GUILDify v2.0

Aguirre-Plans, Joaquim; Piñero, Janet; Sanz, Ferran; Furlong, Laura I.; Fernandez-Fuentes, Narcis; Oliva, Baldo; Guney, Emre

Published in:
Journal of Molecular Biology

DOI:
10.1016/j.jmb.2019.02.027

Publication date:
2019

Citation for published version (APA):
Aguirre-Plans, J., Piñero, J., Sanz, F., Furlong, L. I., Fernandez-Fuentes, N., Oliva, B., & Guney, E. (2019). GUILDify v2.0: A Tool to Identify Molecular Networks Underlying Human Diseases, Their Comorbidities and Their Druggable Targets. *Journal of Molecular Biology*, *431*(13), 2477-2484. https://doi.org/10.1016/j.jmb.2019.02.027

GUILDify v2.0: A tool to identify molecular networks underlying human diseases, their comorbidities and their druggable targets

Joaquim Aguirre-Plans[1], Janet Piñero[2], Ferran Sanz[2], Laura I. Furlong[2], Narcis Fernandez-Fuentes[4], Baldo Oliva[1,*] and Emre Guney[2,5,*]

1- Structural Bioinformatics Group, Research Programme on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia 08003, Spain
2- Integrative Biomedical Informatics Group, Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia 08003, Spain
3- Department of Biosciences, U Science Tech, Universitat de Vic-Universitat Central de Catalunya, Vic, Catalonia 08500, Spain
4- Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, SY23 3EB Aberystwyth, United Kingdom
5- Department of Pharmacology and Personalised Medicine, CARIM, FHML, Maastricht University, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands

***Correspondence to Emre Guney**: emre.guney@upf.edu
***Correspondence may also be addressed to Baldo Oliva**: baldo.oliva@upf.edu

Abstract

The genetic basis of complex diseases involves alterations on multiple genes. Unravelling the interplay between these genetic factors is key to the discovery of new biomarkers and treatments. In 2014, we introduced GUILDify, a web server that searches for genes associated to diseases, finds novel disease-genes applying various network-based prioritisation algorithms and proposes candidate drugs. Here, we present GUILDify v2.0, a major update and improvement of the original method, where we have included protein interaction data for seven species and 22 human tissues and incorporated the disease-gene associations from DisGeNET. To infer potential disease relationships associated with multi-morbidities, we introduced a novel feature for estimating the genetic and functional overlap of two diseases using the top-ranking genes and the associated enrichment of biological functions and pathways (as defined by GO and Reactome). The analysis of this overlap helps to identify the mechanistic role of genes and protein-protein interactions in comorbidities. Finally, we provided an R package, guildifyR, to facilitate programmatic access to GUILDify v2.0 (http://sbi.upf.edu/guildify2)

Introduction

Complex diseases such as cancer, diabetes, neurodegenerative disorders or cardiovascular diseases are rarely caused by a single genetic perturbation and usually involve polygenic modifications on the underlying interconnected cellular network. Understanding the genetic basis of diseases and the interactions of disease-associated proteins in the protein interaction network (PIN) is essential for the

development of new rational therapeutic strategies. Despite recent large-scale genotyping efforts, information on disease-gene associations is still limited, often explaining a small percentage of the phenotypic variance observed among individuals [1]. To address this limitation and infer novel disease-gene associations, various disease-gene prioritisation methods have been suggested, exploiting the “guilt-by-association” principle over certain features of disease-genes such as similarity in sequence and functional annotations, clustering in the linkage interval, or proximity in the PIN [2]. Indeed, albeit the PINs being incomplete [3], the proximity to disease-genes in the PIN has proven extremely useful in prioritising disease-associated genes [4]. Consequently, a number of tools and web servers has been developed to expand the number of disease-associated genes using the interactome [5–9].

Previously, we presented GUILDify, a web server that applies the prioritisation algorithms developed in GUILD software to find novel disease-gene associations based on the connectedness of genes in the PIN [10,11]. GUILDify searches for genes starting from user-provided keywords such as the names of diseases or gene symbols in the BIANA knowledge database. It uses the genes associated to the keywords as seeds and the PIN for the selected organism to apply graph theory algorithms to prioritise new disease genes. Recently, GUILDify has been applied to: (i) find comorbidities across genetic diseases [12]; (ii) construct PINs specific to breast cancer metastasis to lung and brain [13]; (iii) identify candidate genes for body size in sheep [14] and (iv) prioritise preeclampsia pathogenesis [15].

