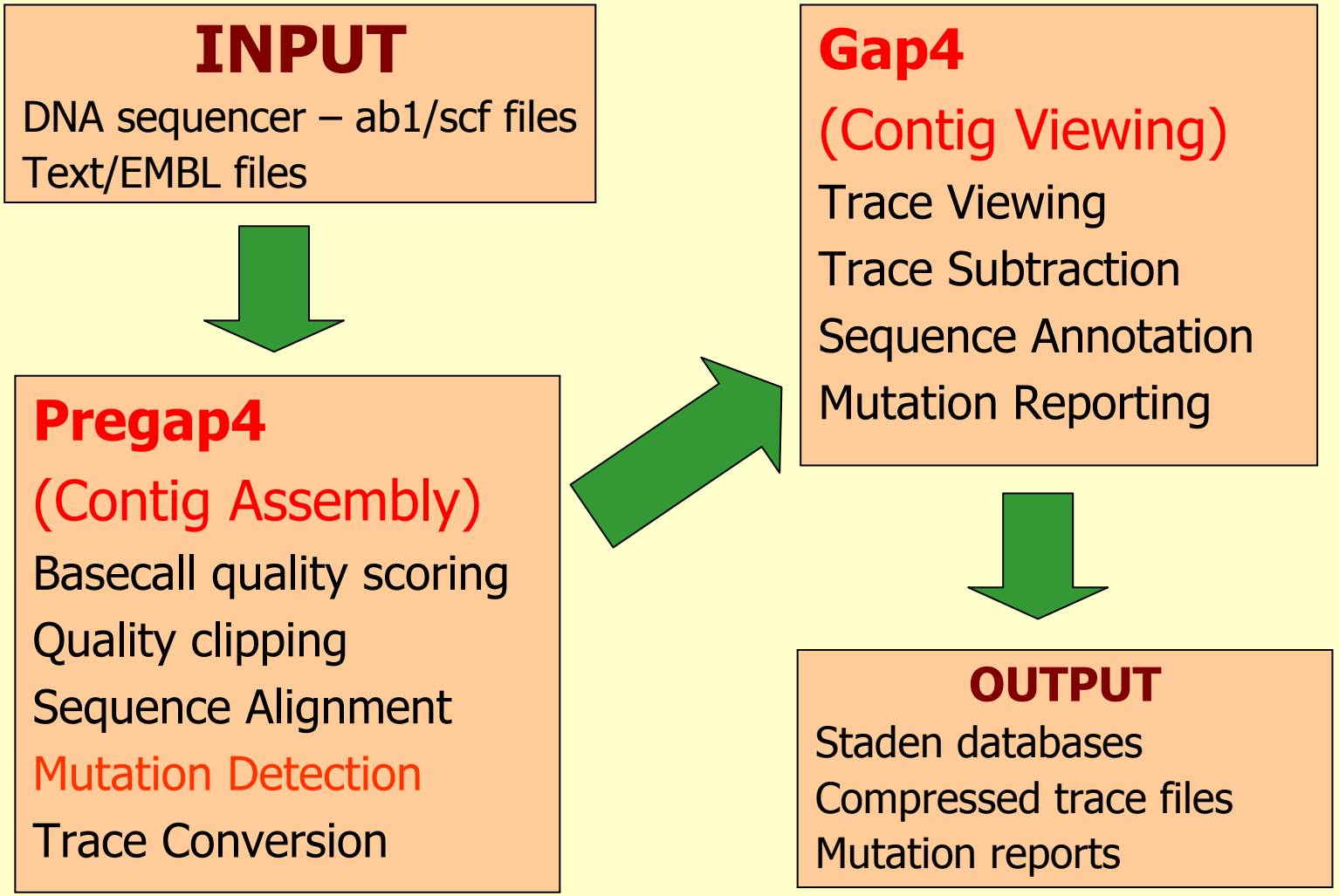
Staden Development

Andrew Wallace
NGRL (Manchester)

