



OPEN

Drug repurposing improves disease targeting 11-fold and can be augmented by network module targeting, applied to COVID-19

Inés Rivero-García, Miguel Castresana-Aguirre, Luca Guglielmo, Dimitri Guala & Erik L. L. Sonnhammer

This analysis presents a systematic evaluation of the extent of therapeutic opportunities that can be obtained from drug repurposing by connecting drug targets with disease genes. When using FDA-approved indications as a reference level we found that drug repurposing can offer an average of an 11-fold increase in disease coverage, with the maximum number of diseases covered per drug being increased from 134 to 167 after extending the drug targets with their high confidence first neighbors. Additionally, by network analysis to connect drugs to disease modules we found that drugs on average target 4 disease modules, yet the similarity between disease modules targeted by the same drug is generally low and the maximum number of disease modules targeted per drug increases from 158 to 229 when drug targets are neighbor-extended. Moreover, our results highlight that drug repurposing is more dependent on target proteins being shared between diseases than on polypharmacological properties of drugs. We apply our drug repurposing and network module analysis to COVID-19 and show that Fostamatinib is the drug with the highest module coverage.

Drug discovery has traditionally been centered around the “one drug—one gene—one disease” paradigm with the aim of achieving a therapeutic outcome while minimizing detrimental off-target effects. This perspective has proven to be successful in some cases, like the BCR-ABL tyrosine kinase inhibitor Imatinib¹, but solely adhering to this model has its downsides. Firstly, the drug development process requires large investments to be successful, averaging at \$2–3 billion and 13–15 years per medication to achieve regulatory marketing approval². Secondly, and as a consequence of the financial risks, not all diseases are targeted in drug discovery, which leaves patients affected by rare conditions with limited therapeutic options³.

Two concepts that could mitigate these problems are polypharmacology and drug repurposing. The concept of polypharmacology refers to the ability of some drugs to target more than one protein⁴. Although it might appear as undesirable at first, polypharmacology can modulate several cellular pathways simultaneously, thereby increasing treatment efficacy⁵. Drug repurposing or repositioning, defined as the use of an approved drug for a new therapeutic indication⁶, can make the market life of a drug more appealing by decreasing development costs to \$40–80 million and 3–12 years for a new indication². The combination of polypharmacological and repositioning strategies could offer therapeutic opportunities for patients with any condition, and its importance is strongly exemplified in the case of fast-evolving pandemic diseases such as the ongoing COVID-19, which has to date caused more than 212 million cases and more than 4 million deaths⁷. Worldwide efforts to find therapeutic candidates for COVID-19^{8–11}, such as the Coronavirus Treatment Acceleration Program by the FDA¹², have been put into action and would benefit from therapeutic opportunities provided by drug repurposing in order to facilitate resource planning and allocation. We here approach drug repurposing from a network and module-based perspective and demonstrate its value for COVID-19.

Networks of functional associations present a convenient model of intracellular relations between proteins, including physical, regulatory, and functional interactions¹³. Such networks exhibit emergent properties¹⁴ on the systems level that manifest themselves in phenotypes and diseases¹⁵, that are not encoded in a single gene. Even for Mendelian diseases, where phenotypes are caused by single mutations, there is a plethora of modifier genes influencing the final outcome¹⁶. In general, genes associated with a given disease tend to cluster together when mapped to a functional association network, forming so-called disease modules¹⁷. For a disease phenotype

Department of Biochemistry and Biophysics, Stockholm University, Science for Life Laboratory, Box 1031, 17121 Solna, Sweden. email: erik.sonnhammer@dbb.su.se

