

# Analysis of potential disease-causing variants in a patient with intellectual disability via whole-exome sequencing

Patricia de Cássia Ruy, Isabela Ichihara Barros, Reginaldo Cruz Alves Rosa, Jessica Rodrigues Praça, Amanda Cristina Corveloni, Cibele Cardoso, Aline Fernanda de Souza, Carlos Alberto Oliveira de Biagi Junior, Ádamo Davi Diógenes Siena, Kamila Peronni Zueli, Maria Florencia Tellechea, Simone da Costa e Silva Carvalho, Greice Andreotti de Molfetta, João Pina, Wilson Araújo da Silva Jr

*Faculty of Animal Science and Food Engineering of USP, USP, Brazil*

## Abstract

To realize the intellectual disability of a patient with none specific syndrome identified, the whole-exome sequencing (WES) technology was applied to search for potential disease-causing variants. Previous studies in our laboratory found a long noncoding RNA deleted in this patient and in vitro neural differentiation showed that patient cells are delayed in the differentiation process when compared to control cells. Besides the intellectual disability, this patient presents motor incoordination, flat feet, brachydactyly, obesity, macrocephaly, hypogenitalism and small mouth and teeth. To better understand this patient phenotype the WES of the patient and his family (father, mother and grandmother) were performed. DNA sequencing (Illumina TrueSeq Rapid Exome) of four blood samples was performed using an Illumina NextSeq 500. The quality of generated reads was obtained with FastQC software and the quality trimming was made by TrimGalore, considering a Phred Score of 30. The alignment of reads with the human reference genome (hg19) used BWA (Burrows-Wheeler Aligner) and the processing of alignment files was made by Picard tools. The variants (germinative mutations) were identified using Strelka. The possible pathogenetic significance of identified variants was assessed using Exomiser program with HPOs (Human Phenotype Ontologies) related to intellectual disability, autistic behavior, etc. This program integrates CADD (Combined Annotation Dependent Depletion), PolyPhen (Polymorphism Phenotyping), SIFT (Sorting Intolerant From Tolerant) and Mutation Taster to assist the variant impact. A filter of variants with maximum allele frequency of 2% was applied. An average of 122 million reads was generated per sample. The mean coverage was between 43x and 57x after removing duplicates. The analysis of the patient exome identified a total of 113, 200 variants. The top ten mutated genes are: CHR\_START-U2 (440), Y\_RNA (423), PRIM2 (274), CHR\_START-AL592188.3 (256), bP-21201H5.1-IGHV1OR21-1 (252), RP11-96F8.1-KSR1P1 (233), snoU13 (213), BAGE2 (164), BX088702.2-CHR\_END (147), CHR\_START-Y\_RNA (134). After annotation, filtering and prioritizing likely causative variants with Exomiser, the top 5 genes highlighted are: ATRX (2 missense X-recessive and 1 missense X-dominant variants), PTEN (1 splicing autosomal dominant and 1 splicing and 1 missense autosomal recessive variants), CHD7 (1 missense autosomal dominant and 2 missense autosomal recessive variants), AKT1 (1 missense autosomal dominant and 1 UTR5 autosomal recessive variants) and HDAC8 (1 missense variant X-dominant). A multi-sample Exomiser analysis will be made with the family samples to analyze the possible heredity of the identified variants.

Funding: CAPES, FAPESP, FAEPA, CNPq, FUDHERP