Here, we present a comprehensive upgrade, GUILDify v2.0, where we updated the underlying biological databases in BIANA knowledge database (protein and drug-target interactions, functional and disease annotations) and: (i) facilitated the use of seven species-specific PINs and 22 human tissue-specific PINs; (ii) increased the quality and number of disease-gene associations by incorporating DisGeNET to our datasets; (iii) incorporated the option to search by drug name, allowing the prioritisation of genes based on known drug targets to uncover the neighbourhood of the PIN affected by the drug; (iv) improved the visualisation of the results using cytoscape.js; (v) refined the definition of top-ranking genes based on whether they had similar functional annotations as the seeds, thus providing the biologically most coherent subnetwork relevant to a given disease; (vi) introduced a feature to measure the genetic and functional overlap of the top-ranking genes of two different diseases, supporting the investigation of disease comorbidities; (vii) implemented a new drug repurposing functionality to propose novel indications for a given drug based on the genetic and functional overlap; and (viii) developed an R package to facilitate the programmatic access to the methods implemented in the web server.

Results and Discussion

Advances

1. Identifying genetic and functional similarities across diseases

In recent works, we have shown that the genetic and functional similarities of diseases in the PIN can

be used to characterise co- and multi-morbidities across diseases [12] and also to repurpose existing drugs targeting these diseases [16]. Motivated by these findings and to provide systematic insights on disease-disease relationships, GUILDify v2.0 now allows users to identify the overlap between two previously submitted results, i.e. sets of genes linked to two different diseases. Accordingly, given two job IDs corresponding to the prioritisation results of two different diseases, GUILDify v2.0 provides: (i) the overlap between the top-ranking genes of the two diseases; (ii) the overlap between the enriched functions among the top-ranking genes of the two diseases; (iii) the enriched functions among the common top-ranking genes; and (iv) a network visualisation of the interactions between common top-ranking genes. Moreover, GUILDify v2.0 also calculates the Fisher's exact test to quantify the significance of the overlap between genes and functions and report one-sided P-value (see details in Supplementary Material). GUILDify v2.0 is the first server that permits the use of gene prioritisation results to explore disease-disease relationships with such simplicity and flexibility.

2. Prioritisation of drug targets

GUILDify v2.0 now allows to search by a drug in addition to a phenotype and returns a list of drug-target associations integrated from DrugBank [17], DGIdb [18], DrugCentral [19] and ChEMBL [20] (see details in Supplementary Material). This new functionality allows the characterisation of the neighbourhood of the drug in the PIN, i.e. neighbouring proteins to those targeted by the drug, and thus providing insights on the potential mechanism of action of the drug. Moreover, the novel feature of assessing the overlap between two network expansion runs (i.e. two job IDs) can also be applied in multiple scenarios to: (i) identify the similarity between the neighbourhood of two drugs in the PIN, which can be useful to identify drug interactions; (ii) compare the neighbourhood of a disease with the neighbourhood of a drug in the PIN, which can be applied to drug repurposing. Such novel features make GUILDify v2.0 one of the most easy-to-use and flexible web servers to inspect the effect of drugs in the PIN.

3. Screening diseases to identify potential new indications of known drugs

Building upon new technical developments mentioned above, GUILDify v2.0 now offers a novel drug repurposing functionality. Given a job ID associated with a drug (or a list of drug targets), this feature automatically calculates the overlap of genes (or functions) between the given drug and a set of pre-calculated diseases. Details on the method and validation of drug repurposing are described in detail at Supplementary Material.

4. Tissue and species-specific PINs

The analysis of the protein interactions in a tissue-specific context is becoming increasingly relevant to understand genetic diseases and find improved treatments [21]. We have included tissue-specific networks derived from 22 different human tissues (see Supplementary Table S1). To create these networks, we filtered the interactions in the global PIN using RNAseq data from GTEx [22], keeping only the interactions between proteins encoded by genes that are expressed in a given tissue (i.e. considering only transcripts with TPM (transcripts per kilobase million) expression values of 1 or

higher (see details in Supplementary Material). We have also included 7 species-specific PINs derived from experimentally determined protein-protein interactions. Although the coverage of interactomic data for some species is low (e.g., 11,943 interactions in rat vs 320,337 interactions in human), these PINs provide a reliable backbone for interactome-based analyses (e.g., in preclinical research) as opposed to PINs generated by predicted interactions based on homology information.