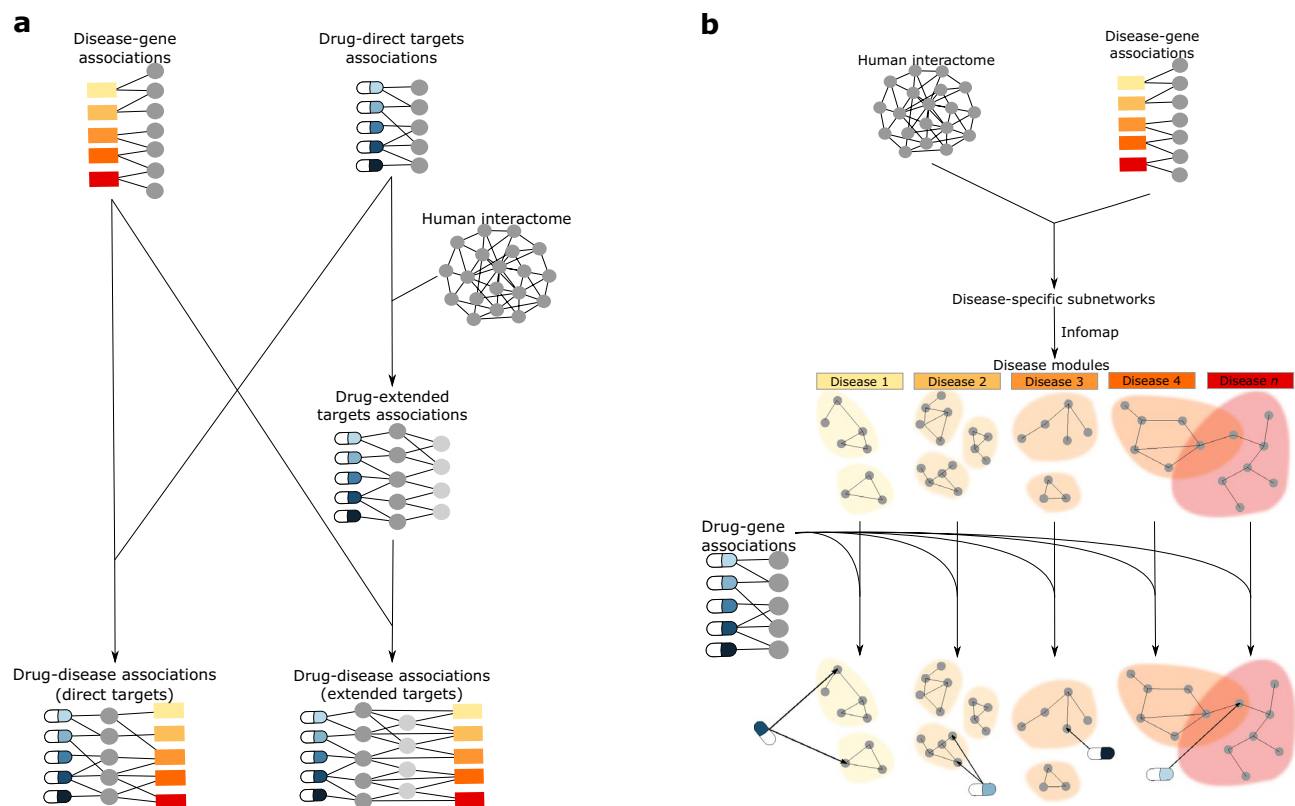


Figure 1. Schematic overview of study workflow. **(a)** The human interactome (FunCoup³⁸) was used to map drug target associations for FDA-approved drugs from DrugBank⁴⁰ and disease-gene associations from DisGeNET⁴². The sets of drug targets were also extended using first order neighbors in the human interactome. The overlap between drug targets and disease gene sets was analyzed to determine the potential disease leverage that could be offered by repurposing. **(b)** In order to find disease modules, the interactions between disease genes retrieved from the human interactome were partitioned using Infomap⁴⁷. The overlap between drug targets and disease genes was later analyzed to map drugs to the disease modules.

to manifest itself, the integrity of the underlying disease module needs to be perturbed¹⁸. The disease module hypothesis has increased our understanding of molecular pathological mechanisms and has been successfully applied to improve therapeutic strategies¹⁹.

Advantages, strategies, and successful implementations of network-based drug repurposing have been described previously^{20–23}. However, the extent of the therapeutic opportunities that can be gained from drug repurposing has to our knowledge not yet been assessed. In this study we combine drug targets with disease-associated genes to determine the extent of therapeutic opportunities to be gained from network-based drug repurposing, both with and without using a network (Fig. 1a). Additionally, the analysis is performed in the context of disease modules found in the human functional interactome (Fig. 1b). Our main findings suggest that drug repurposing can offer an average 11-fold potential increase in drug–disease associations, here referred to as disease leverage, and that polypharmacological drugs tend to have all targets in one or a few modules. These findings lead us to conclude that drug repurposing at the module level could benefit more from the pleiotropic nature of some disease genes than from the polypharmacological action of drugs.

Results

To gain insights into the therapeutic opportunities yet to be exploited from drug repurposing we performed three levels of analyses: at the drug target level, at the disease level and at the disease module level (Fig. 1). With respect to the target level, we quantified the number of direct targets per drug and also extended these to consider their first neighbors in the human functional interactome. Regarding the disease level, we quantified the number of diseases each drug can be linked to by mapping drug targets to diseases. Using the same method, we also analyzed how drugs are linked to disease modules.

Direct drug–protein target associations: most drugs target few proteins.. To establish the specificity of the currently approved drugs we assessed the number of proteins targeted by each of the drugs. The number of direct protein targets per drug ranges from 1 to 251, with a median value of 3 (Fig. 2a). After ranking the drugs according to the number of proteins they target we could identify the spleen tyrosine kinase (SYK) kinase inhibitor Fostamatinib, used e.g., in the treatment of rheumatoid arthritis and immune thrombocytopenia purpura²⁴, at the top of the list. We note that 6 of the 10 top drugs in this ranking are naturally occurring

