The salmonid selenotranscriptome: in silico and in vivo characterization.

Francisco Altimiras, Rodrigo Pulgar y Verónica Cambiazo. fjaltimiras@inta.uchile.cl
Laboratorio de Bioinformática y Expresión Génica, INTA-Universidad de Chile and Center for Genome Regulation (CRG).

Introduction: The selenotranscriptome is the set of selenoprotein transcripts (selenotranscripts). Selenoproteins are a diverse group of proteins implicated in immune and oxidative responses that are present in all three domains of life: bacteria, archaea and eukaryota. Selenoproteins contain selenium (Se) in the form of the amino acid selenocysteine (Sec) which is encoded by the STOP codon (UGA). The presence of an RNA stem-loop structure, the Sec insertion sequence (SECIS) element in the 3' untranslated region (3'UTR) of eukaryotic mRNAs, differentiates the Sec or stop function of UGA codons. Given that the dual meaning of the UGA codon, selenoprotein genes and transcripts are often miss-predicted by standard annotation pipelines. Here we applied a series of bioinformatics tools to predict the salmonid selenotranscriptome from large collections of EST sequences.

Materials and Methods: We applied a computational approach based on the SECISsearch tool and sequence alignment algorithms to predict transcripts with UGA in-frame along with a putative 3'-UTR SECIS downstream element. In doing so, we used a Atlantic salmon EST database of 59,336 contigs. We validated the presence of predicted selenotranscripts in several tissues of Atlantic salmon by RT-PCR followed by amplicon sequencing. In addition, we quantified the abundance of 4 sequenced selenotranscripts by quantitative real time PCR. We used RACE-PCR to obtain 3'-UTR sequences from two putative selenotranscripts. Finally, in current experiments we are evaluating the selenium-dependent cytosolic glutathione peroxidase (cGPx) enzymatic activity in an in vitro model of salmonid phagocytes (shk-1 cell line).

Results: We predicted a set of 31 transcripts with UGA in-frame codons, in 26 of them we also detected 3'-UTR SECIS elements. By using RACE-PCR, we were able to predict one additional SECIS element. Transcripts were grouped according to their UGA pattern (N-term or C-term UGA in-frame codon) and all candidate salmonid selenotranscripts were named after their selenoprotein homologous. We identified one UGA codon in-frame per selenotranscript, exception were the two representatives of selP family (selPa1, selPa2) with a higher Sec number (16-17) and selL with 2 Sec residues that probably form a diselenide bond (UxU). Regarding expression analysis, we found a specific-tissue regulation of selenotranscript expression, which is in agreement with previously reported data.

Discussion: The computational approach for selenotranscripts prediction was successful to identify the salmonid selenotranscriptome. This set of transcripts are larger than other fish selenotranscriptomes previously reported, however, the same core selenoprotein families were found, including several selenoproteins, such as: fep15, selJ, selU and sell, which are missing in mammals.

Supported by: Fondecyt 1090211, Fondap 15090007.