5. Disease-gene information from DisGeNET

We incorporated DisGeNET, one of the largest repositories of genes and variants associated to human diseases [23]. DisGeNET relies on data from UniProt [24], CTD [25], CLINVAR [26], ORPHANET [27], GWAS Catalog [28], PsyGeNET [29] and HPO [30] and is integrated in BIANA [31]. To investigate the increase in the number of disease-gene associations between versions 1 and 2 of GUILDify, we checked the number of associations for the lowest-level non-obsolete diseases from Disease Ontology [32] that were available in our repositories (2,190 terms). GUILDify v1 contains gene associations for 1,505 diseases and 4,171 genes (2.8 genes per disease), while updated GUILDify v2.0 has gene associations for 2,064 diseases and 11,615 genes (5.6 genes per disease on average).

6. Functional-coherency based selection of top-ranking genes

One of the main issues when working with disease-gene prioritisation is to select the most relevant (top ranked) genes associated with a given disease. The user can select top 1% or 2% highest scoring genes among all the proteins in the PIN as top ranked genes. In GUILDify v2.0, we also introduced a cutoff based on the functional validation approach described in Ghiassian *et al.* [5] and provided a new panel visualising the significance of the functional enrichment (P-value) as a function of the number of top-ranking genes included in the validation (implemented in Plotly). In brief, the highest-scoring non-seed proteins are iteratively included in the top-ranking set, provided that they maintain the functional coherency of the existing top-ranking set (see details in Supplementary Material). Note that this approach might be too restrictive for some complex diseases in which the information on known disease-gene associations is limited, failing to represent the functional diversity involved in the disease.

7. Visualisation of the top-ranking subnetwork

GUILDify v2.0 uses the JavaScript-based network visualisation library, Cytoscape.js [33], to show the subnetwork of the top-ranking proteins and the drugs targeting these proteins. The user can decide the cutoff to define the top ranked proteins to be visualised (top 1%, top 2% or functionally-coherent as mentioned above). In addition to seeds (green hexagons), top-ranking proteins (yellow circles) and drugs (blue diamonds), the subnetwork includes the proteins that connect the seeds to the largest connected component induced by seeds (named “linkers” and shown as grey circles, see details in Supplementary Material).

8. R package

We have included an R package in order to provide programmatic access to GUILDify v2.0 through R statistical computing environment (https://www.r-project.org/). The package implements methods to query and retrieve results from the web server as an R data frame, allowing users to run multiple queries for more high-throughput and/or systematic analyses. The package and documentation are available online at: http://sbi.upf.edu/guildify2.

GUILDify v2.0 workflow

1. Input

The interface of GUILDify v2.0 is designed to be simple and intuitive. The input varies slightly depending on the desired task: (i) a new search; (ii) retrieving results from a previous run; and (iii) calculating genetic and functional overlap between two previous runs. For a new search, we require two steps: first the selection of seeds (genes associated with a phenotype or drug) and second the selection of parameters to run the prioritisation algorithms. For the selection of seeds the user has to provide: (i) either keyword(s) describing the phenotype/drug of interest or a set of specific gene names separated by a semicolon; (ii) the species of interest (default value: *Homo sapiens*); (iii) the tissue of interest (default value: *All*); and (iv) the PIN source (default value: BIANA). If the user provides a keyword (or set of keywords) describing a phenotype or drug, the server searches genes containing the keyword in BIANA knowledge database (i.e. integrating information from many resources), otherwise it uses the list of provided gene names. The server shows the selected seeds, which can still be filtered and selected by the user. Then, for the prioritisation parameters the user can select to run the "disease module detection algorithm" (DIAMOnD, downloaded from https://github.com/dinaghiassian/DIAMOnD) [5] or to use one of the several prioritisation algorithms from the GUILD package (default value: NetScore with default parameters). Finally, to retrieve results, the required input is the job ID of a previous run, while for calculating genetic and functional overlap the inputs are two job IDs of previous runs.