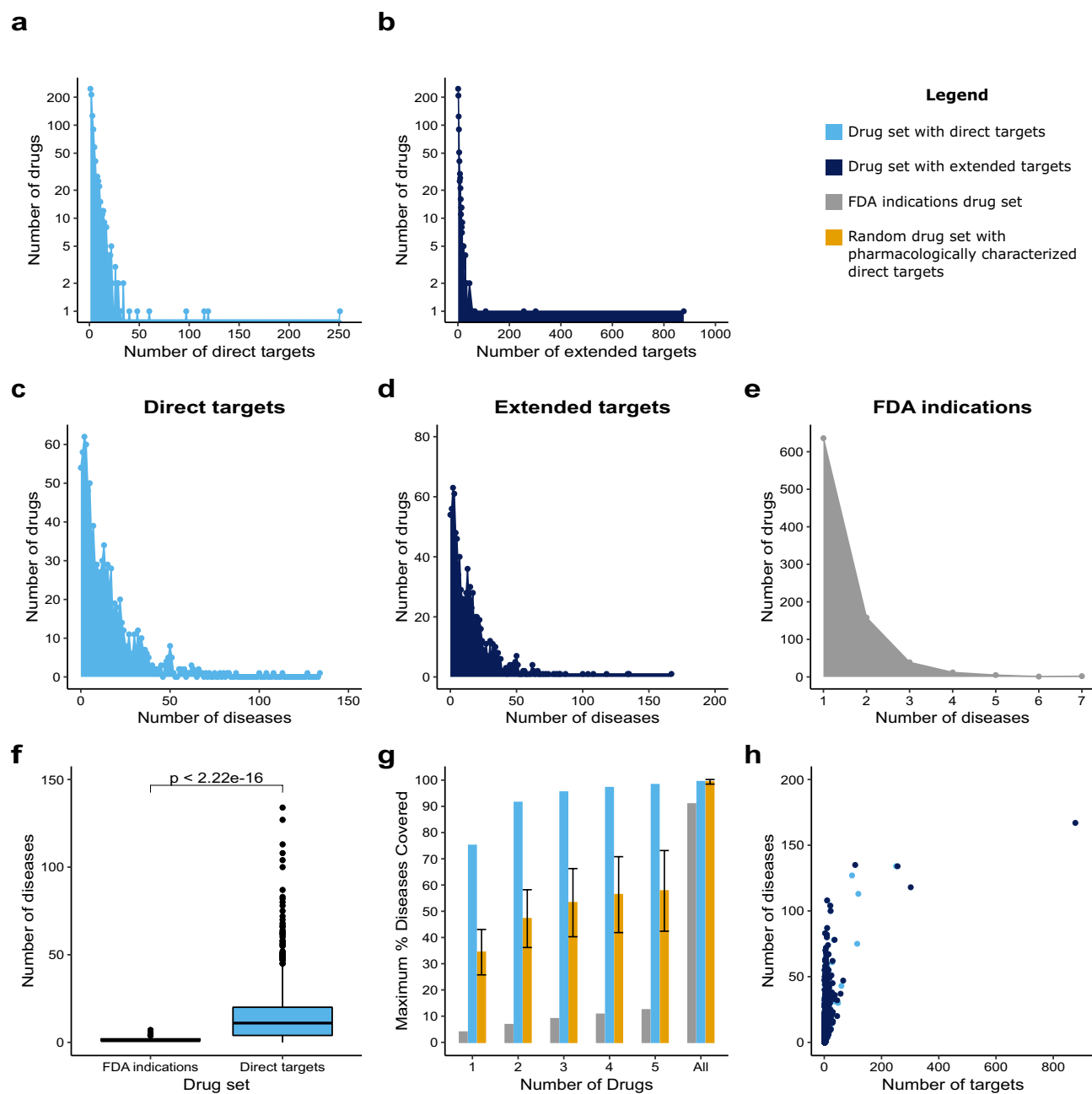


Figure 2. Drug repurposing may offer an average 11-fold increase of disease leverage. (a) Distribution of the number of genes targeted by FDA-approved drugs e.g., drug targets. (b) Distribution of the number of extended drug targets. (c) Distribution of the number of diseases covered by direct targets of FDA-approved drugs. (d) Distribution of the number of diseases covered by the extended targets of the drug set. (e) Distribution of the number of diseases targeted per drug for the “FDA indications” data set. (f) Statistical comparison of the distributions of disease coverage between the “FDA indications” data set and the direct drug target data set. (g) Quantification of the maximum number of diseases covered by 1, 2, 3, 4, 5 and all drugs for the “FDA indications”, the drug set with direct targets and a randomized version of the drug set. In the randomized drug set each drug has the same number of targets as in the original drug set, but the target genes are randomly chosen from the human genome. This random drug set represents the background levels of disease coverage. (h) Correlation between the number of targets and the number of diseases mapped to a drug. For the direct targets: $\rho = 0.636$, $p\text{-value} < 2.2 \times 10^{-16}$. For the extended targets: $\rho = 0.644$, $p\text{-value} < 2.2 \times 10^{-16}$.

ions and small molecules that act as enzyme cofactors, such as NADH or Copper (Sup. Table 1). When only the pharmacologically characterized targets of a drug were considered (e.g., proteins for which the action of the drug on them has been experimentally characterized) the distribution of direct targets per drug has a similar shape,

with the number of direct targets ranging from 1 to 33 and a median value of 1 (Sup. Fig. 1 online). The top-ranked drug is NADH, a nutraceutical drug used e.g., in the management of Parkinson's disease and in dietary supplementation therapies²⁵. Also among the top ranking drugs are the GABA-receptor inhibitors Clotiazepam, Clonazepam, and Flurazepam (Sup. Table 2 online). The top-10 drugs in these two rankings are very different, and it is worth noting for pharmacologically characterized targets that there is only one naturally occurring molecule in humans, NADH, while the “all targets” top-10 list contains 6 such drugs. This highlights the target specificity principle that guides drug design and facilitates marketing approval.

Extending the target sets: drug repurposing perspectives considering network neighbors. Because the pharmacological perturbation of a protein can affect its interacting partners, we also extended the set of direct targets with their network neighbors and performed the same analysis. After target extension, drugs had between 1 and 877 extended targets, with a median of 3 (Fig. 2b). When only the pharmacologically characterized direct targets were considered, the distribution of number of extended targets per drug ranged from 1 to 172 with a median of 1 (Sup. Fig. 1 online). In both cases, the ranking of the top scoring drugs is highly similar to the ranking produced by studying only the direct targets. Some exceptions exist—for example cholic acid, used for the treatment of e.g. peroxisomal and bile acid synthesis disorders, climbed from position 7 in the direct targets ranking to position 9 in the extended targets ranking (Sup. Table 1 online, respectively).

Drug–disease associations: an average 11-fold disease leverage from drug repurposing. To assess the distribution of FDA-approved drug–disease associations, the “FDA indications” dataset was used. The distribution of FDA indications per drug ranged from 1 to 7, with a median value of 1 (Fig. 2e). It is noteworthy that 6 out of the top 10 drugs associated with the most diseases (Sup. Table 3 online) are biotechnological products, e.g., pharmacological macromolecules derived from or produced in biological systems such as tissues or cells. Examples include the soluble TNF receptor recombinant protein Etanercept, used in the treatment of rheumatoid arthritis²⁶, and the immune checkpoint inhibitor antibody Avelumab, which prevents the PD1/PDL1 interaction that restricts the immune defense against tumors²⁷.

