NGS

Libraries were prepared using the standard TruSeq protocol (Illumina, San Diego, CA). The libraries were validated and quantified on an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) using the DNA 1000 Nano Chip kits.Samples were allocated to flow cells using an incomplete block design that allocated all three samples from the same subject to the same or consecutive flow cells, using the initial randomization that provided balance between the discovery and replication cohorts. Sequencing was carried out on Illumina’s HiSeq 2000 (Illumina, San Diego, CA). Samples were sequenced as 51 cycle, single end reads. Protocols were upgraded on 6/28/2011.

The reads from Illumina’s HiSeq 2000 were aligned to the human genome build 37.1 using TopHat (1.3.3) and Bowtie (0.12.7). HTSeq (0.5.3p3) was used to perform gene counting while BEDTools (2.7.1) was used to count the reads mapping to individual exons (*Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111; Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25. doi: 10.1186/gb-2009-10-3-r25; HTSeq: www-huber.embl.de/users/anders/HTSeq/doc/overview.html* [*www-huber.embl.de/users/anders/HTSeq/doc/overview.html*](http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html)*; Quinlan AR and Hall IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 6, pp. 841–842*).

Results are given as counts per gene.