ADVANCEMENT IN ALLELE-CALLING ACCURACY AND THROUGHPUT

Sasha Gutin, Thomas Scholl and Alun Thomas
Myriad Genetic Laboratories

Myriad Genetic Laboratories has developed software applications designed for high performance STR analysis performed on automated fluorescent sequencers (ABI™ model 377). These analysis programs were designed specifically to improve upon aspects of currently available software.

Five major areas of improvement will be described: algorithms for data extraction, identification of appropriate peaks, automated characterization of potential data quality problems, integration of a database with the software, use of information compiled within the database to optimize the system.

Several aspects of raw data extraction were improved. A gel-tracking program with increased accuracy for identifying sample lanes was developed. Dye cross talk analysis that calculates a matrix for each gel lane was implemented in place of the machine based matrix. . An improved signal discrimination algorithm was designed that alleviates commonly encountered problems with high localized background.

The accuracy of fragment allele assignments was improved through approaches that identify and automatically exclude from the analysis artifacts attributable to PCR stutter, plus A addition, dye matrix problems, and detector saturation. The algorithms used to exclude anomalous signals have adjustable thresholds that can be used to refine combinations. These thresholds can be defined to include or to exclude peaks from allele assignment, or to flag peaks for review by an operator. When data is flagged, the program displays the specific data to the operator with a designation of the type of problem to be resolved.

These programs are combined with a database into a coherent genotyping platform that improves allele calling automation and accuracy by compiling results and employing that information to optimize the thresholds and the algorithms used in allele assignments. For example, specific markers can be characterized for their potential to produce stutter peaks and specific allele combinations can be characterized for their ratio of signal amplitudes. With this understanding, the fragment designation algorithms can be modified to improve allele-calling accuracy or to prevent unnecessary operator intervention. Notable, operator intervention was identified as a major source of error. When operator intervention is minimized, benefits are accrued in both accuracy and throughput. The association of sample data for a variety of parameters into the database also yields the potential to develop variety of quality control parameters.

In conclusion, improved software for STR analysis is described. These improvements affect the algorithms used to extract data and the approaches to identify appropriate peaks and assign alleles. The association of these programs to a database permits the accumulation of results that are used to further refine the analysis. The application of this system produced an overall improvement in allele calling accuracy and throughput.