An overlap between a drug's targets and genes associated with a certain disease could indicate a potential therapeutic effect of the drug on the disease phenotype. To estimate how many diseases are targeted by each drug, we systematically studied this overlap in the “direct targets” dataset. The overlap ranged between 0 and 134 diseases, with a median of 11 diseases, per drug (Fig. 2c). The drugs targeting more proteins tend to show an overlap with a larger number of diseases (Fig. 2h). Fostamatinib was once again ranked at the top of this list, together with several drugs with anti-inflammatory properties such as Aspirin, Ibuprofen and Dexibuprofen (Sup. Table 1 online). As drugs with FDA indications only target 1 disease on average, this suggests that drug repurposing on average offers an 11-fold increase in disease coverage (Fig. 2f). If only the pharmacologically characterized drug targets are considered, each drug targets between 0 and 106 diseases (Sup. Fig. 1 online, Sup. Table 2 online). This still increases the average disease leverage fivefold as compared to the FDA indication drug–disease associations (Sup. Fig. 1 online). For both all targets and pharmacologically characterized targets only, the drug targets are four times more shared between diseases than would be expected (Sup. Fig. 2 online, p -value $< 2.2 \times 10^{-16}$).

Combinatorial drug therapy. To further examine the opportunities that drug repurposing opens for disease management, we calculated the maximum number of diseases that a small number of drugs can cover (Fig. 2g). Strikingly, already three drugs can cover 95% of all diseases. These drugs are Fostamatinib, Zinc, and Neonatal foreskin keratinocyte. We note that the disease coverage is dramatically higher than for FDA indications, which only achieves maximally 10% of the diseases for 5 drugs. The disease coverage up to 5 drugs is significantly (p -value $< 2.2 \times 10^{-16}$) higher than expected by random sampling (Fig. 2f). Similar results were obtained when considering the pharmacologically characterized drug targets only (Sup. Fig. 1 online). These results support the fact that drug repurposing has a high potential for combinatorial therapeutic opportunities yet to be exploited.

Increased disease coverage after extending the drug targets. To study the effects of extending the sets of drug targeted proteins on the coverage of the diseaseome we assessed the overlap of the extended drug target sets and disease genes. Together, all drugs in the data set cover 176 of the 177 diseases. Individually, drugs can cover between 0 and 167 diseases, with a median of 11 diseases (Fig. 2d, Sup. Table 1 online). If only the pharmacologically characterized targets are extended the number of diseases covered by a single drug ranges between 0 and 106, with a median of 5 (Sup. Fig. 1 online, Sup. Table 2 online). This suggests that modulating a protein by the pharmacological perturbation of one of its functional neighbors could give further opportunities for disease management by drug repurposing.

Drug–disease module associations: opportunities for drug repurposing at the disease module level. Based on the hypothesis that the perturbation of disease modules is the underlying cause of disease, we sought to examine how drugs target disease modules. The community finder Infomap retrieved disease modules of at least three genes for 157 out of the 177 diseases in the data set (89%). The median number of modules per disease was 2 with a range from 1 to 26. Experimental liver cirrhosis demonstrated the highest number of modules. The identified disease modules had a median size of five genes. The largest disease module was found for COVID-19, where 423 of 572 genes were part of a single connected component (Sup. Fig. 3 online).

To find the most universal drugs at the module level, we counted the number of modules targeted by each drug on a gene-overlap and disease-independent basis. This number ranged from 0 to 158 (32%) modules, with

