

Jordan University of Science and Technology

Experiences with array-based sequence capture; toward clinical applications.

Authors: Almomani R, van der Heijden J, Ariyurek Y, Lai Y, Bakker E, van Galen M, Breuning MH, den Dunnen JT.

Abstract: Although sequencing of a human genome gradually becomes an option, zooming in on the region of interest remains attractive and cost saving. We performed array-based sequence capture using 385K Roche NimbleGen, Inc. arrays to zoom in on the protein-coding and immediate intron-flanking sequences of 112 genes, potentially involved in mental retardation and congenital malformation. Captured material was sequenced using Illumina technology. A data analysis pipeline was built that detects sequence variants, positions them in relation to the gene, checks for presence in databases (eg, db single-nucleotide polymorphism (SNP)) and predicts the potential consequences at the level of RNA splicing and protein translation. In the samples analyzed, all known variants were reliably detected, including pathogenic variants from control cases and SNPs derived from array experiments. Although overall coverage varied considerably, it was reproducible per region and facilitated the detection of large deletions and duplications (copy number variations), including a partial deletion in the B3GALT1 gene from a patient sample. For ultimate diagnostic application, overall results need to be improved. Future arrays should contain probes from both DNA strands, and to obtain a more even coverage, one could add fewer probes from densely and more probes from sparsely covered regions.