**Elucidating drug-drug interactions underlying drug polypharmacy profiles**

**Jorge Hernansanz**1**, Cristina Leal**1**, Gianluca Mazzoni**1**, Søren Brunak**1

1 Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark.

Corresponding author: jorge.hernansanz@cpr.ku.dk

**ABSTRACT**

**Background:** Drug-drug interactions (DDI) are a common characteristic of adverse drug reactions (ADR) and suppose an emerging threat to public health. DDI data remains disperse and its integration into a unique source can benefit other research for an easier access to data.

**Methods:** We generated a complete compendium of DDI data through the integration of **xx** publicly available drug sources. Then, we characterize each source by looking to its drug and DDI distribution. Finally, we performed a network / cluster analysis in order to find significant DDI clusters in a drug combination clinical data

**Results:** Across our databases, there are **X** unique drug/chemical names and **X** unique DDIs. We found 5 different DDI features; clinical significance, management options, evidence level, mechanism of action and side effect, that can be factorized and be used in posterior exploratory analysis. It was observed in non-general databases a bias for some of the ATC class. Finally, it was found two significant clusters from the network analysis performed on the drug combination clinical data.

**Conclusions:** There is little overlap between the different databases, taking into account that some of them are specific for a certain drug class. Network analysis using clinical significance feature proved that DDI information can be valuable to extract significant information from clinical data.

**Keywords:** Drug-drug interactions, WHO-ATC identifiers, Network analysis, DDI characterization

**BACKGROUND Section title**

* *What are DDI and its role in pharmacovigilance*
* *Methods for obtaining DDIs, developing the importance of text-mining or NLP techniques in clinical text abstracts.*
* *The problem with drug identifiers; the challenge of normalizing DDIs to a single identifier.*
* *DDI information; explaining the different features that it can be found at the public databases.*
* *Give highlights of general findings (Of my data) related to the expected overlap of DDI or common features so as it can serve as an introduction for the section of Results.*

A DDI occurs when one drug modifies the pharmacological activity of other drug. It is a common characteristic of adverse drug reactions (ADR) [1], which is the focus of pharmacovigilance [2]. DDIs can be divided into two major groups, pharmacokinetic interactions, which involves changes in the absorption, distribution, metabolism and excretion of the drugs [3], and pharmacodynamics interactions; in which directly one drug changes the pharmacological effect of another one being additive or antagonist [4].

DDI data can be found in both public website sources (Drugbank, Crediblemeds…) and private commercial sources (Lexicomp, Stockley’s drug interactions…). Current DDI can be used to train statistical models that predicts new DDIs. Some DDI extraction challenges were hold in the past with the goal of covering different approaches when extracting DDI data from biomedical texts [5] [6]. Also, other project evaluated the performance of unsupervised and supervised machine learning methods for predicting potential DDIs [7].

Another challenge DDIs face is its integration and common consensus. A previous study already tried to integrate 14 databases with DDI information mapped to Drugbank identifiers [8]. Moreover, other research focusing on the overlapping of specific DDI databases has been done [9]. In a society where co-administration of drugs is becoming more and more frequent, a synthesis of this knowledge would help clinicians guide patients to effective and less noxious medication therapies [10].

In this paper, our objective is to extract all DDI data from publicly available databases, and gather all of them by mapping each drug to a common identifier; the ATC code. Having a common dataset, we will characterize the drug pairs and will factorize its DDI information so that it can be used in future exploratory analysis. Finally, we filtered our compendium to a clinical data generated by the laboratory and tried to identify significant patterns across it. Its importance in the pharmacovigilance field makes trivial to unify all this data in a unique compendium, normalized to a golden standard identifier.

