Exome sequencing: Methods

Pablo Cingolani

1 Methods

We performed exome sequencing of twelve samples using Illumina Genome Analyzer GAIIx with a Paired End Cluster Generation Kit. Obtaining 242,719,043 pair-ended reads, length 100 bp. A total of 48.5 Giga-bases. Sequence extraction and base calling was done using Illumina Pipeline v1.6 software. Image analysis and base-calling was done using *SCS* (Sequencing Control Software).

Read mapping was done using the standard Illumina pipiline which uses Eland mapping algorithm. Reference genome used was hg37.61, which was the latest version at the moment. Reads mapped successfully accounted for 86% of the original reads.

Multi-sample variant calling was performed using SamTools 0.1.12a [2] package and BcfTools (same authors).Variant analysis and statistics were performed using our in-house development tool SnpEff[1] (version 1.0301).

We obtained over **563,000** variants form all 'case' samples. These variants were filtered using the variants obtained from the 'control' group and DbSnp-132. The remaining variants were further filtered using quality values over 30 and at least coverage of 3. After filtering, we obtained over **36,000** novel high quality variants.

References

[1] Pablo Cingolani. SnpEff: Variant effect analysis. *snpeff.sourceforge.net*, 2011.

[2] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078, 2009.