2. Output

GUILDify v2.0 outputs the ranking of the nodes in the PIN and the visualisation of the subnetwork involving the top-ranking genes in a cytoscape.js panel. In addition, the output page has: (i) a panel showing the P-values of functional enrichment of the ranked nodes; (ii) two panels with functions enriched among the top-ranking nodes and seeds, respectively; and (iii) one panel with the drugs that target the top-ranking proteins.

For the "*Overlap between two results*" option, the server provides: (i) the list of the common top-ranking genes and the significance of the overlap assessed by a Fisher's exact test (see details in Supplementary Material); (ii) the network visualisation of the common top-ranking genes including the "linkers" (see above); (iii) the list of enriched functions of the common genes; iv) the list of common enriched functions of both results and the significance of the overlap; and v) the drugs targeting the proteins of the common PIN. Using this functionality, the users can identify the overlap between any two queries such as between two diseases, two drugs or a disease and a drug. Although we do not

provide the overlap between interactions of top-ranking proteins in a separate table, these interactions can be investigated in the network visualisation panel.

Case studies

1. Exploring the mechanistic links between rheumatoid arthritis and asthma

In multiple studies, rheumatoid arthritis and asthma are linked as a potential comorbidity, although the mechanisms underlying this association remain unclear [34]. Using the new functionality of GUILDify v2.0, we can assess the overlap between diseases and thus propose a potential mechanism to explain the association between them. Querying for "rheumatoid arthritis" and "asthma" returns 156 and 96 seeds, respectively coming from DisGeNET, OMIM, and UniProt. There are already 12 seeds in common (Fisher's exact test, one-sided P-value = $1.4 \cdot 10^{-9}$) and 18 common functions out of the total enriched functions of the seeds (P-value = $9.3 \cdot 10^{-23}$). After running GUILDify v2.0, we select 290 and 181 top ranked genes using functional-coherency based cutoff for rheumatoid arthritis and asthma, respectively. We find that the number of common genes increases to 55 (yielding a P-value = $5.9 \cdot 10^{-48}$), while the number of common functions (biological processes) increases to 31 (P-value = $8.1 \cdot 10^{-46}$). The link between these diseases is significant even when the seeds are removed from the top-ranking genes (see Supplementary Material). Among the shared top-ranking genes, we find Tumor Necrosis Factor (TNF), which has been proposed as a potential drug target for asthma and rheumatoid arthritis, and highlighted as a potential precursor of the comorbidity [12]. We also find HLA-DRB1 and several interleukins (IL18, IL1B, IL3), taking part of the immune response potentially involved in both diseases. Furthermore, the most common enriched functions relate to inflammatory processes such as "inflammatory response", "positive regulation of interferon-gamma production" and "positive regulation of T-helper 1 cell cytokine production". These functions appear again if we check the functions enriched by the common genes, along with other functions such as "T-helper 1 type immune response" or "negative regulation of type 2 immune response", highlighting the involvement of type 1 immune response in both diseases. As negative controls, we repeated the analysis using other disease pairs that are not likely to be comorbid such as "rheumatoid arthritis" - "breast cancer" and "asthma" - "breast cancer", finding drastically reduced number of genes in the overlap between these disease pairs (see Supplementary Material). The results can be further explored in Figure 1 and in the pre-calculated examples section of the web. Additionally, we compared the functional relevance of the top-ranking genes identified by NetScore with DIAMOnD, based on the analysis in *Sharma et al.* [35] (see Supplementary Material). We checked the enrichment of top-ranking genes among the pathways containing the seed genes of asthma and rheumatoid arthritis, showing that both methods significantly recover the pathways in each disease. Furthermore, NetScore identified more genes that belonged to the pathways shared between asthma and rheumatoid arthritis compared to DIAMOnD.

2. Study of the mechanism of non-small cell lung carcinoma drugs

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer. Typically induced by exposure to toxic substances, the NSCLC pathology has been specially associated with a mutation in

the Epidermal Growth Factor Receptor (EGFR) [36]. In a recent study, 9 drugs were proposed to treat this disease [37], 6 of them having drug-target interactions reported: Afatinib, Ceritinib, Crizotinib, Erlotinib, Gefitinib and Palbociclib. Given that we can now identify potentially new relationships between drugs and diseases using drugs as queries, we investigate whether the neighbourhood of the targets of these drugs in the PIN significantly overlaps with the neighbourhood of the genes associated with NSCLC. We used GUILDify v2.0 to define this neighbourhood. We observe that the genetic overlap is always significant, except for one of the drugs (Palbociclib, see Table 1).