**MATERIALS AND METHODS Section title**

* Brief description of the compendia we are creating and its purpose
* *Mentioning as a reference for DDI extraction the Github project mentioned in [3]*
* *Explaining the DDI resources*
* *Parsing DDIs to ATC identifiers; explaining the different methodologies used without giving numbers. Mentioning the rest APIs, web scraping…*
* *Factorizing the different features; explaining the different methodologies without giving numbers.*

**ANATOMICAL THRAPEUTICAL CLASSIFICATION Sub heading**

Anatomical Therapeutic Chemical (ATC) Classification System is a drug-related index that classifies within different levels the chemical group of each drug. Drug may have several ATC codes but there is only one drug for each ATC code. It consists of 14 different anatomical groups, each of them reaching 5 different levels of differentiation.

Some databases use ATC codes alongside with their own identifier, such as Drugbank or KEGG. Nevertheless, the ATC classification is developed and maintained by the World Health Organization Collaborating Center (WHOCC) for drug statistics methodology.

**OVERVIEW OF DDI SOURCES Sub heading**

We created a drug- drug interaction dataset using content from 15 openly available database resources. Many of them are continuously updating, but others are just fixed projects just one time updated.

The method each database was extracted is mentioned below. Several kind of sources were met such as regular XML format files, REST APIs, or information on website that was downloaded directly from there.

Each database contains specific information / features related to DDI pairs (Table 2). From the 15 databases, we recollected five kinds of DDI features. *Clinical significance* feature associates to the DDI the level of change in the physiological effect of the drugs affected. *Evidence level* gives an estimation of how well documented is that DDI. *Management options* refers to how to proceed with the administration of the drug combination. *Mechanism of action* reflects the target/s involved due to the DDI. *Side effect* states the adverse reaction produced by the DDI.

Clinical significance, evidence level, and management features were transformed to digits in order to characterize in a numerical scale the magnitude of that information for each DDI, which could be used in posterior exploratory analyses.

Mechanism of action and Side effect features were also treated to be used in exploratory analyses. Mechanism of action output was restricted / formatted to a general vocabulary of actions easy to deal with. Side effect output was filtered to the ones that had Meddra identifiers.

**DRUGBANK Sub sub heading**

The DRUGBANK database [11] is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information. Information retrieval was performed with the R package “DBparsed” [12], obtaining a brief description for each DDI. The version used in this study (5.0) was downloaded from the Drugbank website on January 2020.

**KEGG DRUG Sub sub heading**

The KEGG DRUG database [13] is a comprehensive drug information resource for approved drugs in Japan, USA, and Europe. Information is unified based on the chemical structure and/or the chemical component of active ingredients. Information retrieval was performed with its REST API [14], such as drugs’ mechanism of action and management DDI (“Contraindicated”, “precaution”). The version used in this study (93.0) was downloaded from the KEGG API in February 2020.

**TWOSIDES ….**

The TWOSIDES project [15] consists of a comprehensive database of DDI side effects elaborated by Tatonetti’s laboratory. Twosides reports a total of 40 million DDI-related side effects, with their corresponding propensity scores for the evidence level. We downloaded the dataset from their lab repository in February 2020.

**NDF-RT …..**

NDF-RT [16] is the National Drug File – Reference Terminology from the U.S. Department of Veteran Affairs, Veterans Health Administration (VHA). It is an extension of the VHA National Drug File (NDF) that combines its hierarchical drug classification with a multi-category reference model. “VA Drug interactions” was the category extracted, which contains clinical significance concept for each of the DDIs. The version used in this study dates from July 2014 as this was the last release VA was maintaining NDF-RT interactions, resulting in its removal from their posterior updates. Data was downloaded from the National Institutes of Health (NIH) repository in March 2020.

**CREDIBLEMEDS ….**

CREDIBLEMEDS [17] is a clinically oriented information source that is used to guide clinical decision-making and safe use records drugs with risk of QT prolongation and/or torsades de pointes (TdP). Crediblemeds comprises a small DDI dataset and contains information such as Precipitant and Object drugs, as well as mechanism of action and management options features for each of the DDIs. The newest version of it dates from May 2010 with no identifiers, and it was extracted from [] as they already had parsed the drugs to their Drugbank identifiers.