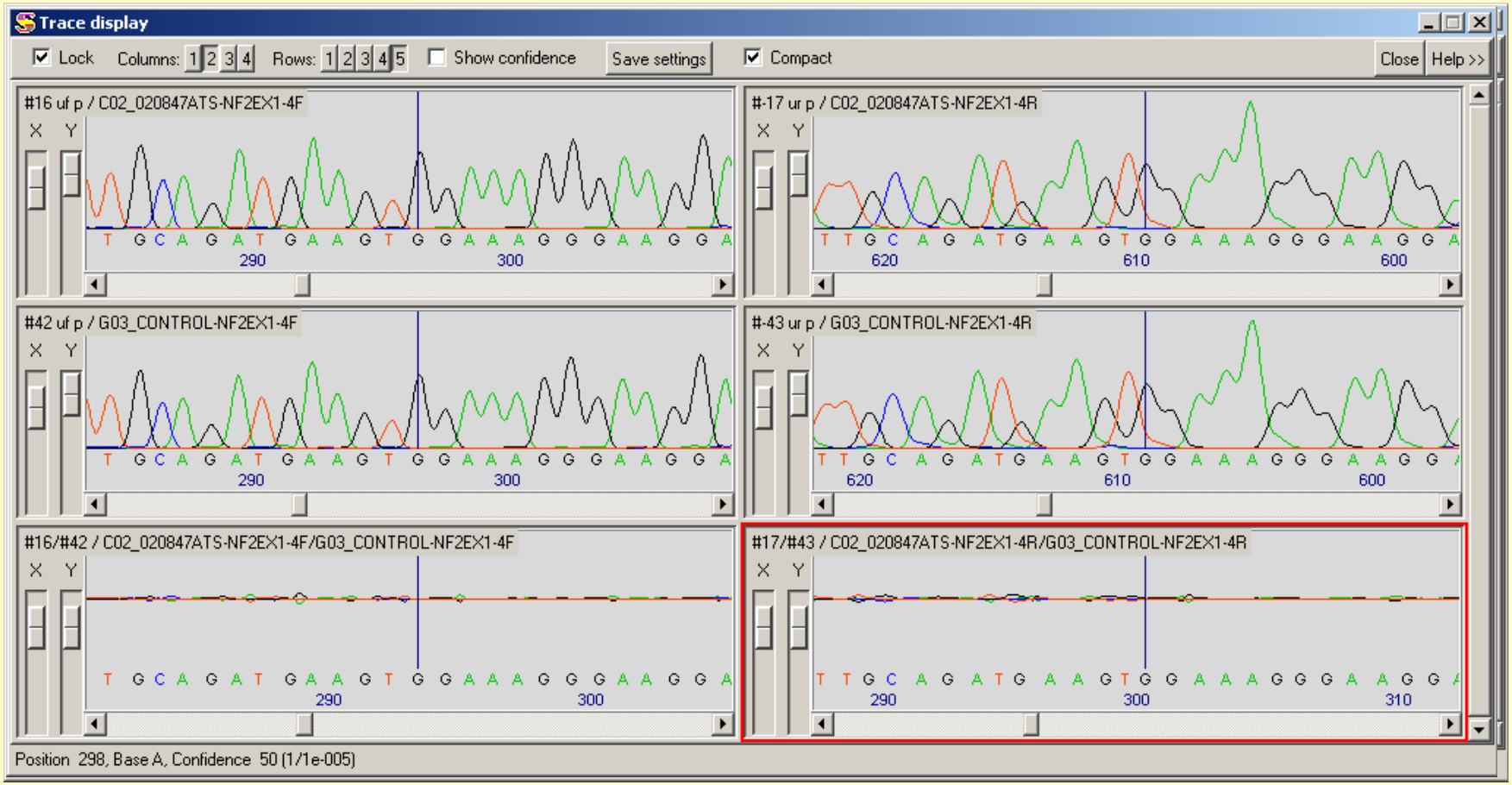
What is Staden?

- Staden is a package of sequence analysis programs
- Developed by Rodger Staden of MRC-LMB Cambridge
- Originally produced for genome sequencing projects
- Incorporates several features useful for mutation scanning sequence data