We confirm the significance by applying the same approach to breast cancer, showing that Ceritinib, Crizotinib and Palbociclib produce a significant genetic overlap, although the number of common genes in each case is substantially lower than it is in NSCLC (see Table 1). These results are consistent with the fact that Palbociclib is primarily indicated for breast cancer and it has been recently repurposed for NSCLC [38]. The small but significant overlap of Ceritinib and Crizotinib suggests that these two drugs might also be considered as potential repurposing candidates. We note that using the top-ranking nodes increases the significance of the genetic overlap (with lower P-values) compared to the overlap using only seeds (genes associated with a pathophenotype and direct targets of drugs). The significant overlap between the top ranked genes identified using these drugs and the top ranked genes for NSCLC (but not for the top ranked genes for breast cancer) suggests that GUILDify v2.0 can help understanding how drugs exert their action on certain diseases. Indeed, the characterisation of the neighbourhood in the PIN that is affected by drugs opens a wide range of possibilities for drug repurposing research.

Methods

Datasets

GUILDify v2.0 uses BIANA [31] for the integration of biological interaction databases with information on drugs, genes, proteins, functions, pathways and diseases. To create the tissue-specific PINs, we use the RNAseq data from GTEx V7 [22]. Phenotype-gene associations are extracted from DisGeNET, OMIM, Uniprot, and Gene Ontology. Drug-target associations are taken from DrugBank [17], DGIdb [18], DrugCentral [19] and ChEMBL [20]. See Supplementary Material for details on the datasets.

Prioritisation algorithms

GUILDify v2.0 uses four different network-based prioritisation algorithms: NetShort, NetZcore, NetScore and DIAMOnD. For details on these algorithms see references [5,10,11] and the Supplementary Material.

Overlap and functional enrichment analysis

We use one-sided Fisher's exact test to calculate the overlap between two sets of genes or functions and use Benjamini-Hochberg multiple hypothesis testing procedure (where applicable). The functions

enriched among seeds and top-ranking nodes as well as common functions between two diseases are calculated as explained in a previous work [12] (see details in Supplementary Material).

Conflicts of Interest Statement

None declared.

Acknowledgements

J.A.P. and E.G. would like to acknowledge the technical support from GRIB IT team, in particular A. Gonzalez Pauner and M. A. Sánchez Gómez.

The authors received support from: ISCIII-FEDER (PI13/00082, CP10/00524, CPII16/00026); IMI-JU under grants agreements no. 116030 (TransQST) and no. 777365 (eTRANSAFE), resources of which are composed of financial contribution from the EU-FP7 (FP7/2007- 2013) and EFPIA companies in kind contribution; the EU H2020 Programme 2014-2020 under grant agreements no. 634143 (MedBioinformatics) and no. 676559 (Elixir-Excelerate); the Spanish Ministry of Economy (MINECO) [BIO2017-85329-R] [RYC-2015-17519]; “Unidad de Excelencia María de Maeztu”, funded by the Spanish Ministry of Economy [ref: MDM-2014-0370]. The Research Programme on Biomedical Informatics (GRIB) is a member of the Spanish National Bioinformatics Institute (INB), PRB2-ISCIII and is supported by grant PT13/0001/0023, of the PE I+D+i 2013-2016, funded by ISCIII and FEDER.

References

[1] M.F. Wangler, S. Yamamoto, H.-T. Chao, J.E. Posey, M. Westerfield, J. Postlethwait, P. Hieter, K.M. Boycott, P.M. Campeau, H.J. Bellen, Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research, Genetics. 207 (2017) 9–27. doi:10.1534/genetics.117.203067.

[2] Y. Bromberg, Chapter 15: Disease Gene Prioritization, PLoS Comput. Biol. 9 (2013) e1002902. doi:10.1371/journal.pcbi.1002902.