**INTERAKTION DATABASEN ….**

The Danish DDI database [18] is an electronic search tool that describes evidence-based interactions documented by clinical/case studies. It provides a common frame of reference for the on the handling of drug interactions in the Danish healthcare system. Data was retrieved from a public document provided by the Danish Medicines Agency. Information provided includes clinical significance, evidence level, and management administration features for each DDI. The version used in this study (XML\_dato\_3 release 2.7) was downloaded in March 2020.

**ONC HIGH-PRIORITY / NON-INTERRUMPTIVE ….**

ONC HIGH-PRIORITY (ONC-HP) [19] consists of a set of high-severity DDIs for use in electronic health records (EHR). This dataset is characterized by contraindicated and highly clinical significant DDIs. Nevertheless, most of the DDIs included in this resource lack of primary literature supporting their evidence. We retrieve information regarding clinical significance, evidence level, and management options features. It is a one-time updated project published in April 2012.

ONC NON-INTERRUMPTIVE (ONC-NI) [20] consists of a set of low priority DDIs feasible for non-interruptive alerts that aims to reduce alert fatigue for the provider’s workflow of EHRs. We retrieved information regarding the clinical significance and evidence level features. Like ONC HIGH-PRIORITY, this is also a one-time updated project published in September 2012.

These two projects comes from research organized by the Office of The National Coordinator for Health Information Technology (ONC) where in both cases the set of DDIs was a consensus between the different commercial drug providers that participated. The projects were already treated by [8], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**DDI CORPUS 2011 / 2013 ….**

DDI CORPUS 2011 [5] relates to a project included in the DDI Extraction 2011 workshop focused on natural language processing (NLP) techniques for drug-drug extraction from text selected from the Drugbank database. This project dates from September 2011

DDI CORPUS 2013 [6] relates to a project included in the SemEval 2013 DDI Extraction challenge for the evaluation of NLP techniques applied to recognition of pharmacological substances and drug-drug extraction from Drugbank and Medline databases. This project dates from July 2013.

Isabel Segura’s Lab carried out both projects and its raw data can be accessed from her own GitHub. Those projects were already treated by [8], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**NLM CV CORPUS ….**

NLM CV CORPUS [21] is a DDI corpus used for training NLP techniques that identifies possible DDIs where cardiovascular drugs are involved. This project was already treated by [8], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**PK CORPUS ….**

PK CORPUS [22] is a specific-pharmacokinetic DDI corpus from a pharmacokinetic ontology used for DDI text mining analysis in drug product labels. This project was already treated by [8], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**HIV / HEP / CANCER DRUG INTERACTIONS ….**

HIV / HEP /CANCER DRUG INTERACTIONS [23] [24] [25] correspond to three DDI resources from the Liverpool university. They report the clinical significance between a selected small amount of disease-specific drugs and a bigger set of standardized drugs, independently of if there is interaction or not. We extracted information related to the clinical significance and the evidence level for each DDI. The three of them were downloaded in April 2020.

**PARSING DDI PAIRS Sub heading section**

Here we describe a summary of the mapping procedure and its results for the different databases parsed. Table 1 gives statistics about the unique number of drugs and DDI pairs, as well as the owner of the source and the version / update used when retrieved.

DrugBank and KEGG names their drugs with their own type of identifier, and provide a complementary section of data that links their identifier with the ATC code. This complementary section of DrugBank was also used for the four corpus, Crediblemeds dataset, and the ONC projects, that were already encoded with Drugbank identifiers.

NDF-RT uses its own type of identifier, but does not provide in a direct way a link of its drugs with the ATC code. For it, we used a comprehensive repository of biomedical ontologies called BioPortal [26] that provided us for indexes to parse the NDF-RT identifier to the Concept Unique Identifier (CUI), and from CUI to ATC code.

Interaktion Databasen encodes its drugs and every other component with its own identifiers. For the mapping procedure it was used a CSV file retrieved from the Danish Medicine Agency that accounts for the ATC code for each drug identifier stored in the database.

