**ABSTRACT**

DisGeNET [1] is a discovery platform designed to answer questions concerning the molecular mechanisms underlying human diseases. DisGeNET data can be explored using a suite of tools which includes a web interface, a CytoScape plugin [2], and a SPARQL endpoint [3]. In this contribution, we present disgenetR: an R package for exploring DisGeNET. disgenetR contains a variety of functions for leveraging DisGeNET using the powerful visualization and statistical capabilities of the R environment. disgenetR is specially designed to harness the large amount of information contained in DisGeNET, facilitating its analysis and interpretation. The package offers different types of visualization of DisGeNET data, such as heatmaps and networks, and it is especially well suited to explore the genetic basis of diseases as well as disease comorbidity. Furthermore, to allow answering more sophisticated research questions that need the interrogation of multiple, heterogeneous and disparate resources, the disgenetR package permits benefiting of the potential of the Semantic Web technologies, without the need of special expertise in this area. This is achieved through a set of functions that connect DisGeNET with other resources present in the Linked Open Data, covering different information such as gene expression, drug activity, and biological pathways, just to mention a few examples. The disgenetR package also expedites the integration of DisGeNET data with other R/Bioconductor packages, and allows the construction of complex bioinformatic workflows. We illustrate the functionality of disgenetR through several use cases to show how the package can be applied to aid particular user’s needs. The source code and documentation of disgenetR package are available at https://bitbucket.org/albag/disgenetR.

---

**THE DISGENET DATABASE**

**THE GENES**

**THE DISEASES**

**THE VARIANTS**

**LINKING TO OTHER RESOURCES**

---

**Summary of DisGeNET statistics**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Genes</th>
<th>Diseases/Phenotypes</th>
<th>SNPs</th>
<th>GDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURATED</td>
<td>3203</td>
<td>7697</td>
<td>35094</td>
<td>32834</td>
</tr>
<tr>
<td>PREDICTED</td>
<td>2743</td>
<td>2064</td>
<td>40154</td>
<td>40254</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td>1567</td>
<td>9661</td>
<td>45148</td>
<td>46111</td>
</tr>
</tbody>
</table>

The DisGENET score

\[ S = \sum_{i=1}^{C} C_i + \sum_{j=1}^{M} M_j + \sum_{k=1}^{L} L_k \]

\( C_i \) - Curated Sources (UniProt, ClinVar, Orphanet, the GWAS Catalog, CTD)

\( M_j \) - Predicted Sources (MSD, RGD, CTD(animal models))

\( L_r \) - Text Mining Data (GAD, LGHDN, BefRaw)

---

**REFERENCES**

