Standard Exome v5 Sequencing Methods - Main Text

Whole exome libraries are constructed and sequenced on an Illumina HiSeq 4000 sequencer with the use of 151 bp paired-end reads. Output from Illumina software is processed by the Picard data-processing pipeline to yield BAM files containing well-calibrated, aligned reads. All sample information tracking is performed by automated LIMS messaging.

Library Construction

Library construction was performed as described by Fisher et al. with some slight modifications. Initial genomic DNA input into shearing was reduced from 3µg to 50ng in 10µL of solution and enzymatically sheared. In addition, for adapter ligation, dual-indexed Illumina paired end adapters were replaced with palindromic forked adapters with unique 8 base index sequences embedded within the adapter and added to each end.

In-solution hybrid selection for exome enrichment

In-solution hybrid selection was performed using the Illumina Rapid Capture Exome enrichment kit with 38Mb target territory (29Mb baited). The targeted region includes 98.3% of the intervals in the Refseq exome database. Dual-indexed libraries are pooled into groups of up to 96 samples prior to hybridization. The liquid handling is automated on a Hamilton Starlet. The enriched library pools are quantified via PicoGreen after elution from streptavadin beads and then normalized to a range compatible with sequencing template denature protocols.

Preparation of libraries for cluster amplification and sequencing

Following sample preparation, the libraries prepared using forked, indexed adapters were quantified using quantitative PCR (purchased from KAPA biosystems), normalized to 2 nM using the Hamilton Starlet Liquid Handling system, and pooled by equal volume using the Hamilton Starlet Liquid Handling system. Pools were then denatured using 0.1 N NaOH. Denatured samples were diluted into strip tubes using the the Hamilton Starlet Liquid Handling system.

Cluster amplification and sequencing

Cluster amplification of the templates was performed according to the manufacturer's protocol (Illumina) using the Illumina cBot. Flowcells were sequenced on HiSeq 4000 Sequencing-by-Synthesis Kits, then analyzed using RTA2.7.3

References

Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C. *A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries*. Genome Biology 2011, 12:R1.