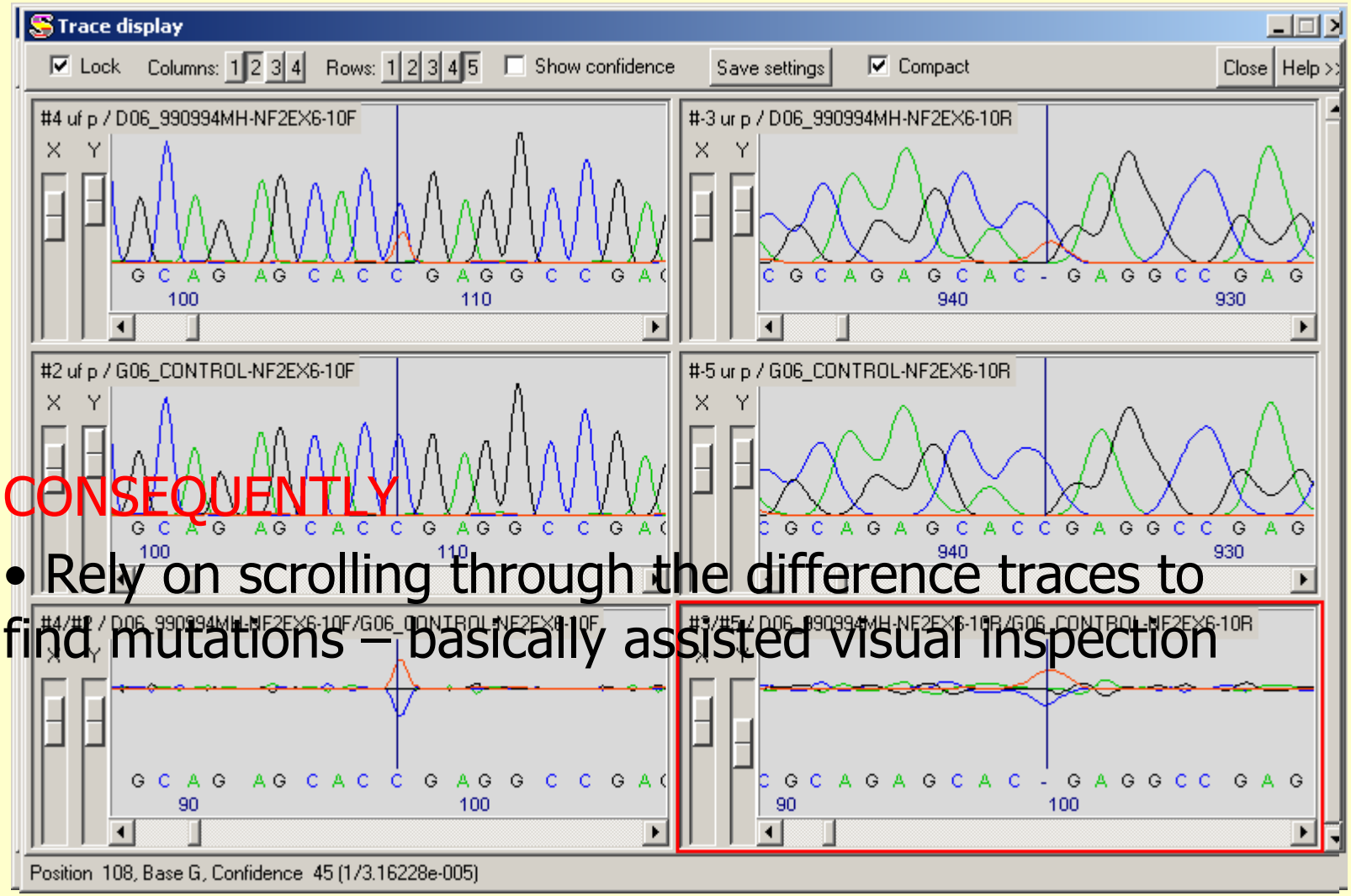
Staden Package & Data Flow



Detecting Mutations using Trace Subtraction



Automated Mutation Scanning



CONSEQUENTLY

- Rely on scrolling through the difference traces to find mutations – basically assisted visual inspection

Staden – Plus Points

- **Comparative**
- **Trace subtraction**
- **Free!**
- **Platform independent**
- **Iterative development**

Staden – Minus Points

- **Unreliable heterozygote scanning algorithm**
- **Complex user interface**
- **Produces a multitude of files**
- **Manual file management - a big problem!**
- **Currently unsupported – "open source"**

Staden Development

- NGRL (Manchester) took a decision to fund diagnostic development
- IP agreement with MRC technologies
- Certus Technologies appointed to carry out development
- Shopping list of “system requirements”
- First step was to “reverse engineer” Staden

Development Shopping List

- No false negatives
- Low false positive rate
- Use raw data
- HGVS mutation naming in Mutation reports
- Quality measure of sequence data
- Integration of sub-tests > patient tests
- Reference sequence editor, full annotation of reference sequences & database of sequence variations
- Full audit trail
- Fully automated scanning (no file management, files processed as produced by sequencer)

