an exposure to 500 Gy of gamma radiation. Under this circumstance, genomic DNA is fragmented, but the karyotype is gradually restored, leading to a complete re-establishment of the chromosomal band pattern in less than 48 hours. The aim of this study is to characterize trans-splicing at structural and functional levels, observing its dynamics after *T. cruzi* is exposed to ionizing radiation stress. A time-series experiment was designed to evaluate changes in the transcriptome of epimastigotes of the CL Brener strain not exposed versus 4, 24 and 96 hours after exposure to 500 Gy of gamma radiation with two biological replicates at each time point. Following total RNA extraction, libraries were prepared with Truseq mRNA Stranded and then sequenced in the Illumina Hiseq2500 platform. The pipeline applied for the annotation of splice acceptor sites was mainly comprised of: FastQC (for quality check), Cutadapt (for identifying and trimming the spliced leader sequence), BWA-mem (for mapping to the reference genomes), Python in-house scripts (for the actual calling of splice sites) and R scripts (for statistical analyses). A total of 48,719 different splice sites were identified using this protocol. They were assigned to 17,053 annotated genes, while there were 5,658 genomic regions with splice sites but no genes annotated. Further inspection showed that 1,384 genes out of those annotated may have a wrong CDS start and 1,228 new CDSs were found in those regions with no annotation. Most of the sites are present in the control (~71%) and there is a clear distinction, in terms of coverage, between the conditions non-irradiated versus irradiated. Some genes even have a change in their main splice site over time. In summary, this work enables a refinement of genome annotation and a better understanding of the dynamics of SLTS in *T. cruzi*.