

IMPROVED MOLECULAR DIAGNOSIS USING THE GENSEARCH[®] SOFTWARE PACKAGE: THE EXAMPLE OF DISEASE-CAUSING MUTATION INTERPRETATION IN FAMILIAL HYPERCHOLESTEROLEMIA

Amsellem S, Benlian P*.

INSERM U538, Faculté de Médecine Saint Antoine, Paris.

Pascale.benlian@sat.ap-hop-paris.fr

Detection of novel point mutations has greatly improved in precision and throughput with the development automated capillary gene sequencing. However, at the step of DNA sequence tracks analysis with available software packages, and interpretation of disease-causing of variations, the process remains manual and time-consuming. In addition, intronic variants or small insertion or deletions may be skipped from analysis or interpretation. Finally, reference set-up for a new locus requires tedious steps.

Gensearch[®] (Phenosystems[®]) was developed to overcome these difficulties. An algorithm compares the electronic signal of a diagnostic peak processed by the PHRED base-caller, with that of a reference peak, and neighboring peaks on the same track, allowing automated analysis of peak significance in terms of true DNA variation, at a specified location in a gene locus. The semi-automated referencing system is made by a download of reference sequences from public databases, to automatically format any locus by its exon-intron organization even when alternatively spliced. The automation of peak analysis process provides a user-friendly output with interpretation of exonic, intronic or frameshift mutations, or synonymous SNPs annotated according to the HGVS nomenclature. It also provides links to interpret the medical significance in terms of predicted protein or splice site structure or function; build haplotypes from series of SNPs, in comparison with a personal or public database at this locus. It allows the user to validate a newly identified DNA variant, using criteria required by the HUGO consortium. This type of semi-automated process might be useful for the diagnosis of novel mutations at loci with high allelic heterogeneity.

Familial hypercholesterolemia (FH) is a dominantly inherited disease of cholesterol transport. To date, 3 loci (LDLR, APOB, PCSK9) have been involved as causes of FH. In most cases, LDLR mutations are the leading cause for FH (50-75% of cases). More than 900 disease-causing alleles have been reported (<http://www.ucl.ac.uk/fh>; <http://www.umd.necker.fr:2004/>). In highly admixed populations such as the French population, allelic heterogeneity prevails. We used FH as a model to test software efficiency to provide faster and reliable results in terms of interpretation of detected DNA variants as disease-causing mutations at the 3 loci. Genomic DNA sequencing of the LDLR locus (18 amplicons), part of APOB exon 26 and the PCSK9 (13 amplicons), (ABI 3100, Big Dye terminator chemistry, Applied Biosystems) was performed previously for diagnostic purposes in 636 unrelated families (i.e. ≈14.000 amplicons). The Seqscape[®] software package, then Gensearch[®] were run blinded on this dataset. In 436 families, 210 mutations were found on either locus, including 201 different LDLR mutations (58% MS, 20% del/ins, 10% NS, 8% splice). In every case, interpretation of heterozygous del/ins mutations was made possible using Gensearch[®], a case when alignment-based softwares, might skip these mutations. In three cases, a base substitution previously undetected (lack of sensitivity of base-calling) was confirmed. Real-time comparison with a private mutation database, or a direct link to publicly available tools (BLAST, Splice site prediction, etc.) was possible. Moreover, the negative mutation carrier status was supported by a holistic quality analysis of the sequenced locus.

Computed-assisted analysis of DNA sequences using Gensearch[®] appeared robust, time-saving to obtain relevant information for disease-causing mutation interpretation on a large dataset.