Staden Mutation Scanner Algorithm

FIXED

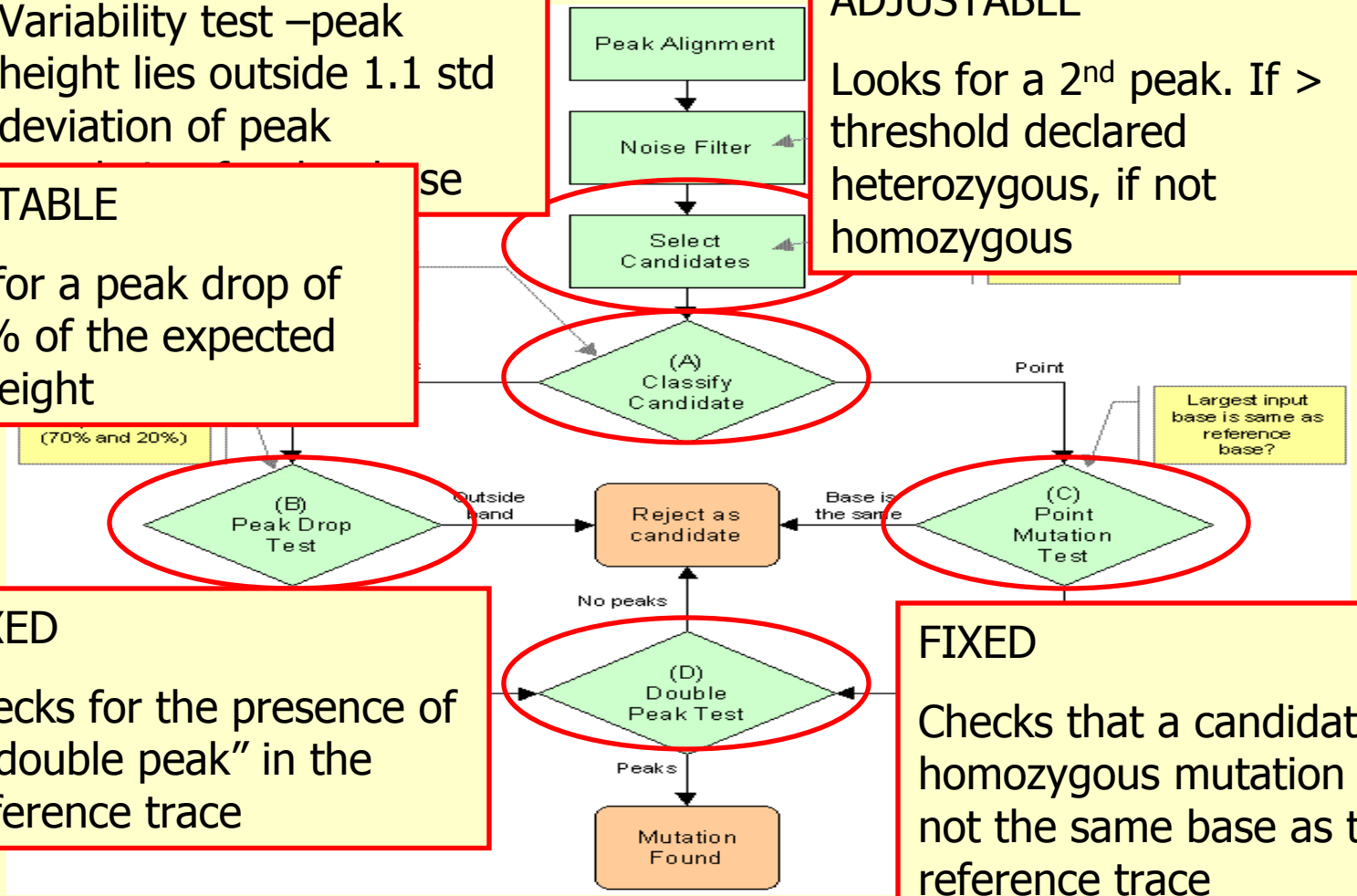
Variability test –peak height lies outside 1.1 std deviation of peak

ADJUSTABLE

Looks for a peak drop of 20-70% of the expected peak height

ADJUSTABLE

Looks for a 2nd peak. If > threshold declared heterozygous, if not homozygous



FIXED

Checks for the presence of a "double peak" in the difference trace

FIXED

Checks that a candidate homozygous mutation is not the same base as the reference trace

