RNA-Seq Reveals Transcriptional Differences Between Early Stage Metastatic and Advanced Stage Non-Metastatic Oral Squamous Cell Carcinomas

Frederico O Gleber-Netto (AC Camargo Cancer Center, Brazil), Willem Hans Schreuder (AC Camargo Cancer Center, Brazil), Rodrigo Duarte Drummond (AC Camargo Cancer Center, Brazil), Renan Valieris (AC Camargo Cancer Center, Brazil), Israel Tojal (AC Camargo Cancer Center, Brazil), Maria Galli de Amorim (AC Camargo Cancer Center, Brazil), Ana Flavia Costa (AC Camargo Cancer Center, Brazil), Melissa Pool Pizzi (AC Camargo Cancer Center, Brazil), GENCAPO, Rossana VM López (Instituto do Câncer do Estado de São Paulo, CTO/ICESP, Brazil), Fábio Daumas Nunes (Faculdade de Odontologia da USP, Brazil), Marcos Brasilino de Carvalho (Hospital Heliópolis, Brazil), Luiz Paulo Kowalski (AC Camargo Cancer Center, Brazil), Diana Noronha Nunes (AC Camargo Cancer Center, Brazil), Emmanuel Dias-Neto (AC Camargo Cancer Center, Brazil).

BACKGROUND: Detection of local lymph node metastasis in oral cavity squamous cell carcinoma (OSCC) is pivotal for prognosis. With current preoperative assessment of lymph node metastasis approximately 20-40% of occult nodal metastasis will remain undetected, potentially leading to inappropriate treatment. Gene expression profiling has been reported to have predictive value for assessment of lymph node status in OSCC, however, the molecular signatures from different studies show little overlap and no single set of genes with optimal predictive value. Also, all published gene-expression studies were performed using cDNA microarrays and were therefore limited known transcripts. Our aim was to determine gene expression signatures, derived from massive parallel sequencing of RNA (RNAseq), capable of detecting tumors with lymph node metastatic potential.

HYPOTHESIS: The transcriptome analysis of polar tumor groups composed of large non-metastatic tumors from one end of the spectrum versus small but already metastatic tumors from the other end of the spectrum, using deep RNA-Seq may detect gene expression signatures predictive of lymph node metatstases in OSCC.

METHODS: Frozen tumor samples (tumor percentage > 80%) of 10 patients with early-stage metastatic OSCC (pT1-2N+; Group I) and 10 patients with advanced stage non-metastatic OSCC (pT3-4N0; Group II) were identified and included. Localizations and histopathological characteristics were evenly distributed between the groups. All patients had a follow up of at least 3 years and received neoadjuvant treatment. RNA samples (RIN≥6) were depleted from genomic DNA contamination and rRNA, cDNA libraries with a mean size of 260 nt were constructed, submitted to emulsion PCR and sequenced using the platform SOLiD5500XL. After aligning and processing the sequence data (Lifescope and HTseq softwares), genes with a total reads < 5, a mean reads per sample < 0.2 and cpm <1 were excluded. For determining genes differently expressed between two groups DESeq2 software was used. Genes with a p-value <0.01 and false discovery rate < 0.1 were considered to be differently expressed. To validate the predictive value, the genetic signature was compared to data sets of non-metastatic and metastatic OSCC generated by The Cancer Genome Atlas (TCGA).

RESULTS: A total of 3×10^9 gene fragments were generated, representing a good transcriptome coverage with a mean of 153×10^6 fragments per sample (range 99×10^6 – 296×10^6). Most sequences were generated from group I (61%), being statistically significantly (p=0.01) different from group II (39%). A total of 18,815 unique transcripts were sequenced, mostly derived from protein-coding genes (83.5%). A total of 150 genes suggested a different expression between both groups and this set of genes was able to differentiate both groups according to the lymph node status of the patient. In silico validation with TCGA-data produced a gene-signature of 7 genes. The predictive value of this gene-subset remained when again applied to the sequence data of the original samples.

Financial support: CNPq, FAPESP