**Short Communication**

The MINTAct Archive for Mutations Influencing Molecular Interactions

Pablo Porras Millán1, Margaret Duesbury1, Maximilian Koch1, Sandra Orchard1, IMEx Consortium Curation Team1

1European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, CB10 1SD, Hinxton, UK

*Correspondence: pporras@ebi.ac.uk

Received 2017-10-13; Accepted 2017-10-27

**SUMMARY**

The MINTAct archive for mutations affecting interaction holds results of over 28,000 events on the influence of a change of protein sequence on physical interaction outcome. All data has been manually curated from experimental evidence found in more than 4,100 publications, following the IMEx consortium ([www.imexconsortium.org](http://www.imexconsortium.org)) high-detail curation standards. The dataset contains data from about 300 different organisms, with a predominance of events related to human proteins, and it is freely available at the IntAct database website using this link: [www.ebi.ac.uk/intact/resources/datasets#mutationDs](http://www.ebi.ac.uk/intact/resources/datasets#mutationDs).

**KEYWORDS**

Variation, Mutagenesis, Protein-protein interactions, Molecular interactions, Data curation, IMEx, Databases

**BODY**

Molecular interaction (MI) networks have become a fundamental resource in systems biology, providing the building blocks for pathway analysis, model inference and protein complex characterization. MI databases have a key role hosting the data that enables scientific work on the study of interactomes. The IntAct database at the European Bioinformatics Institute ([www.ebi.ac.uk/intact](http://www.ebi.ac.uk/intact)) [1] has, as its primary aim, the collation of highly detailed experimental interaction data to make it available for the scientific community. In the last few years, IntAct has also consolidated its role as a common curation platform for most members of the IMEx consortium ([www.imexconsortium.org](http://www.imexconsortium.org)). Its high-detail curation model aims to record every aspect of the experimental setup used to obtain MI evidence. Kinetic parameters, construct details – such as specific mutations that affect the interaction outcome –, and host system are examples of the representation depth provided.

We have undertaken a re-curation effort to further standardize the representation of mutations influencing interaction outcome. IMEx partners have traditionally captured this data, but flexibility in the way sequence replacement should be captured has generated inconsistencies that needed to be addressed. In light of that need, we have also adapted the PSI-MI standard format [2] to its 3.0 version1, enabling new fields that can capture the sequence changes in a consistent manner. Our effort has resulted in what to our knowledge represents the biggest dataset of its type: a high-quality, experimentally derived and manually curated list featuring mutations affecting interactions.

Our mutations dataset currently comprises over 28,000 fully annotated mutations affecting around 14,000 interaction pieces of evidence, with a broad scope of species (~300 organisms represented). About half of the data, however, is related to human proteins (see more detailed statistics on Table 1).

The data is currently available through our standard PSI-MI XML (2.5 and 3.0) and PSI-MITAB data downloads and we also provide a tab-delimited, text file centered on the mutations for ease of use, summarizing the information about each event and referencing the interactions affected. This file contains full details about the sequence changes in the affected protein and the effect it has in the interaction, using the PSI-MI controlled vocabularies (see [http://purl.obolibrary.org/obo/MI_0118](http://purl.obolibrary.org/obo/MI_0118) for more details). Additionally, each record contains information about the affected protein and its interacting partners, the source publication and figure legend where the evidence is provided, along with the interaction accession to enable search in IntAct if further details are needed.

At the moment, we are working to enrich the dataset by mapping the mutations to known variants in the genome and to phenotypic annotations linked to such variants. Initial work on this direction has shown that about 60% of the mutations described in human proteins can be fully or partially mapped to known variants.

The dataset and further documentation can be freely accessed at [www.ebi.ac.uk/intact/resources/datasets#mutationDs](http://www.ebi.ac.uk/intact/resources/datasets#mutationDs). We believe this free, openly available resource provides the scientific community with an invaluable tool to study the impact that specific protein variation has on the interactome as a whole. Our mutations dataset has immediate applications in the study of interaction interfaces, the impact of variation in the interactome and the characterization of previously un-annotated variants and their relation with disease, among other key questions.
<table>
<thead>
<tr>
<th>Organism</th>
<th>Annotated features</th>
<th>Sequence changes</th>
<th>Affected proteins</th>
<th>Affected interactions</th>
<th>Source publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>16,861</td>
<td>7,955</td>
<td>2,095</td>
<td>8,268</td>
<td>2,219</td>
</tr>
<tr>
<td>mouse</td>
<td>2,236</td>
<td>1,406</td>
<td>482</td>
<td>1,248</td>
<td>509</td>
</tr>
<tr>
<td>yeast</td>
<td>2,029</td>
<td>1,144</td>
<td>363</td>
<td>1,069</td>
<td>326</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>1,172</td>
<td>546</td>
<td>187</td>
<td>590</td>
<td>189</td>
</tr>
<tr>
<td>E. coli</td>
<td>979</td>
<td>614</td>
<td>143</td>
<td>374</td>
<td>148</td>
</tr>
<tr>
<td>rat</td>
<td>562</td>
<td>354</td>
<td>142</td>
<td>341</td>
<td>160</td>
</tr>
<tr>
<td>fly</td>
<td>359</td>
<td>232</td>
<td>100</td>
<td>214</td>
<td>92</td>
</tr>
<tr>
<td>Others</td>
<td>3,670</td>
<td>2,396</td>
<td>841</td>
<td>1,951</td>
<td>855</td>
</tr>
<tr>
<td>Total</td>
<td>27,688</td>
<td>14,647</td>
<td>4,353</td>
<td>13,926</td>
<td>4,182</td>
</tr>
</tbody>
</table>

Table 1: Summary statistics for the mutations dataset.

ACKNOWLEDGEMENTS

Supported by the Biotechnology and Biological Sciences Research Council under grant agreement BBSRC MIDAS grant (BB/L024179/1).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES


NOTES