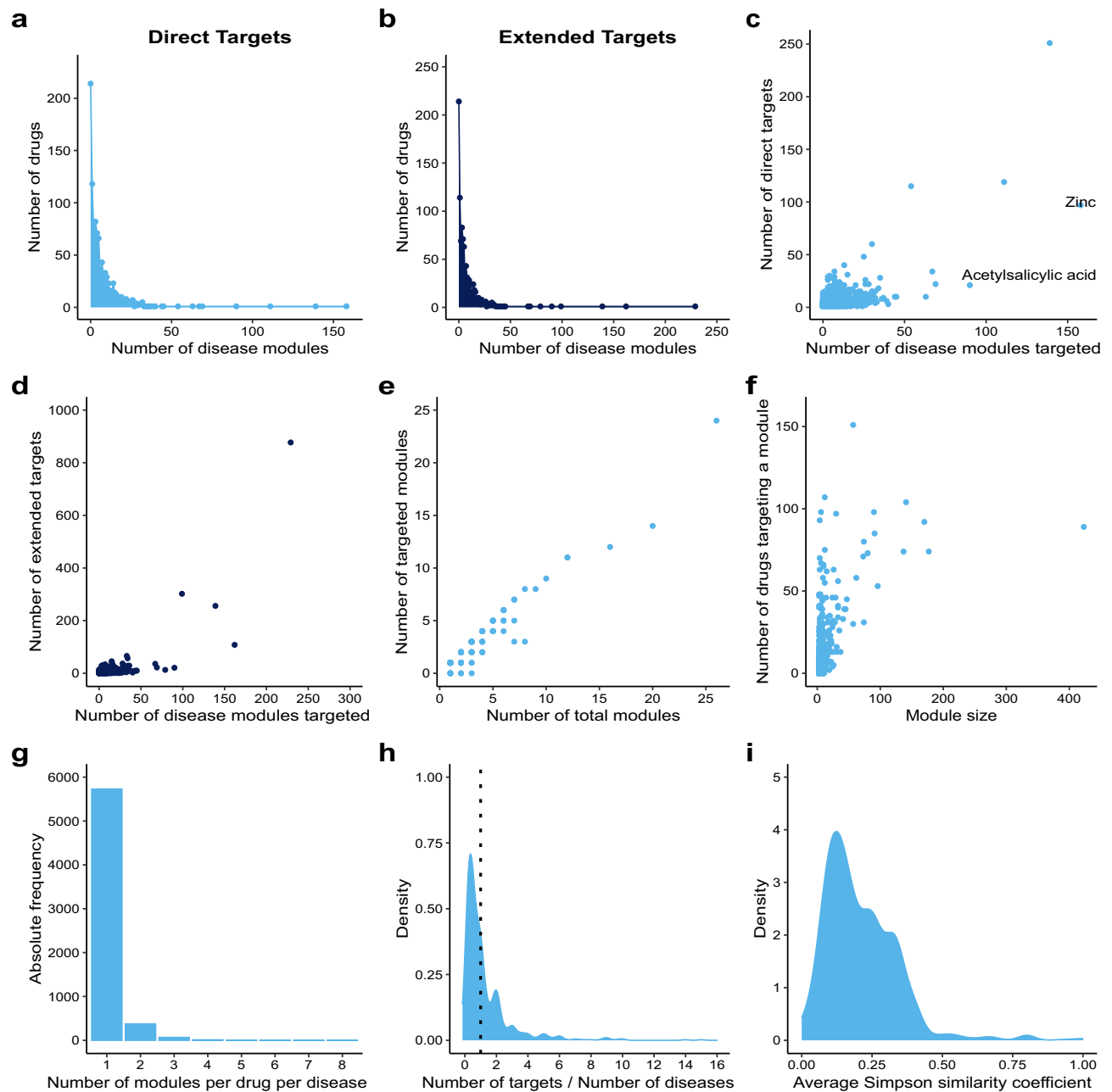


Figure 3. Most drugs target few modules. **(a)** Distribution of the number of modules targeted per drug considering the direct targets. **(b)** Distribution of the number of modules targeted per drug considering their extended targets. **(c)** Correlation between the number of disease modules targeted by a drug and its number of direct gene targets ($\rho = 0.495$, $p\text{-value} < 2.2 \times 10^{-16}$). **(d)** Correlation between the number of disease modules targeted by a drug and its number of extended gene targets ($\rho = 0.503$, $p\text{-value} < 2.2 \times 10^{-16}$). **(e)** Correlation between the total number of modules and the number of drug-targeted modules in a disease ($\rho = 0.903$, $p\text{-value} < 2.2 \times 10^{-16}$). **(f)** Correlation between the size of a disease module (number of genes) and the number of drugs targeting it ($\rho = 0.488$, $p\text{-value} < 2.2 \times 10^{-16}$). **(g)** Number of modules targeted per drug in a disease targeted by that drug. **(h)** Density plot of the number of direct targets per disease for all drugs (average ratio = 0.52). **(i)** Density plot of the Szymkiewicz–Simpson similarity coefficients between disease modules targeted by single drugs (average Szymkiewicz–Simpson similarity coefficient = 0.211).

a median value of 4 (Fig. 3a). Overall, the top drugs in this ranking were highly similar to the best ranked ones based on diseases, with Zinc as the top-ranked drug (Sup. Table 1 online). The similarity may be explained by the fact that the number of diseases and disease modules covered by single drugs is highly correlated (Pearson's correlation coefficient = 0.98). When only the pharmacologically characterized drug targets are considered the distribution ranged between 0 and 44, with a median of 1 and Neonatal foreskin keratinocyte as its best performing agent (Sup. Fig. 4 online, Sup. Table 2 online).

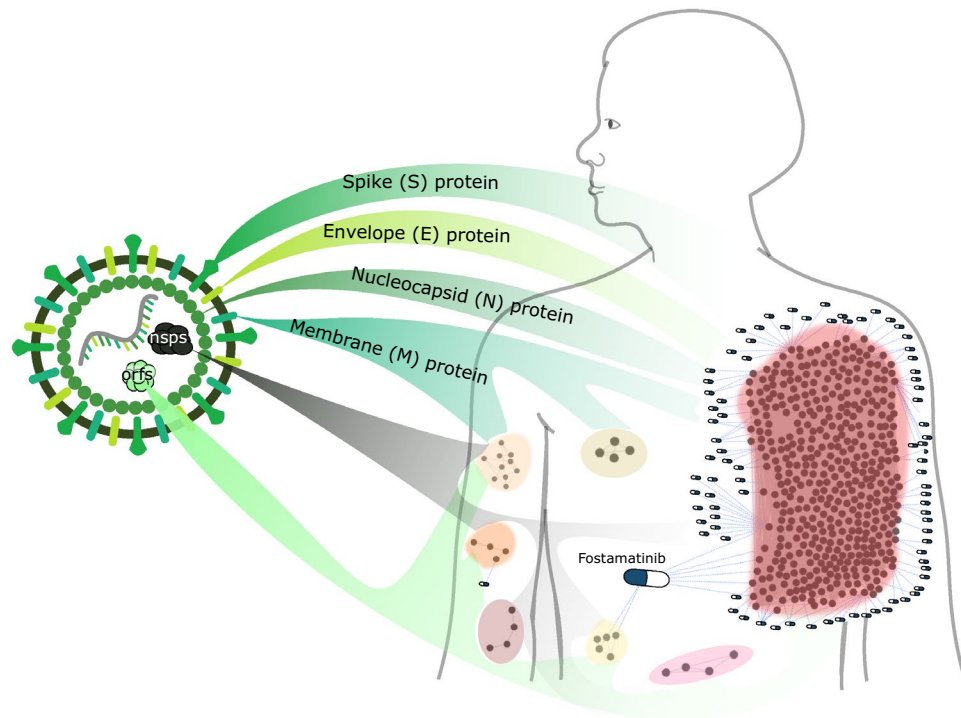


Figure 4. Drug targeting of the COVID-19 disease modules. Bipartite network linking drugs with their targets in the COVID-19 disease modules, where nsps are the SARS-CoV-2 non-structural proteins and orfs are its other open reading frames. Several polypharmacological drugs can be observed, highlighting Fostamatinib as the only one that is linked to two modules. The remaining polypharmacological drugs are linked to one module only.