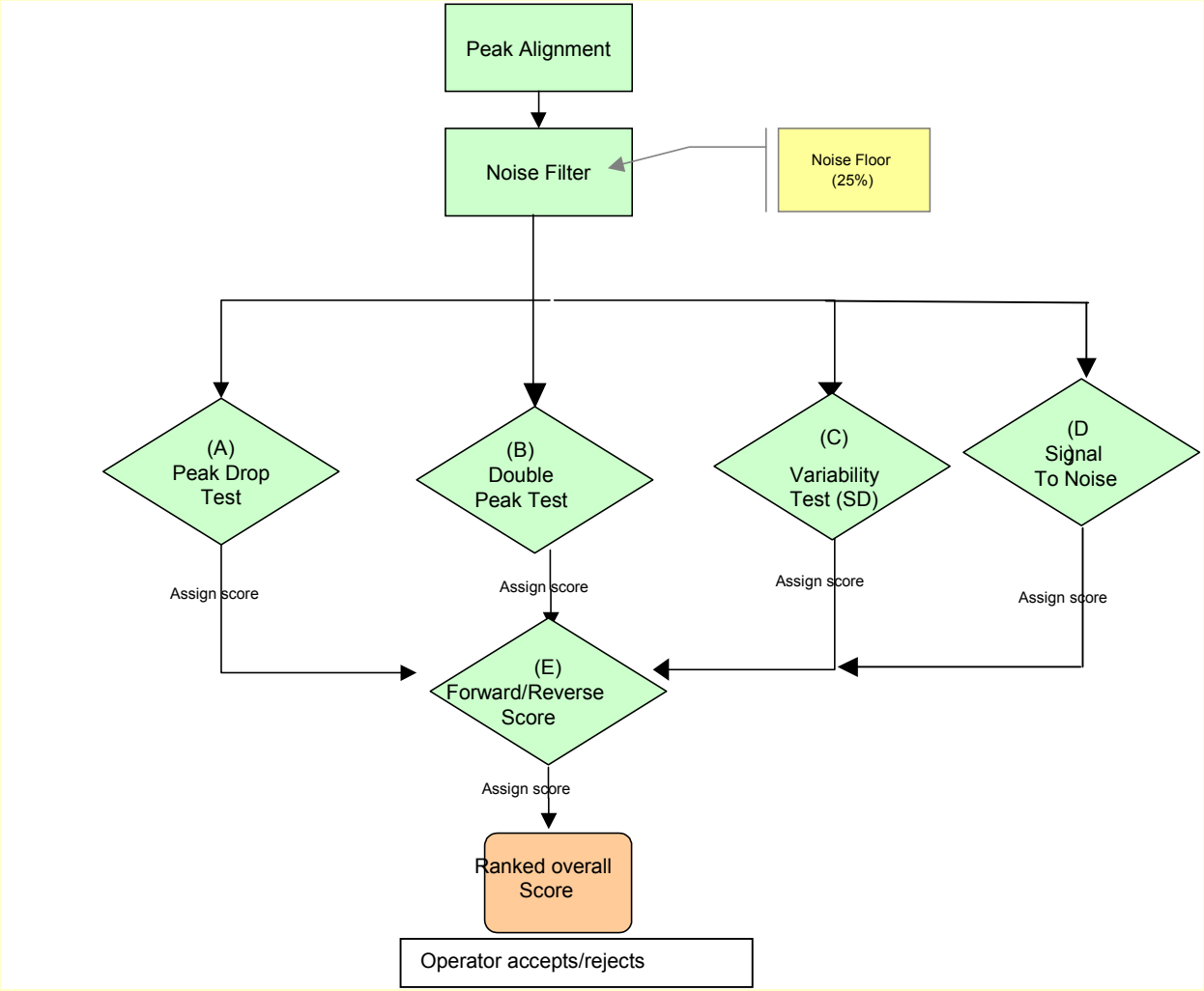
Re-design of Mutation Scanner Algorithm

- **No preset thresholds** - carry out tests in parallel rather than follow a decision tree
- **Collect ALL data** on test vs reference trace comparisons NOT select
- **Reconcile data -** from forward and reverse reads
- **Rank data according to mutation score** – use a presentation tool to sort the data according to the candidates with the highest mutation scores