[3] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabási, Uncovering disease-disease relationships through the incomplete interactome, Science (80-.). 347 (2014) 1257601-1-1257601–8. doi:10.1126/science.1257601.

[4] X. Wang, N. Gulbahce, H. Yu, Network-based methods for human disease gene prediction, Brief. Funct. Genomics. 10 (2011) 280–293. doi:10.1093/bfgp/elr024.

[5] S.D. Ghiassian, J. Menche, A.L. Barabási, A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome, PLoS Comput. Biol. 11 (2015) e1004120. doi:10.1371/journal.pcbi.1004120.

[6] D. Nitsch, L.C. Tranchevent, J.P. Gonalves, J.K. Vogt, S.C. Madeira, Y. Moreau, PINTA: A web server for network-based gene prioritization from expression data, Nucleic Acids Res. 39 (2011) 334–338. doi:10.1093/nar/gkr289.

[7] K. Zuberi, M. Franz, H. Rodriguez, J. Montojo, C.T. Lopes, G.D. Bader, Q. Morris, GeneMANIA prediction server 2013 update., Nucleic Acids Res. 41 (2013) 115–122. doi:10.1093/nar/gkt533.

[8] A. Gottlieb, O. Magger, I. Berman, E. Ruppin, R. Sharan, Principle: A tool for associating genes with diseases via network propagation, Bioinformatics. 27 (2011) 3325–3326. doi:10.1093/bioinformatics/btr584.

[9] T. Kacprowski, N.T. Doncheva, M. Albrecht, NetworkPrioritizer: A versatile tool for network-based prioritization of candidate disease genes or other molecules, Bioinformatics. 29 (2013) 1471–1473. doi:10.1093/bioinformatics/btt164.

[10] E. Guney, J. García-garcía, B. Oliva, GUILDify : A web server for phenotypic characterization of genes through biological data integration and network-based prioritization algorithms, Bioinformatics. 30 (2014) 1789–1790. doi:10.1093/bioinformatics/btu092.

[11] E. Guney, B. Oliva, Exploiting Protein-Protein Interaction Networks for Genome-Wide Disease-Gene Prioritization, PLoS One. 7 (2012) e43557. doi:10.1371/journal.pone.0043557.

[12] C. Rubio-Perez, E. Guney, D. Aguilar, J. Piñero, J. Garcia-Garcia, B. Iadarola, F. Sanz, N. Fernandez-Fuentes, L.I. Furlong, B. Oliva, Genetic and functional characterization of disease associations explains comorbidity, Sci. Rep. 7 (2017) 6207. doi:10.1038/s41598-017-04939-4.

[13] F. Halakou, E. Sen Kilic, E. Cukuroglu, O. Keskin, A. Gursoy, Enriching Traditional Protein-protein Interaction Networks with Alternative Conformations of Proteins, Sci. Rep. 7 (2017) 7180. doi:10.1038/s41598-017-07351-0.

[14] A. Kominakis, A.L. Hager-Theodorides, E. Zoidis, A. Saridaki, G. Antonakos, G. Tsiamis, Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep, Genet. Sel. Evol. 49 (2017) 41. doi:10.1186/s12711-017-0316-3.

[15] E. Tejera, M. Cruz-Monteagudo, G. Burgos, M.E. Sánchez, A. Sánchez-Rodríguez, Y. Pérez-Castillo, F. Borges, M.N.D.S. Cordeiro, C. Paz-Y-Miño, I. Rebelo, Consensus strategy in genes prioritization and combined bioinformatics analysis for preeclampsia pathogenesis, BMC Med. Genomics. 10 (2017) 50. doi:10.1186/s12920-017-0286-x.

[16] J. Aguirre-Plans, J. Piñero, J. Menche, F. Sanz, L.I. Furlong, H.H.H.W. Schmidt, B. Oliva, E. Guney, Proximal pathway enrichment analysis for targeting comorbid diseases via network endopharmacology, Pharmaceuticals. 11 (2018) 61. doi:10.3390/ph11030061.

[17] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maclejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, Di. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: A major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082. doi:10.1093/nar/gkx1037.

[18] K.C. Cotto, A.H. Wagner, Y. Feng, S. Kiwala, C. Coffman, G. Spies, A. Wollam, N.C. Spies, O.L. Griffith, M. Griffith, DGIdb 3.0 : a redesign and expansion of the drug-gene interaction database, Nucleic Acids Res. 46 (2018) 1068–1073. doi:10.1093/nar/gkx1143.

