

TITLE: **Frame shift/Edit Repair**

PAGE: 1 of 2

SOP #: PA07

REVISION LEVEL: .2

EFFECTIVE DATE: July 2006

AUTHOR:
Ramana Madupu

PRIMARY REVIEWER:
Bill Nelson

1 OVERVIEW

Gene models whose pair-wise alignments show frameshifts or in-frame stops relative to known genes are evaluated to allow detection and repair of any sequencing errors, or annotation of identified mutations.

1.1 Scope

The pair-wise alignments of all gene models are checked for potential frameshifts and in-frame stops. Any gene identified in this analysis is subjected to review.

1.2 Related Documents

SOP PA01 – Gene Prediction

SOP PA02 – Homology Searches

[Naming Convention Guidelines](#)

1.3 Revision History

| Author | Date | Change |
|-------------|--------------|----------------------------------|
| Bill Nelson | 14 July 2006 | Basic edits and addition of SOP# |

2 REQUIREMENTS

A genome sequence.

Frame shift reports

3 PROCEDURE

3.1 Evaluation Criteria

A gene model should be examined if the frameshift (FS) and/or in-frame stop (PM) is:

- present in the majority of BER alignments.
- in the same approximate region(s) in each alignment

3.1.1 Sequence Verification

If the bulk of the alignments is suggestive of a FS, PM or sequence ambiguity, the sequence around this region is verified. The sequence coverage and quality is checked by examining the underlying chromatograms for sequencing or editing errors.

TITLE: **Frames shift/Edit Repair**

SOP #: PA07

REVISION LEVEL: .2

PAGE: 2 of 2

3.1.1.1 Low Sequence Coverage/Poor Sequence Quality

If the region of interest has low sequence coverage (<3X) or the sequence quality is poor, the region is re-sequenced. If an error has occurred, the assembly is edited and the gene model re-evaluated.

3.1.1.2 High Sequence Coverage/Good Sequence Quality

If the sequence is correct, the curator determines whether any special cases such as programmed frameshift, incorporation of selenocysteine, or expression of a multi-domain protein as separate proteins explains the apparent FS or PM.

3.1.2 Gene Model Edit

The gene model is edited to reflect the findings of the analysis. Frequently a gene with a FS or PM is called by the gene prediction software as two adjacent genes. Surrounding genes are checked to see if two gene models should be merged. Models can be extended to include alignment regions in other frames. Annotation is curated in accordance with the naming convention guideline.

3.1.3 Other Gene Disruptions

Other type of gene disruptions (truncations, interruptions, internal deletions and fragmented genes) are also frequently identified in this analysis. These genes are evaluated and curated according to naming convention guideline.

4 DATA MANAGEMENT

4.1 *Quality Control*

The final dataset is reviewed for gene models that don't have a canonical start or stop codons, models with lengths indivisible by 3, models with in-frame stop codons, and genes with annotation that suggests a FS or PM, but doesn't meet the other criteria.