TWOSIDES uses the RXNORM identifier for its drugs, which is the normalized clinical drug dictionary of the Unified Medical Language System (UMLS). As it didn’t provided us with a direct link to the ATC code, we retrieve form BioPortal an index to parse RXnorm to CUI identifier, and used the previous index retrieved of CUI – ATC identifier.

HIV, HEP and CANCER interactions are just online websites where it is only registered the regular drug name. Therefore, drugs were name-linked to Drugbank identifiers so as to be linked to the ATC code (From the index obtained from Drugbank), and those that were not mapped to DrugBank ones were manually annotated their ATC code using as index the WHO ATC list of 2017.

SPACE FOR THE DESCRIPTION OF CLINICAL DATA. TO BE ADDED BY THE SUPERVISORS.

**NETWORK ANALYSIS Sub section …**

Cytoscape [27] is an open source software platform used for visualizing and integrating networks. It provides features called “apps” used for network analysis. For our project, we used “MCODE” [25] app to identify clusters in our DDI network. It is a cluster algorithm that detects densely connected regions in large interaction networks. Although it is mainly used for protein-protein interactions, it can be used for DDI in order to identify significant drug clusters according to a specific DDI feature.

**RESULTS Section**

**CHARACTERISING DDIS Sub section**

An overlapping analysis was performed involving the 15 databases parsed. It was used an R package called “UpsetR” [] for the visualization of the intersecting sources, ranging from DDIs that are unique for a database to the DDIs that can be found in most of them (Additional file 1). We found our DDI data to have a significant limited overlap where we can observe there is a enormous quantity of DDIs that just appear in one, two, or three databases, whereas very few DDIs appear in higher degree intersections (Figure 1 a, b). The highest degree of overlapping achieved by our compendia is of nine databases with only five DDIs in that section.

An analysis for the distribution of DDIs was run inside each database. For it, we categorized drugs according to their first ATC level and checked, both, the quantity of each kind of drug (Figure 2) and its weight on the interactions (Figure 3) for each database.

Attending the distribution of drugs, we observe that the databases follow a similar one where the level 1 ATC class most relevant corresponds to the Cardiovascular System drugs. Also, Nervous System drugs and Antiinfectives For Systemic Use drugs have a major overall presence than most of classes. Some databases account for a more personalized distribution of drugs where we observe the NLM CV Corpus and ONC NI with a bigger proportion of the Cardiovascular System class (Expected from the first one as it targets drug labels related to cardiovascular drugs), or Crediblemeds with Musculo-skeletal System ATC class compared to the rest of databases.

Looking at the distribution of interactions inside each database, we observe again that the most predominant ATC class correspond to the one of Cardiovascular System. Both NLM CV Corpus and ONC-NI display a big set of ATC class C receiving most of interactions from the rest of classes. It is significant to observe Crediblemeds that most drugs from ATC class J are interacting with the class C (The use of antineoplastic drugs can lead to cardiotoxicity [], provoking the more frequent co-administration of this two kind of drugs) and that ATC class M, although having a good percentage of drugs inside the database, plays an insignificant role at the interactions. HIV, HEP and CANCER databases consists of 41, 22, and 70 respective related-drugs interacting with hundreds of diverse drugs, explaining why most of interactions for each one are dominated by one ATC class which had a normal weight in the distribution of figure 2. Therefore, we observe that drugs

**CLUSTER ANALYSIS Sub section**

We aimed to represent in a network how DDIs are distributed according to clinical significance feature (Only using DDIs that contain this data), and find any significant cluster of highly interacting drugs. DDIs were filtered to a drug-combination clinical dataset provided by Brunak lab, allowing the analysis to have a real implementation in a clinical environment. Drugs were enriched to their first level ATC code and. As observed in additional table 2, Clinical significance feature ranks interactions in four levels (0,1,2,3) being 0 a DDI that it is not expected to present pharmacokinetic changes and 3 a DDI with a mayor change in the pharmacokinetics one of the drugs. For the analysis, DDIs with a score of zero in clinical significance feature were removed to reduce noise in the network.

