Updated on 09/27/2018

Whole genome (WGS) or whole exome (WES) sequencing analysis

Sequencing data is first checked by FastQC for quality control and then aligned to the human (or corresponding organism) "indexed" reference genome (hg38) using PEMapper [REF 1]. Sequences with 150bp reads are mapped at a 95% stringency. 100bp reads are mapped at 90% stringency. Variants are called using PECaller in a joint calling framework. Variants are annotated using bystro.io [REF 2].

QC checks, including sequencing depth, mapping percentage, Ts/Tv ratio, Silent/Replacement site ratio, theta, theta in exons, fraction of non-reference homozygote calls, and number of novel variants called. Samples deviating by more the three standard deviations from the study mean will be carefully examined and likely excluded from analysis. Samples are also excluded if they have >5% missing calls at sites or missing >10% of calls across all samples.

A VCF file is generated containing all variants for all samples. This file is suitable for downstream analyses of all types. Bystro annotation information is also provided.

Required:

1. Raw data files (fastq)
2. Metadata spreadsheet with sample information

Deliverables of WGS/WES data analysis service:

1. Variant call format (VCF) file with all variants identified in sequenced individuals
2. Bystro variant annotation information

References:

1. Johnston, H. Richard, Pankaj Chopra, Thomas S. Wingo, Viren Patel, Brain International Consortium on, Syndrome Behavior in 22q11.2 Deletion, Michael P. Epstein, Jennifer G. Mulle, Stephen T. Warren, Michael E. Zwick and David J. Cutler. "Pemapper and Pecaller Provide a Simplified Approach to Whole-Genome Sequencing." *Proceedings of the National Academy of Sciences*, (2017): 201618065.
2. Alex V. Kotlar, Cristina E. Trevino, Michael E. Zwick, David J. Cutler and Thomas S. Wingo. "Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale."*Genome Biology, (*2018): 19:14

Note: For sequencing data acquisition please contact Emory Integrated Genomics Core (EIGC@emory.edu).

Questions? Comments?

Please email us at EICC@emory.edu