We also investigated how the extension of the drug targets affected drug targeting at the disease module level. Single drugs with extended targets covered 0 and 229 disease modules (46%), with a median value of 4 (Fig. 3b). The best performing drug in this category was Copper (Sup. Table 1 online). For pharmacologically characterized targets, between 0 and 74 modules were targeted per drug with a median of 1. Fostamatinib was then the highest ranked drug (Sup. Fig. 4 and Sup. Table 2 online). The high correlation between the number of diseases and the number of modules targeted by single drugs was maintained when using the extended targets (Pearson's correlation coefficients = 0.98).

Given that most drugs target few modules, are the targets of a drug typically in the same module? To address this question, we calculated the correlation between the number of direct targets of a given drug and the number of modules it targets. If the targets of a drug were present in different modules we could expect a high positive correlation between these two variables. However, the Spearman correlation is only 0.495, thus targets of one drug tend to be part of the same disease module, albeit with some outliers such as Acetylsalicylic acid, which targets 90 disease modules with just 21 targets (Fig. 3c). Similar results were obtained for the extended drug target sets (Fig. 3d).

As an example of module targeting, Fig. 4 shows a bipartite network of drugs and disease genes in COVID-19. This network is composed by the human proteins that interact with SARS-CoV-2 proteins²⁸ and the 90 drugs (Sup. Table 4 online) that target them. We found seven modules in the COVID-19 network. Although there are several polypharmacological drugs targeting this disease, Fostamatinib is the only one that targets different modules. Fostamatinib has been found to reduce the levels of membrane-bound MUC1^{29,30}, the levels of Neutrophil Extracellular Traps³¹ and platelet activation³². Altogether, this results points toward the potential beneficial effect of Fostamatinib in severe COVID-19 patients, which is being currently evaluated by two clinical trials (Clinicaltrials.gov identifiers: NCT04579393 and NCT04581954³³). Moreover, the imbalance in the number of drugs targeting the different modules points towards the importance of employing drugs that together can target all modules (e.g., pathophysiological mechanisms) involved in the disease.

With the hypothesis that different modules represent different pathobiological molecular mechanisms, we investigated if diseases with a higher number of modules have more modules that are targeted by drugs. The correlation between the total number of modules and the number of drug-targeted modules in a given disease supports this idea (Spearman correlation coefficient = 0.90, Fig. 3e). Since there is also strong correlation between the number of genes associated with a disease and its number of modules (Spearman's correlation coefficient = 0.77, Pearson's correlation coefficient = 0.89), a disease with more genes has higher chances of having more modules as well as a higher chance of being targeted by a drug.

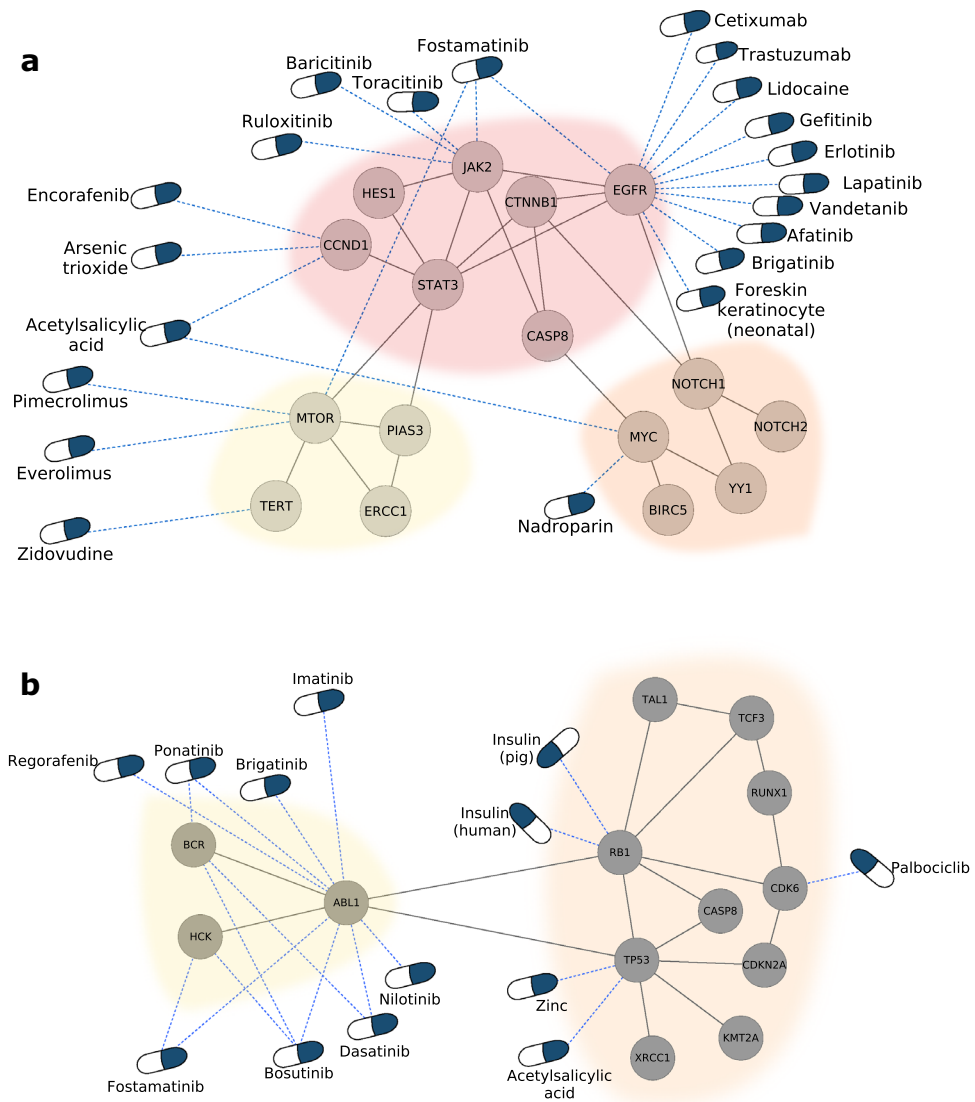


Figure 5. Network examples of disease module targeting by drugs. **(a)** Disease modules for uterine cervical neoplasm accompanied by the drugs that target them. The biggest module is the one associated with most drugs. There are two polypharmacological drugs: Acetylsalicylic acid and Fostamatinib, both of them targeting two out of the three disease modules. **(b)** Disease modules for childhood acute lymphoblastic leukemia with the drugs that target them. In this case the smallest module is the one associated with most drugs. There are four polypharmacological drugs (Ponatinib, Fostamatinib, Bosutinib and Dasatinib), all of them associated only with the small module.

