Whole-exome sequencing and bioinformatics analysis

Abdelaziz Tlili,Abdullah Fahd Al Mutery,Mona Mahfood,Walaa Kamal Eddine Ahmad Mohamed and Khalid Bajou

Abstract

Sequencing library construction, exome capture, sequencing, and standard data analyses for the affected children in this family was performed by Sengenics. Exome capturing and enrichment was carried out using SureSelect All Exon V5 kit (Agilent Technologies, Santa Clara, CA, USA) following the manufacturers' protocols. Whole exome sequencing was carried out on Illumina HiSeq 2500 system (Illumina, San Diego, CA, USA). Paired end (2×100 bases) DNA sequence reads that passed the quality control i.e phred score > 20 were mapped to the human reference genome build hg19/GRCh37 using the BWA (Li, Durbin 2010) and SAM tools (Li, Handsaker et al. 2009) was used for processing BAM files. Genome analysis tool kit (GATK) v2.7.2) (McKenna, Hanna et al. 2010) was used for calling variants from BAM files. Variants were annotated with gene, existing variations, consequences from dbSNP (build 137), SIFT v5.0.2 (Kumar, Henikoff et al. 2009) and polyphen v2.2.2 (Adzhubei, Schmidt et al. 2010) using Ensembl Variant Effect Predictor v73 (VEP) (McLaren, Pritchard et al. 2010). Known variants were annotated by dbSNP and unannotated variants with serious predicted consequences were identified based on SIFT and polyphen which were considered as novel variants. Variants were filtered for increased accuracy using following steps: a) variants were filtered at the read depth (DP) >= 10 b) Variants with >10% i.e > 0.1 minor allele frequency based on 1000 Genome project ((http://www.1000genomes.org/data)

References:

ADZHUBEI, I.A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V.E., GERASIMOVA, A., BORK, P., KONDRASHOV, A.S. and SUNYAEV, S.R., 2010. A method and server for predicting damaging missense mutations. *Nature methods,* **7**(4), pp. 248-249.

KUMAR, P., HENIKOFF, S. and NG, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols,* **4**(7), pp. 1073-1081.

LI, H. and DURBIN, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England),* **26**(5), pp. 589-595.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. and 1000 GENOME PROJECT DATA PROCESSING SUBGROUP, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England),* **25**(16), pp. 2078-2079.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. and DEPRISTO, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research,* **20**(9), pp. 1297-1303.

MCLAREN, W., PRITCHARD, B., RIOS, D., CHEN, Y., FLICEK, P. and CUNNINGHAM, F., 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics (Oxford, England),* **26**(16), pp. 2069-2070.

Citation: Abdelaziz Tlili,Abdullah Fahd Al Mutery,Mona Mahfood,Walaa Kamal Eddine Ahmad Mohamed and Khalid Bajou Whole-exome sequencing and bioinformatics analysis. **protocols.io**

dx.doi.org/10.17504/protocols.io.jhscj6e

Published: 21 Aug 2017

Protocol