MutDB: Enhanced Biochemical Analysis of Structural and Functional Features of Human Genetic Variation

Kishore K. Kamati
kishore@compbio.iupui.edu

Other Authors: Rakesh Sathyesh, Matthew Mort, Arti Singh, Adebayo Olowoyeye, Jessica Dantzer, Charles Moad, Vidhya G. Krishnan, Peter H. Baenziger, Maricel G. Kann, Predrag Radivojac, Randy Heiland, and Sean D. Mooney

Indiana University School of Medicine
Center for Computational Biology and Bioinformatics

Grants: Supported by NLM K22LM009135 (PI: Mooney) P01AG018397(PI: Econ), a grant from IU Biomedical Research Council, an RSFG grant from IUPUI, Lilly endowment
SNPs Are Important for Genetic Studies

The most commonly known form of genetic variation is a single nucleotide polymorphism (SNP). SNPs are single base substitutions that occur at a frequency of at least 1-10% in a population. Given their frequency and utility, they are of interest to understand how they change molecular function.

Adapted from: http://snp.ims.u-tokyo.ac.jp/samplesMethods.html#SNP
What is MutDB?

- MutDB aims to cover annotation of SNPs/Mutations on gene, transcript and protein level
- Web portal for mutation research community that helps identify the molecular cause of disease

MutDB
(http://mutdb.org)

Databases (UCSC, dbSNP, Swiss-Prot)

Our Annotations (Molecular Features)

Bioinformatics Tools

- MutDB portals for annotation of SNPs mutations help identify the molecular cause of disease
SNPs / Mutations - Gene Level

- MutDB enables visualizing 188 KEGG human pathways
- Genes / Proteins with mutations are highlighted dynamically using SOAP based web services
Protein Data Bank (PDB) - greater than 45,000 structural chains

Annotation pipeline:
- Sequences with mutations are searched against the PDB using the sequence similarity search tool, BLAST
- 100% identical subsequences are then aligned to the sequence of the structure
- Mutation positions are then mapped to 'coordinate space' and saved in a relational database

SNPs / Mutations

SNP Selection Tool

<table>
<thead>
<tr>
<th>Total SNPs</th>
<th>Chromosome</th>
<th>Number of Currently Selected SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>3693</td>
<td>chr1</td>
<td>0</td>
</tr>
</tbody>
</table>

SNP Info

| rs241231   | A/G        | X    | X    | unknown | -  | 4491481 | 0.192132  | 0.243211  | 0.192 0.234898  | 0.249648  | 0.190 0.215 0.22 | 0.011 0.017 0.26 |
| rs12567893 | C/G        | X    | X    | unknown | +  | 4490944 | 0.248988  | 0.249648  | 0.190 0.215 0.22 | 0.011 0.017 0.26 |
| rs12755100 | C/T        | X    | X    | unknown | +  | 4490930 | 0.0198   | 0.0975008 | 0.208 0.388 0.326 |
| rs241232   | C/T        | X    | X    | unknown | -  | 4490828 | 0.428917  | 0.17461  | 0.208 0.388 0.326 |
| rs34995056 | T/A        | X    | X    | unknown | +  | 4490633 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs241233   | A/G        | X    | X    | unknown | -  | 4490609 | 0.491403  | 0.0994683 | 0.292 0.311 0.5 0.2 |
| rs12491112 | T/A        | X    | X    | unknown | +  | 4490320 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs241234   | A/G        | X    | X    | unknown | -  | 4489270 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs34679959 | T/A        | X    | X    | unknown | +  | 4489260 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs34846714 | C/T        | X    | X    | unknown | +  | 4489249 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs34865748 | C/T        | X    | X    | unknown | +  | 4489224 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs241235   | C/T        | X    | X    | unknown | -  | 4489121 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs241236   | C/T        | X    | X    | unknown | -  | 4489023 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs4553444  | X          | X    | X    | unknown | +  | 4489521 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs11807023 | X          | X    | X    | unknown | -  | 4487995 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs241237   | C/T        | X    | X    | unknown | +  | 4487828 | 0.473211 | 0.117426  | 0.457 0.222 0.322 0.475 |
| rs12748762 | X          | X    | X    | unknown | -  | 4487739 | 0.130353 | 0.223565 | 0.190 0.215 0.22 | 0.011 0.017 0.26 |
| rs37517488 | C/T        | X    | X    | unknown | +  | 4487637 | 0       | 0        | 0.292 0.311 0.5 0.2 |
| rs10915531 | C/T        | X    | X    | unknown | +  | 4487454 | 0.072888 | 0.176441 | 0.018 0.045 0.085 |
| rs241238   | A/G        | X    | X    | unknown | +  | 4487378 | 0.347567 | 0.230725 | 0.433 0.045 0.368 |
| rs241239   | A/G        | X    | X    | unknown | +  | 4487036 | 0.178465 | 0.235457 | 0.258 0.0 0.036 |
Future

- We have recently found evidence that bioinformatic methods can predict the molecular mechanism of disease mutations (see poster). We are currently developing a web portal for predicting disease mechanism.