Is the number of drugs targeting a module affected by the size of the module? The correlation between these two variables was found to be 0.535 (Fig. 3f), suggesting a general trend, although there are several diseases in which the smallest module has the highest number of drugs. As an example of the general trend, 76% (19 of 25) of the drugs targeting uterine cervical neoplasm target the largest module of 7 genes, while the rest of the drugs are divided between two modules of 5 and 4 genes respectively (Fig. 5a). An example of the opposite trend is childhood acute lymphoblastic leukemia, where more than 60% (8 of 13) of the drugs target the smallest module of 3 genes, while the largest module with 10 genes is targeted by only 5 drugs (Fig. 5b). For drugs targeting a single disease, the most common pattern is that each drug targets a single module, yet exceptions exist where up to 8 modules are targeted by a single drug, which is the case for Fostamatinib in experimental liver cirrhosis and malignant neoplasm of breast (Fig. 3g). Further examples can be found in Sup. Fig. 5 online.

Lastly, we asked whether drugs targeting more than one disease do this by targeting the same or distinct proteins in different diseases. Our results show that on average, drugs have 0.75 targets per disease (Fig. 3h), e.g., there is a clear trend that multiple diseases are targeted because they share the same target. We further calculated the overlap between disease modules targeted by a single drug. The distribution of these Szymkiewicz–Simpson coefficients shows that the modules targeted by the same drug are very different from each other (Fig. 3i). Putting

these results together, we conclude that drug repurposing opportunities are mainly due to the pleiotropic nature of disease genes, rather than polypharmacological properties of drugs.

Discussion

The main aim of this study is to assess the disease leverage that can be achieved by drug repurposing, either via direct targeting or via a network of functional interactions. We also studied how network modules within disease gene sets are targeted. Our findings demonstrate that drug repurposing can offer an 11-fold increase in disease coverage on average, and that disease modules can be used to pinpoint untargeted pathological mechanisms and identify polypharmacological drugs that can perturb a module at multiple targets, reducing the chances of drug resistance³⁴.

At first, we established some basic characteristics of approved drugs. The low number of direct targets per drug, median value of 3, is in line with the traditional aim of drug discovery⁶ to minimize off target effects making the mode of action more easily explainable and avoiding adverse drug reactions. After having examined direct drug targets, we sought to quantify disease coverage by available drugs to answer our main research question: what is the potential of drug repurposing with respect to disease coverage. Our results show that drug repurposing can offer an average 11-fold increase in disease leverage, giving significant results when compared to a background set of target genes. While most of this repurposability potential is explained by gene pleiotropy, there is an average twofold increase in disease leverage from polypharmacological properties of drugs alone. Moreover, expanding the sets of direct drug targets prior to repurposing may increase disease coverage even further and could provide novel therapeutic strategies.

In an attempt to facilitate identification of drug repurposing candidates, using the disease module hypothesis, we have generated a table that maps each disease gene to disease(s), module(s), and drug(s), either directly or via network extension (Sup. Table 5 online). Pharmacological characterization is also indicated. This resource could be used to identify repurposing candidates given a disease or a set of genes. Moreover, it can be used to find additional drugs that target other modules, given an existing drug–disease combination. However, additional studies should be performed to assess if a given drug can be repurposed for a particular disease because the fact that a drug has targets in a set of disease genes does not guarantee a therapeutic effect on that disease. Gene expression signature studies and structural predictions could help to assess drug–disease compatibility and to select the most suitable candidates for further testing³⁵.

Genes associated with a single disease typically do not form a single connected network component. This could be due to the incompleteness of the network¹⁸ but also to the existence of multiple mechanisms, each linked to a distinct subset of genes forming distinct disease modules. Methods such as Disease module detection (DIAMOND)³⁶ and Seed connector algorithm (SCA)³⁷ aim to connect the genes of a disease into larger modules by adding connector genes. In the case of DIAMOND, submodules are then detected by clustering methods. We here followed a similar approach that allows disease genes to be part of multiple distinct modules. This idea is supported by the facts that diseases are often heterogeneous and that disease genes can be associated to either the cause or the effect of the disease. To define disease modules we used InfoMap which is a widely used module detection algorithm that has performed well in benchmarks³⁸.

The identification of disease modules was done in a disease-independent manner, which resulted in low similarity between modules of different diseases (average Szymkiewicz–Simpson similarity coefficient = 0.015). We also found that drugs tend to target very few disease modules, which once again points towards a very targeted approach in drug design. Moreover, drug targets tend to be part of the same disease module, and diseases with more modules tend to be targeted by more drugs, despite lack of a clear correlation between module size and the number of drugs that target it. The asymmetry of drug targeting across modules opens the possibility of finding drugs with the widest coverage at the disease module level. Such drugs, like Fostatinib in the COVID-19 case, can potentially achieve higher efficacy because they can interfere with more mechanisms.

