**INTEGRATIVE METHODOLOGY TO UNCOVER DRIVERS CANDIDATES ASSESSING GENOMIC, TRANSCRIPTOMIC, MIRNA AND METHYLATION DATA IN PENILE CARCINOMA.**

Fabio Albuquerque Marchi (A.C.Camargo Cancer Center, Brazil), David Correa Martins Jr (Federal University of ABC - UFABC, Brazil), Mateus Camargo Barros-Filho (A.C.Camargo Cancer Center, Brazil), Hellen Kuasne (A.C.Camargo Cancer Center, Brazil), Ariane Fidelis Busso (Molecular Epidemiology Laboratory – UNICAMP, Brazil), Jose Carlos de Souza Trindade Filho (Faculty of Medicine – UNESP, Brazil), Cristovam Scapulatempo Neto (Cancer Hospital of Barretos, Brazil), Ademar Lopes (A.C.Camargo Cancer Center, Brazil), Helena Brentani (Institute of Psychiatry, FMUSP), Silvia Regina Rogatto (A.C.Camargo Cancer Center, Brazil; Faculty of Medicine - UNESP, Brazil; Vejle Hospital - University of Southern Denmark, DK).

**BACKGROUND:** Integrative analysis of multiple cancer -omics data is challenging due to the heterogeneity among studies, high dimensionality and small number of samples. In recent years, algorithms have been published aiming to differentiate driver alterations from passengers. However, in rare diseases there is a lack of candidates identified by high-throughput analysis. Penile carcinoma (PeCa) is a rare neoplasm in developed countries (0.5-1.1:100,000 in Europe) and an important public health problem in poor and developing countries, such as Brazil (2.9-6.8:100,000). To our knowledge there is no study involving a multidimensional integrative approach to identify driver genes in PeCa samples.

**HYPOTHESIS:** Multidimensional data analysis can contribute to the identification of driver genes and deregulated pathways in patients with penile carcinoma.

**PATIENTS AND METHODS:** Thirty-one fresh frozen PeCa and 19 normal glans were assessed by genome-wide aCGH, transcriptome, methylome, and miRNA analysis. Data were processed and grouped in a *m x n* matrix, where *m* represents the number of genes and *n* the number of samples, for each molecular level. The algorithm was designed in four steps: 1) Identification of target genes, data stratification according to the alteration and the selection of candidates by correlation, considering the best combination in all analyzed levels; 2) Transcriptome matrix partitioned into different modules of co-expressed genes that exhibit the same behavior and the same conditional probability distribution (CPD) using a biclustering algorithm; 3) Modules enriched to biological pathways (Gene Ontology database (geneontology.org) and filtered according to the number of passenger genes (genomic alteration with frequency < 20% and no included in the first step); 4) Each target candidate, previously selected, was assigned as a regulator of a module based in the regression tree algorithm. The best target-module association was represented by score. The top 10 driver candidates were selected for validation by RT-qPCR in an independent set of 29 samples.

**RESULTS**: *BIRC5, DTX2, PPARG, PLCB3, NRAS, RB1* and *TNFSF10* genes were chosen as driver candidates and confirmed by RT-qPCR. Important biological pathways related to tumor process were identified, such as regulation of cell growth (GO:0001558, p=6,2x10-3), homeostasis (GO:0048872, p=8,2x10-3), and transcription regulation (GO:0003700, p=8,9x10-4). The algorithm was validated in an independent analysis using 255 Glioblastoma Multiforme (GBM) samples of four-level alterations available on TCGA (https://tcga-data.nci.nih.gov/). Among the top 15 genes identified, *PTEN, RB1, TP53, EGFR* and *CDKN2A* were associated with tumor development.

**CONCLUSION:** Our results suggest that the integrative algorithm proposed in this study was able to identify driver candidates, contributing to an understanding of penile carcinoma development.

**FINANCIAL SUPPORT**: National Institute of Science and Technology in Oncogenomics (FAPESP 2008/57887–9 and CNPq 573589/08–9), CAPES (Coordination for the Improvement of Higher Education Personnel) and FAPESP (2009/52088-3) and CNPq (302606/2011-4).