Characterization of the Transcriptional Landscape of Circulating Extracellular Vesicles in HER-2+ Breast Cancer Patients

Maria Galli de Amorim (AC Camargo Cancer Center, Brazil), Renan Valieris (AC Camargo Cancer Center, Brazil), Rodrigo Drummond (AC Camargo Cancer Center, Brazil), José Humberto Fregnani (Hospital de Câncer de Barretos, Brazil), Israel Tojal da Silva (AC Camargo Cancer Center, Brazil), Emmanuel Dias-Neto (AC Camargo Cancer Center, Brazil), Diana Noronha Nunes (AC Camargo Cancer Center, Brazil).

BACKGROUND: Circulating extracellular vesicles (EVs) derived from cancer patients represent a potential source of tumor biomarkers. In this study we characterized the transcriptional landscape of EVs derived from the plasma of individuals without breast cancer (n=5) and patients with $HER2^+$ breast cancer (n=30).

HYPOTHESIS: Differentially expressed transcripts found between controls and patients from different stages are potential diagnostic and prognostic biomarkers. The comprehensive analysis of EVs RNA cargo may also shed light on the biological function of these vesicles in cancer.

METHODS: In order to follow the natural progression of the disease and to identify transcripts potentially involved with tumor agressiveness, we selected three groups of patients with invasive ductal $HER2^+$ breast cancer with increasing tumor staging: 10 patients with disease restricted to the primary site, 10 patients with lymph node disease and 10 patients with metastatic disease. Blood-derived EVs were isolated by differential ultracentrifugation, RNA was extracted from EV pellets and libraries were constructed with Ion Total RNA-Seq kit v2 for sequencing on the Proton platform. Given the limited quantities of RNA obtained from EVs, an adaptation of the protocol was performed where the RNA was extensively fragmented to augment the number of molecules available for adapter ligation and thereby reduce the formation of adapter dimer artifacts, as well as to allow a single library construction to characterize the entire transcritome (both small and long RNAs) from each sample. The bioinformatics pipeline used to determine the transcriptome consisted of an initial filter using FastQ Screen, to eliminate rRNA, repetitive elements and contaminants from other species. miRDeep2 software was used to annotate miRNAs contained in miRBase v20 and unmapped reads were aligned against hg19 genome reference using STAR aligner. tRNAs were annotated using GtRNAdb database, followed by Ensembl Homo_sapiens.GRCh37.73 for all other transcript annotations. DESeq2 was used for differential expression analysis.

RESULTS: Our modified library construction protocol was successful in generating sequencing data from very small quantities (<1ng) of input RNA. Given that the whole transcriptome was sequenced, we believe that our results represent an unbiased view of the EVs RNA content. We have generated an average of 6.7 million reads per sample. An average of 6.390 protein coding transcripts, 176 miRNAs, 18 snoRNAs, 20 snRNAs, 93 lincRNAs, 151 Y-RNAs and 193 tRNAs were identified per sample. Differential expression analysis identified a total of 529 unique transcripts that were significantly differentially expressed ($p<0.01$) between controls and patients. The majority of differentially expressed transcripts were protein-coding, with RNAs classified as miscellaneous (miscRNAs) as the second most represented category. Y-RNAs were the most abundant among miscRNAs. In addition, several tRNAs fragments were found differentially expressed. Although a full understanding of the biological role of these molecules is lacking, the enrichment of Y-RNAs and tRNAs fragments in EVs may be important in the regulation of gene expression in recipient cells that uptake EVs.
Financial support: FAPESP 2011/09172-3.