This study explores and showcases the therapeutic potential that could be obtained from drug repurposing in conjunction with network analysis. Disease modules can be helpful for GWAS interpretation³⁹, identifying key pathological proteins⁴⁰, guiding the design of effective therapeutic strategies⁴¹, and detecting drug repurposing opportunities⁴². These strategies could increase treatment efficacy by targeting multiple disease modules and genes, and selecting the drug targets that lead to a therapeutic, rather than symptomatic, treatment of disease³. Looking forward, the use of the human interactome information for drug repurposing strategies may lead to economic, social and medical benefits in the treatment of human disease.

Methods

Human interactome. The human interactome was retrieved from FunCoup v4.1⁴³. FunCoup is an online resource of functional association networks in multiple species. It is constructed by naïve Bayesian integration of ten different types of evidence of functional interactions, including: domain–domain and protein–protein interactions, mRNA and protein co-expression, genetic interactions, co-regulation by transcription factors and micro RNA and co-evolution. Additionally, the functional associations are transferred to multiple species using orthologs identified by InParanoid⁴⁴. The retrieved *Homo sapiens* interactome from FunCoup contains 5,315,787 interactions among 17,402 nodes and was used because it is the largest network comprised exclusively of experimental data. The human interactome data was used to expand the drug target gene sets and identify interactions between disease genes needed for identification of disease modules.

Drug–gene targets data set. Drug–gene target data was retrieved from DrugBank (version 5.1.5)⁴⁵. To ensure the clinical utility of the results, only FDA-approved drugs were considered. Subsequently, drug–gene mappings for which a gene lacked an Ensembl ID or was not present in FunCoup were removed. Drugs with

Data set	Number of drugs	Number of genes	Number of drug–gene mappings
Drugs with direct targets	985	1590	5286
Drugs with pharmacologically characterized direct targets	762	527	1374
Drugs with extended targets	985	6017	1374
Drugs with pharmacologically characterized extended targets	762	4119	44,386
FDA indications	853	NA	NA

Table 1. Drug–gene targets data sets. FDA indications only connect drugs to diseases, not to targets.

identical gene targets and indications were grouped together keeping only one drug as a group representative while removing the rest of the group from the analysis. This yielded a drug set consisting of 985 drugs. Additionally, the Therapeutic Target Database, TTD (2020 version)⁴⁶, was used to subset the drug–target gene collection, keeping only experimentally characterized drug–gene mappings e.g., those for which an exact pharmacological action of the drug on the gene has been experimentally determined. This data set was referred to as “pharmacologically characterized targets” and consists of 762 drugs. These two drug sets were expanded to include the first neighbors of the drug targets, from FunCoup with a link confidence score $\text{pfc} \geq 0.99$. An additional drug set that mapped each drug with its FDA-approved indications was obtained from TTD⁴¹ and named “FDA indications”. Table 1 summarizes the sizes of all drug data sets.

Disease genes data set. Disease genes were retrieved from DisGeNET v6.0⁴⁷. Only the disease–gene associations reported by the Comparative Toxicogenomics Database, CTD⁴⁸, were kept, as those have been manually curated by the authors of CTD. Conditions classified as phenotypes or disease groups were removed. Diseases with fewer than 20 genes were removed in order to keep only well-characterized diseases, as in previous work¹⁷. Lastly, to correct for the fact that diseases do not have an unambiguous nomenclature in DisGeNET, e.g., cerebral artery atherosclerosis and cerebral atherosclerosis, the remaining diseases were merged under the same name if their Szymkiewicz–Simpson similarity coefficient⁴⁹, calculated as shown in Eq. (1), was equal or greater than 0.95. This threshold was selected because it maximizes the number of disease–gene associations while minimizes the overlap between diseases.

$$\text{overlap}(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)} \quad (1)$$

In addition, the 572 human genes associated with COVID-19 reported by Gordon et al.⁵⁰ and Li et al.⁵¹ were retrieved from IntAct²⁸ and added to the dataset. The final data set contained 13,560 disease–gene associations between 177 diseases and 5766 genes.

Direct and extended target-based drug ranking. For the direct target ranking, drugs were sorted by the number of direct gene targets. In the extended target ranking, direct drug targets were expanded using their first order neighbors retrieved from FunCoup with confidence $\text{pfc} \geq 0.99$. From these, high quality first neighbors were retrieved using MaxLink⁵². Maxlink is a guilt-by-association algorithm that identifies genes tightly linked to a set of query genes using a hypergeometric test to ensure the statistical significance of the association. MaxLink was run independently for each drug, with the drug direct targets and the gene interactions with $\text{pfc} \geq 0.99$ in the human FunCoup v4.1⁴³ network as inputs. The neighbor genes with an Benjamini–Hochberg FDR-corrected⁵³ p-value ≤ 0.05 were returned. The drugs were then sorted by the number of expanded targets.

Disease-based drug ranking. Drugs were mapped to a disease if one or more of its gene targets was part of the disease gene set. All drugs were then sorted based on the number of diseases they had been mapped to. A similar procedure was done for the extended drug targets. The drugs in the “FDA indications” data set were ranked based on number of indications.

Disease coverage by drugs. To analyze how many diseases can be covered by n drugs ($n = 1, \dots, 5$ or the full data set) the direct targets drug set and the “FDA indications” drug set were reduced in an iterative procedure. In each iteration, the drug covering the highest number of diseases in terms of gene-overlap was removed, and the number of diseases was added to a counter. The covered diseases were removed from the diseases-to-target space e.g., diseases not yet covered during the iteration. The procedure was repeated n times. A permutation test was performed to assess the statistical significance of the disease coverage. The test involved performing the drug coverage estimation procedure with randomly sampled gene sets of the same size as the drug target sets, repeated 1000 times.

Finding disease modules. For each disease in the disease data set, the interactions between disease genes were retrieved from FunCoup using the link confidence cutoff of $\text{pfc} \geq 0.80$. Infomap⁵⁴ was used to find modules in the subnetworks, using the pfc scores as edge weights. For each disease, one or more disease modules were retrieved. