# **Expanded View Figures**

#### User file

DNA mutations mapped to Ensembl protein sequences.



### Identification of hotspot residues

Identification of hotspot mutation residues.

- Exclusion of common polymorphisms.
- · Calculation of associated p-values.



Parameters that can be adapted: Length of surrounding sequence windows, minimum number of mutations, fraction of mutations that need to map to the same residue, p-value threshold, switch to overrepresentation instead of fraction as a criteria for hotspot definition.

### Annotation of functional residues

Retrieval of UniProt KB annotations for the significant proteins.



Alignment and overlap of functional protein segments with the identified hotspots.

## Figure EV1. DominoEffect R package.

We used the in-house Perl script to map genomic mutations onto protein sequences, and we are happy to share it upon request. The obtained mutation dataset is then be analyzed with the <code>DominoEffect</code> R package. In the first step, the package identifies hotspot mutation residues and in the second step it obtains UniProt-SwissProt KB annotations for the respective proteins and provides information on the protein functional segment affected by the hotspot mutation. In the latter step, it also assures that the Ensembl and UniProt sequence alignments are in agreement. When identifying hotspot residues, the package allows the user to flexibly change different parameters. Hotspot identification and residue annotation can be run separately or jointly by calling the <code>DominoEffect</code> function. The package provides a test mutation dataset and comes with a vignette that explains the options in detail.