**SUPPLEMENT MATERIALS**

**PeptoVar** **MODES**



**Fig. 1** Three modes of the PeptoVar program.

**THE ALGORITHM EXPLANATION**

The base steps of PeptoVar workflow:

1) reading input data

2) building graph for the reference nucleotide sequence

3) genomic polymorphisms are introduced to the graph as alternative paths

4) attaching prefixes (sequence to the closest upstream codon start) and suffixes (nucleotides sequence to the closest downstream codon end) to branching paths according to the reading frames (including frames induced from upstream frame shifts)

5) identifying and removing synonymous branches

6) translation of the graph into peptide set by tree traversal algorithm

7) subtraction of the peptide sets to obtain unique peptides for a sample (in transplantation mode)

8) writing the results

The public databases of genomic polymorphisms usually stores information about which genomic variations are non-synonymous and thus should be taken into consideration for population-wide peptidome construction. However this data is in fact context dependent. The same variation can be synonymous and non-synonymous depending on particular combination of upstream frame shifts (Fig. 2). More over, if two polymorphisms are found in the same codon their synonymy depend on each other. For example, in the polymorphic codon (C,T)T(A,T) the third position is synonymous if the first position is ‘C’ and non-synonymous if the first position is ‘T’ (Fig. 3). The PeptoVar program save the dependency data into a separate file. If many variations occur consecutively PeptoVar optimizes combinatorial task removing all synonymous branches before translation to reduce computational time and memory usage (Fig. 4).



**Fig. 2** The variation type (synonymous or non-synonymous) depends on frame shifts in upstream.

PeptoVar considers full set of frame shift combinations in upstream to classify each polymorphism properly.

**(C,T)T(A,T)**

**CT(A,T) => Leu (L)**

**TT(A,T) => Leu (L), Phe (F)**

**Fig. 3** Polymorphisms translation dependency in the polymorphic codon.



**Fig. 4** Combinatorial explosion of translation to protein in case variations occur consecutively. For each variation PeptoVar takes prefixes and suffixes and determines functional paths including translation frames induced from upstream frame shifts.

**PEPTIDE POSITIONING ON A GENOME**

To compare peptides of two samples the peptide positions (coordinates of translated sequences) on the reference genome should be defined. But a peptide will not have defined genome position if its translated sequence starts and/or ends on a genome insertion. For that peptides PeptoVar calculates genome positions as “peptide shadows” on the reference genome (Fig. 5):

*start = leftmost\_translated\_reference\_position - number\_of\_translated\_insertion\_nucleotides;*

*end = rightmost\_translated\_reference\_position + number\_of\_translated\_insertion\_nucleotides;*



**Fig. 5** Peptide positioning on the reference genome sequence.