After this preprocessing of data, we execute MCODE app to produce the clusters. Cluster analysis tuning involved setting parameters. “Degree cutoff” was set to two in order to prevent single-connected nodes in clusters. As our clinical data is directed and therefore two drugs may share two edges, “K-Core” parameter was set to three to force each node to have three connections per node and to connect at least with two drugs. Following the FAQ section of MCODE [28] it is stated to apply the “MCODE level 3” parameters in case of seeking for dense interconnected clusters. This changed parameter “Node Score Cutoff” to 0.1, “fluff” to zero, and disabling “haircut” option.

MCODE returned two clusters (Figure 4). The first one consisted of ten nodes and ninety edges all of them with a clinical significance level 3. The cluster reveals that this ten drugs (With a predominance of N ATC class) are very likely to interact between them and alter in a significant way their pharmacokinetic metabolism. The second one is less densed but still targeted by MCODE. It is more heterogeneous in terms of clinical significance levels. We can observe from this cluster that a great majority of interactions is set in level two, and that ATC level N remains again as the class that groups more nodes in the cluster.

**DISCUSSION**

**CONCLUSION Section**

In this project, we integrate a wide range of publicly available sources of DDI information. We analyzed the overlap across the different databases and we found that there is actually little overlap between DDI pairs and there is heterogeneity between the features extracted from the sources. It was curious to see that Drugbank, KEGG, and Twosides appear in nearly all intersections for the most overlapped DDIs, justifying that those databases also store the most important ones.

Then, we looked to the distribution of the interactions inside the databases and checked for significant patterns. It was an interesting finding to see how a small percentage of drugs can be the dominant ATC class in a source, such as in the case of the Liverpool interactions, or that the non-specific DDI sources such as Drugbank, KEGG and Twosides present a very diverse distribution of interactions between classes.

Finally, we perform a cluster analysis where we filter our compendia to DDIs that had clinical significance information, and that were located into a drug-combination clinical dataset. Its analysis a revealed a highly dense cluster of 10 drugs that were to interact between each of them.

As future work, this project may continue in the process of extracting more DDI databases such as SIDER, SUPERCYP… Moreover, the characterization of the different DDI features will serve as future exploratory analysis for clinical data as it was for the one of our project.