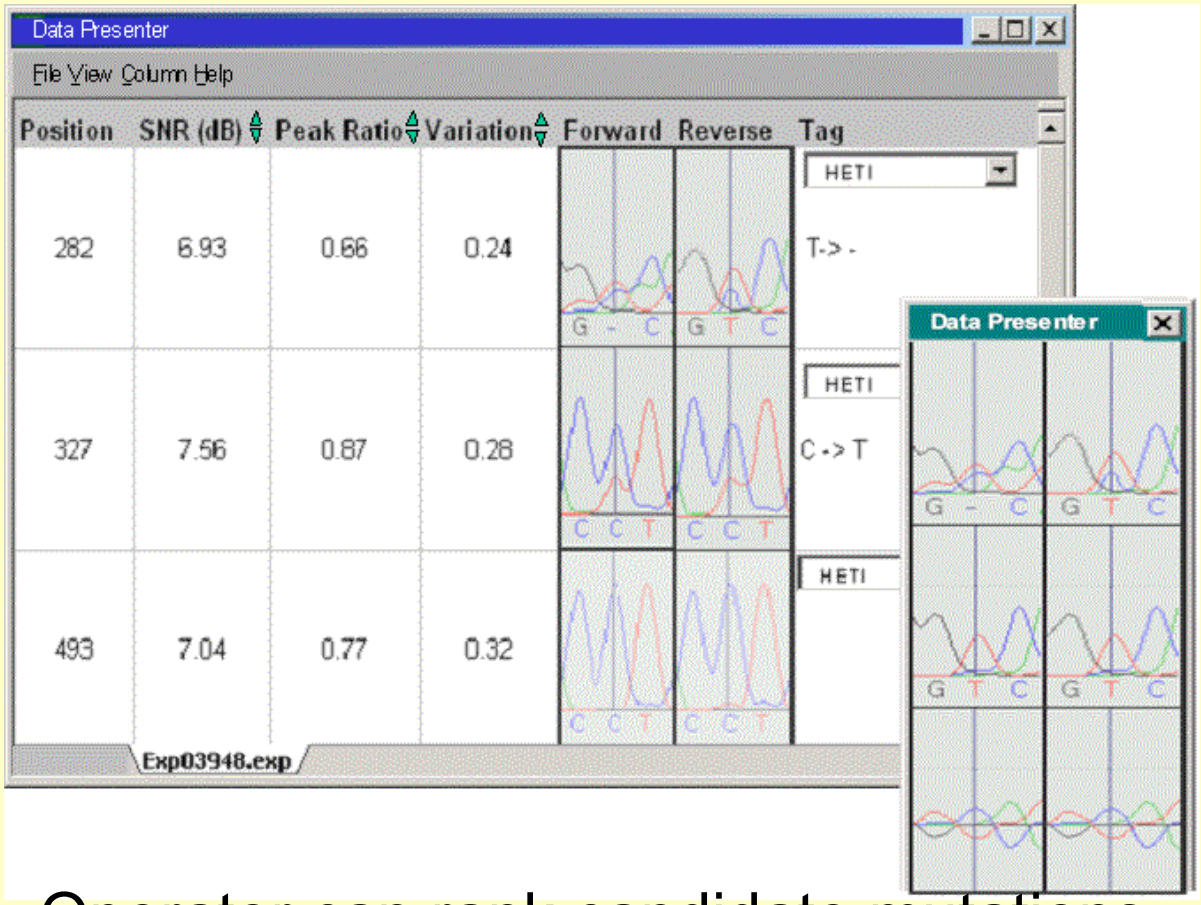
User then scans down the candidates until they are satisfied that the differences are "noise"

Approach has been tested on mutations not detected by current version of Mutation Scanner

Revised Mutation Scanning Algorithm



Data Presenter -User Interface



Operator can rank candidate mutations according to any test parameter

Operator accepts/rejects candidates

Data Presentation Tool

- To be supplied as an “add-on” tool running alongside Gap4
- Work with the modified Mutation Scanner incorporated into PreGap4
- Accepted candidates produce a mutation report which includes standard HGVS name
- Freely available for evaluation – ?Autumn ‘04