[19] O. Ursu, J. Holmes, J. Knockel, C.G. Bologa, J.J. Yang, S.L. Mathias, S.J. Nelson, T.I. Oprea, DrugCentral : online drug compendium, Nucleic Acids Res. 45 (2017) 932–939. doi:10.1093/nar/gkw993.

[20] A. Gaulton, A. Hersey, A. Patr, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibri, M. Davies, N. Dedman, A. Karlsson, P. Magari, J.P. Overington, G. Papadatos, I. Smit, The ChEMBL database in 2017, Nucleic Acids Res. 45 (2017) 945–954. doi:10.1093/nar/gkw1074.

[21] M. Kitsak, A. Sharma, J. Menche, E. Guney, S.D. Ghiassian, J. Loscalzo, A.L. Barabási, Tissue Specificity of Human Disease Module, Sci. Rep. 6 (2016) 35241. doi:10.1038/srep35241.

[22] G. Consortium, Genetic effects on gene expression across human tissues, Nature. 550 (2017) 204–213. doi:10.1038/nature24277.

[23] J. Piñero, Á. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno, J. García-García, F. Sanz, L.I. Furlong, DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants, Nucleic Acids Res. 45 (2017) D833–D839. doi:10.1093/nar/gkw943.

[24] The UniProt Consortium, UniProt: The universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169. doi:10.1093/nar/gkw1099.

[25] A.P. Davis, C.J. Grondin, R.J. Johnson, D. Sciaky, B.L. King, R. McMorran, J. Wiegers, T.C. Wiegers, C.J. Mattingly, The Comparative Toxicogenomics Database: Update 2017, Nucleic Acids Res. 45 (2017) D972–D978. doi:10.1093/nar/gkw838.

[26] M.J. Landrum, J.M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, D.R. Maglott, ClinVar: Public archive of interpretations of clinically relevant variants, Nucleic Acids Res. 44 (2016) D862–D868. doi:10.1093/nar/gkv1222.

[27] A. Rath, A. Olry, F. Dhombres, M.M. Brandt, B. Urbero, S. Ayme, Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users, Hum. Mutat. 33 (2012) 803–808. doi:10.1002/humu.22078.

[28] J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, J. Morales, Z. MayPendlington, D. Welter, T. Burdett, L. Hindorff, P. Flicek, F. Cunningham, H. Parkinson, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog), Nucleic Acids Res. 45 (2017) D896–D901. doi:10.1093/nar/gkw1133.

[29] A. Gutiérrez-Sacristán, À. Bravo, M. Portero, O. Valverde, A. Armario, M.C. Blanco-Gandía, A. Farré, L. Fernández-

Ibarrondo, F. Fonseca, J. Giraldo, A. Leis, A. Mané, M.A. Mayer, S. Montagud-Romero, R. Nadal, J. Ortiz, F.J. Pavón, E. Perez, M. Rodríguez-Arias, A. Serrano, M. Torrens, V. Warnault, F. Sanz, L.I. Furlong, Text mining and expert curation to develop a database on psychiatric diseases and their genes, Database. 1650 (2017) 48–55. doi:10.1093/database/bax043.

[30] S. Köhler, N.A. Vasilevsky, M. Engelstad, E. Foster, J. McMurry, S. Aymé, G. Baynam, S.M. Bello, C.F. Boerkoel, K.M. Boycott, M. Brudno, O.J. Buske, P.F. Chinnery, V. Cipriani, L.E. Connell, H.J.S. Dawkins, L.E. DeMare, A.D. Devereau, B.B.A. De Vries, H. V. Firth, K. Freson, D. Greene, A. Hamosh, I. Helbig, C. Hum, J.A. Jähn, R. James, R. Krause, S.J.F. Laulederkind, H. Lochmüller, G.J. Lyon, S. Ogishima, A. Olry, W.H. Ouwehand, N. Pontikos, A. Rath, F. Schaefer, R.H. Scott, M. Segal, P.I. Sergouniotis, R. Sever, C.L. Smith, V. Straub, R. Thompson, C. Turner, E. Turro, M.W.M. Veltman, T. Vulliamy, J. Yu, J. Von Ziegenweidt, A. Zankl, S. Züchner, T. Zemojtel, J.O.B. Jacobsen, T. Groza, D. Smedley, C.J. Mungall, M. Haendel, P.N. Robinson, The human phenotype ontology in 2017, Nucleic Acids Res. 45 (2017) D865–D876. doi:10.1093/nar/gkw1039.

