Validation of a Software Tool for Clinical Sequencing Analysis

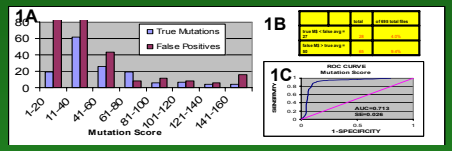
J. Uhrmacher[1], F. Monzon-Bordonaba[1], A. Carter[1] and J.Kant[1, 2]
Departments of Pathology[1] and Human Genetics[2], University of Pittsburgh Medical Center, Pittsburgh, PA 15213

Introduction: The steady increase in volume of clinical tests involving DNA sequence analysis has emphasized inefficiencies associated with manual inspection of electropherograms (EPs). Software for 'automatic' analysis of sequence traces raises apprehensions that programming algorithms by themselves or in conjunction with low quality technical results might miss important nucleotide variants seen on manual inspection. We evaluated version 2.5 of the Mutation Surveyor Program (MS2.5) by SoftGenetics, LLC (State College, PA) against manual sequence analysis to see if this tool could be used to streamline analysis of clinical resequencing runs. MS2.5 compares a sample sequence with both a GenBank reference sequence and an actual aligned reference EP, then calculates a quality score (q-value) for the overall EP based on average base resolution quality as well as a confidence score (mutation score, M$) for each significant nucleotide variant encountered. The ability of MS2.5 to accurately interrogate electropherograms and identify sequence variants, to determine when repeat analysis was required, and to serve as a user-friendly system to facilitate construction of clinical reports was assessed.

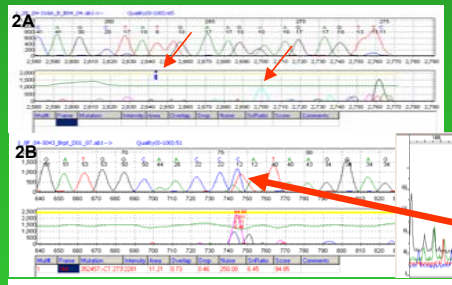
Materials and Methods: Archived EP files from clinical *SDHD* and *SDHB* resequencing runs over a nine month period were analyzed and compared with manual interpretations. Sequencing assays sample full exon and flanking intron/untranslated sequences bidirectionally from forward and reverse strands. Cycle sequencing was performed using ABI Big Dye Terminator Ready Reaction v.1.1 kits; results were analyzed on an ABI 3100 Capillary Electrophoresis Analyzer with 50cm arrays and POP6 polymer. Laboratory professionals other than faculty who did the original manual analysis performed the software reanalysis.
Each sequence run was re-analyzed in its entirety, rather than each clinical case individually. Data was exported to an Excel 2002 SP3 spreadsheet and sorted into individual cases. Variants called by MS2.5 software were required to have an identical locus and variant in both directions, confident mutation score, and clarity of the EP display (MS2.5 Graphic Analysis Display) on review.
Variants were categorized as follows:

- 'True mutations': variants of known and uncertain clinical significance
- Non-pathogenic variants
- Intron polymorphisms

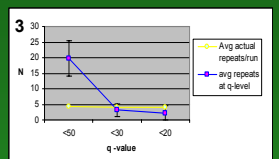
Other variants (e.g. one direction only) were regarded as 'False positives."
Run statistics were compiled and compared to the diagnostic determination in each case.



↑ **Figure 1. Distribution of Mutation Scores.**
1A. Frequency distribution of N=602 software calls separated into "true mutation" (TM) and "false positive" (FP) categories on basis of the EP display. Most scores were <60. Scores of <40 were most likely FP, 60-80 were most likely a TM while >140 again more likely a FP.
1B. Mutation Score Overlap.
1C. Receiver Operating Characteristic (ROC) Curve for Mutation Score discriminating true mutations. The area under the curve (AUC) indicates that using the mutation score alone to distinguish a mutation is only a fair method (AUC and standard error by Hanley and MacNeil approximation).



↑ **Figure 2. Mutation Surveyor Graphic Analysis Display.**
The sample EP is the top panel and the MS Mutation plot is on the bottom.
2A. A case of a single, unmatched call of a true mutation: the variant was called on the complementary strand, but the EP shows two of the reasons for no matching call. This variant lies near the 3' end of the strand and outside of the software analysis range (blue line). Generally, at these extremes there are non-linearities and poor base-calling or the sequencing reaction just terminates. Additionally in this sample the mutation score does not reach a set threshold value. Both can be corrected by software options, but this occurred in about 10% of true variants. The need to have a protocol running duplicates and to review the EP opposite a called variant is clear.
2B. A high scoring false positive. The variant is an anomaly caused by an electrophoretic "bubble" which can be easily seen in both the EP and the raw data (on right). These anomalies do not repeat and do not match the opposite strand.



↑ **Figure 3. Quality (q) score evaluation and utility.**
Baseline quality and peak resolution of the EP contribute to the q-score. The average theoretical number of repeats required from each run using different q-values as a critical threshold is mapped against the mean of actual repeat requests (following its own internal set of decision rules). Intersection of the critical-q curve with the empirical values suggest a q-score utility as a decision tool lying in the range of 25-30.

Results: 199 cases with a total of 1820 sequence files were reanalyzed. Average EP quality (q-score) was 65.2 ± 6.4. Average size (reflecting length-of-read) for *SDHD* and *SDHB* products indicated >84% of theoretical maximum data capture. With this approach MS2.5 made a total of 661 calls. There were 14 unmatched files due to failed reactions or bad data (0.9%). There were 57 total indel calls which did not contribute to mutation score statistics.
"False positive" results were generally lower scoring calls by the software though some at the extremes of the EP were very high scoring (**Figure 1A**). Mutation scores for "true mutations" averaged 49.7 ± 12.3 and for false positives averaged 27 ± 13.4. Within each run, the mutation score by itself was only a fair discriminator between true mutations and EP artefacts, since the distributions were overlapped, some true mutations had low scores and some false positives had very high scores (**Figures 1B,1C**).
The software identified all sequence variants reported clinically in at least one direction in all cases (100% concordance with manual review). No pathogenic mutations were missed. Single calls were generally polymorphisms at the extremes of sequence on one strand only (identical what is seen in manual review where the variant was out of analysis range) or when the mutation score did not reach a set discrimination threshold. Generally these were single direction conflicts readily seen in the graphic display to be electrophoresis artifacts, e.g. compressions or bubbles (**Figure 2B**, **Figure 3**). Miscalls in base numbering were only seen with frameshift mutations which resulted in software calls of nonidentical bases between the strands or differences in cDNA numbering which could be understood from the surrounding sequence.
The q-value minimum score which indicated inspection of the whole EP and a need for repeat analysis was determined to be <30 (**Figure**

Conclusions:

- By focusing on differences in electropherograms which highlight potential nucleotide variants, sequence analysis software reduces analysis time for large runs from ~1 day to <2 hours. Quality scores and analysis of length data assist in evaluating individual lanes or entire runs for suitability.
- The use of a software tool such as MS2.5 also enables analysis to be structured through a computer network instead of using hard copy EP's, thus saving tech time and facilitating archiving and retrieval of primary data.
- Identification of true mutations is aided by bidirectional sequencing and requires checking of forward and reverse calls at the same locus. Indels should be examined for correct base calling and numbering.
- Though not perfect, with modest manual oversight, MS2.5 sequence analysis software was a clinically-friendly and useful tool.

UPMC | University of Pittsburgh Medical Center
Affiliated with the University of Pittsburgh schools of the health sciences