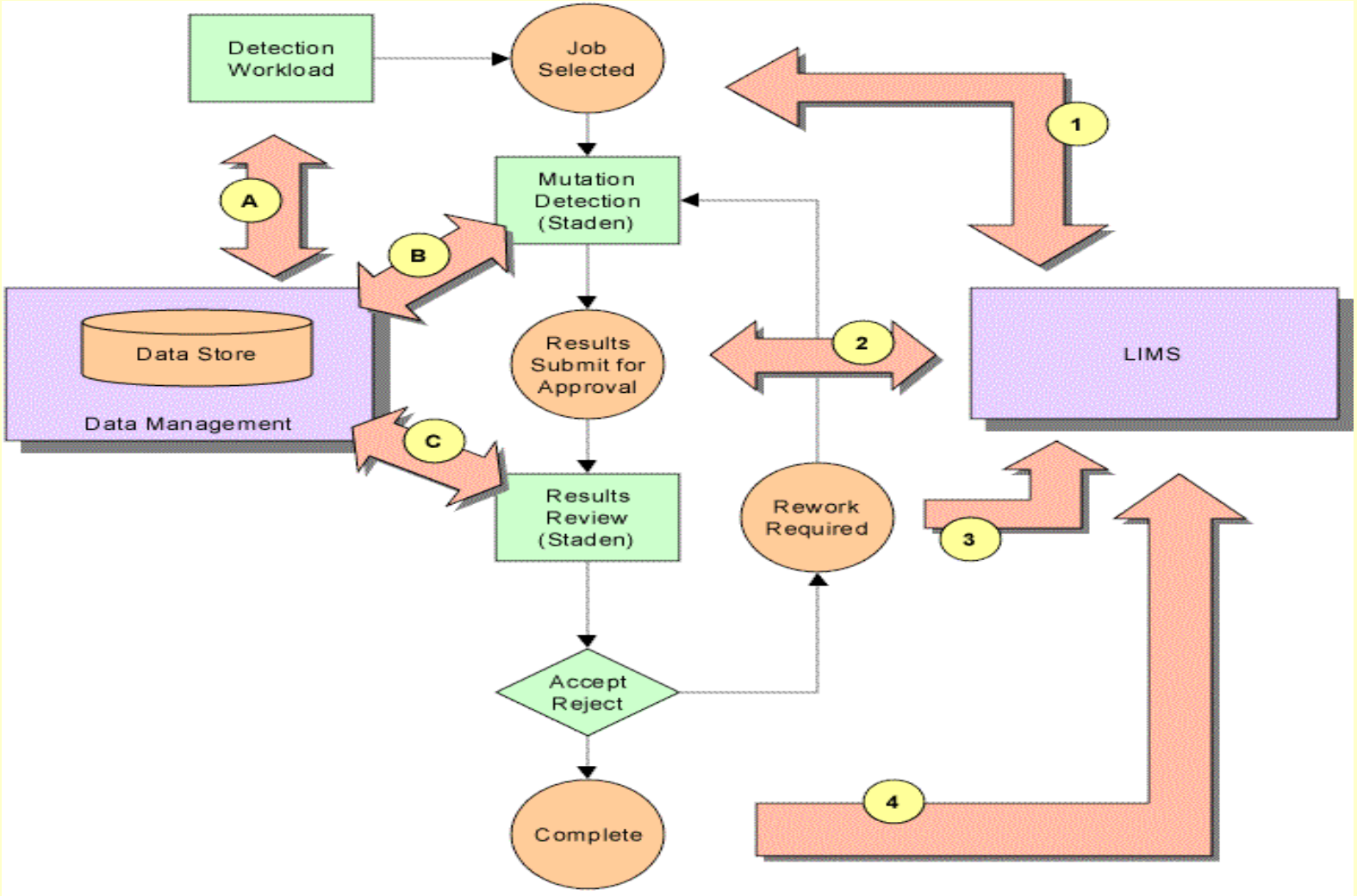
Comparison with Alternatives

In conjunction with NGRL(W) Plan to carry out a comparison between software packages

- SeqScape – non-comparative
- Mutation Surveyor – comparative, uses 4 criteria
- Paracel Agent - non-comparative, uses confidence estimates
- Gensearch - comparative
- CSA – comparative

Propose to measure ease of use, operator time, false negative rate, false positive rate on an identical data set

Future Plans



Conclusions

- Staden package under active development
- Modified Mutation Scanner algorithm
- Improved Data Presentation
- Will add-on to current Windows version of Pregap4 & Gap4
- Plans for further remodelling
- Comparison with other software

Acknowledgements

- Ian Bamsey
- Robert Pumphrey
- Rodger Staden
- Mark Jordan
- David Gokhale