[31] J. Garcia-Garcia, E. Guney, R. Aragues, J. Planas-Iglesias, B. Oliva, Biana: a software framework for compiling biological interactions and analyzing networks, BMC Bioinformatics. 11 (2010) 56. doi:10.1186/1471-2105-11-56.

[32] W.A. Kibbe, C. Arze, V. Felix, E. Mitraka, E. Bolton, G. Fu, C.J. Mungall, J.X. Binder, J. Malone, D. Vasant, H. Parkinson, L.M. Schriml, Disease Ontology 2015 update : an expanded and updated database of human diseases for linking biomedical knowledge through disease data, Nucleic Acids Res. 43 (2015) 1071–1078. doi:10.1093/nar/gku1011.

[33] M. Franz, C.T. Lopes, G. Huck, Y. Dong, O. Sumer, G.D. Bader, Cytoscape.js: A graph theory library for visualisation and analysis, Bioinformatics. 32 (2015) 309–311. doi:10.1093/bioinformatics/btv557.

[34] M.C. Rolfes, Y.J. Juhn, S.I. Wi, Y.H. Sheen, Asthma and the risk of rheumatoid arthritis: An insight into the heterogeneity and phenotypes of asthma, Tuberc. Respir. Dis. (Seoul). 80 (2017) 113–135. doi:10.4046/trd.2017.80.2.113.

[35] A. Sharma, J. Menche, C. Chris Huang, T. Ort, X. Zhou, M. Kitsak, N. Sahni, D. Thibault, L. Voung, F. Guo, S.D. Ghiassian, N. Gulbahce, F. Baribaud, J. Tocker, R. Dobrin, E. Barnathan, H. Liu, R.A. Panettieri, K.G. Tantisira, W. Qiu, B.A. Raby, E.K. Silverman, M. Vidal, S.T. Weiss, A.L. Barabási, A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma, Hum. Mol. Genet. 24 (2014) 3005–3020. doi:10.1093/hmg/ddv001.

[36] G. Bethune, D. Bethune, N. Ridgway, Z. Xu, Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update, J. Thorac. Dis. 2 (2010) 48–51. doi:10.3978/j.issn.2072-1439.2010.02.01.017.

[37] C. Rubio-Perez, D. Tamborero, M.P. Schroeder, A.A. Antolín, J. Deu-Pons, C. Perez-Llamas, J. Mestres, A. Gonzalez-Perez, N. Lopez-Bigas, In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities, Cancer Cell. 27 (2015) 382–396. doi:10.1016/j.ccell.2015.02.007.

[38] J. Zhou, S. Zhang, X. Chen, X. Zheng, Y. Yao, G. Lu, J. Zhou, Palbociclib, a selective CDK4/6 inhibitor, enhances the effect of selumetinib in RAS-driven non-small cell lung cancer, Cancer Lett. 408 (2017) 130–137. doi:10.1016/j.canlet.2017.08.031.

TABLE AND FIGURES LEGENDS

Figure 1. GUILDify v2.0 example study on the comorbidity between asthma and rheumatoid arthritis. First, we run the prioritisations of the two diseases by searching (1) and selecting (2) the genes. After obtaining the ranking of proteins from the prioritisation (3), we use both job IDs to check their overlap (4) and inspect the genetic and functional relationships between them (see details at http://sbi.upf.edu/guildify2 in the pre-calculated examples section.

Table 1. Results of the genetic and functional overlap between the subnetwork of genes associated with “non small cell lung carcinoma” and “breast cancer” (top ranking genes and seeds) and the subnetwork of genes associated with the targets of drugs Afatinib, Ceritinib, Crizotinib, Erlotinib, Gefitinib and Palbociclib (drug targets and top-ranking genes obtained with GUILDify v2.0). P-values shown have been corrected using the Benjamini-Hochberg correction for multiple tests. Results with non-significant P-value are highlighted in red.