**REFERENCES**

1. Montastruc F, Sommet A, Bondon-Guitton E, et al. The importance of drug-drug interactions as a cause of adverse drug reactions: a pharmacovigilance study of serotoninergic reuptake inhibitors in France. European Journal of Clinical Pharmacology. 2012 May;68(5):767-775. DOI: 10.1007/s00228-011-1156-7.
2. Lopes P, Nunes T, Campos D, Furlong LI, Bauer-Mehren A, Sanz F, et al. (2013) Gathering and Exploring Scientific Knowledge in Pharmacovigilance. PLoS ONE 8(12): e83016. <https://doi.org/10.1371/journal.pone.0083016>
3. Palleria C, Di Paolo A, Giofrè C, et al. Pharmacokinetic drug-drug interaction and their implication in clinical management. *J Res Med Sci*. 2013;18(7):601‐610.
4. Niu J, Straubinger RM, Mager DE. Pharmacodynamic Drug-Drug Interactions. *Clin Pharmacol Ther*. 2019;105(6):1395‐1406. doi:10.1002/cpt.1434
5. Segura-bedmar, I. (2011). Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction. *Proceedings of the 1st Challenge Task on Drug-Drug Interaction Extraction*, *January*.
6. Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., & Declerck, T. (2013). The DDI corpus: An annotated corpus with pharmacological substances and drug-drug interactions. *Journal of Biomedical Informatics*, *46*(5), 914–920. <https://doi.org/10.1016/j.jbi.2013.07.011>
7. Kastrin, A., Ferk, P., & Leskošek, B. (2018). Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PLoS ONE*, *13*(5), 1–23. <https://doi.org/10.1371/journal.pone.0196865>
8. Ayvaz, S., Horn, J., Hassanzadeh, O., Zhu, Q., Stan, J., Tatonetti, N. P., Vilar, S., Brochhausen, M., Samwald, M., Rastegar-Mojarad, M., Dumontier, M., & Boyce, R. D. (2015). Toward a complete dataset of drug-drug interaction information from publicly available sources. *Journal of Biomedical Informatics*, *55*(May), 206–217. <https://doi.org/10.1016/j.jbi.2015.04.006>
9. Peters, L. B., Bahr, N., & Bodenreider, O. (2015). Evaluating drug-drug interaction information in NDF-RT and DrugBank. *Journal of Biomedical Semantics*, *6*(1). https://doi.org/10.1186/s13326-015-0018-0
10. Hines, L.E., Malone, D.C. and Murphy, J.E. (2012), Recommendations for Generating, Evaluating, and Implementing Drug‐Drug Interaction Evidence. Pharmacotherapy, 32: 304-313. doi:[10.1002/j.1875-9114.2012.01024.x](https://doi.org/10.1002/j.1875-9114.2012.01024.x)
11. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, Di., … Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
12. Mohammed Ali, Ali Ezzat (). dbparser: DrugBank Database XML Parser. R package version 1.1.3.9000.
13. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2009). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research*, *38*(SUPPL.1), 355–360. https://doi.org/10.1093/nar/gkp896
14. <https://www.kegg.jp/kegg/rest/keggapi.html>
15. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med*. 2012;4(125):125ra31. doi:10.1126/scitranslmed.3003377
16. Olvey, E.L., Clauschee, S. and Malone, D.C. (2010), Comparison of Critical Drug–Drug Interaction Listings: The Department of Veterans Affairs Medical System and Standard Reference Compendia. Clinical Pharmacology & Therapeutics, 87: 48-51. doi:10.1038/clpt.2009.198
17. <https://www.crediblemeds.org/healthcare-providers/drug-drug-interaction>
18. <https://laegemiddelstyrelsen.dk/en/sideeffects/danish-drug-interaction-databases/>
19. Phansalkar, S., Desai, A. A., Bell, D., Yoshida, E., Doole, J., Czochanski, M., Middleton, B., & Bates, D. W. (2012). High-priority drugedrug interactions for use in electronic health records. *Journal of the American Medical Informatics Association*, *19*(5), 735–743. https://doi.org/10.1136/amiajnl-2011-000612
20. Phansalkar, S., van der Sijs, H., Tucker, A. D., Desai, A. A., Bell, D. S., Teich, J. M., Middleton, B., & Bates, D. W. (2013). Drug-drug interactions that should be noninterruptive in order to reduce alert fatigue in electronic health records. *Journal of the American Medical Informatics Association*, *20*(3), 489–493. https://doi.org/10.1136/amiajnl-2012-001089
21. Johann Stan, A Machine-Learning Approach for Drug–Drug Interaction Extraction from FDA Structured Product Labels, Presented at the 2014 National Library of Medicine Training Conference, Pittsburgh PA, USA, 17-Jun-2014.
22. Wu, H., Karnik, S., Subhadarshini, A. *et al.* An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinformatics* **14,**35 (2013). https://doi.org/10.1186/1471-2105-14-35
23. <https://www.hiv-druginteractions.org/checker>
24. <https://www.hep-druginteractions.org/>
25. <https://cancer-druginteractions.org/>
26. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, Smith B; NCBO team. The National Center for Biomedical Ontology. J Am Med Inform Assoc. 2012 Mar-Apr;19(2):190-5. Epub 2011 Nov 10.
27. Paul Shannon, 1, Andrew Markiel, 1, Owen Ozier, 2 Nitin S. Baliga, 1 Jonathan T. Wang, 2 Daniel Ramage, 2, Nada Amin, 2, Benno Schwikowski, 1, 5 and Trey Ideker2, 3, 4, 5 (1971). Cytoscape: A Software Environment for Integrated Models. *Genome Research*, *13*(22), 426. https://doi.org/10.1101/gr.1239303.metabolite
28. Bader, G.D., Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,**2 (2003). <https://doi.org/10.1186/1471-2105-4-2>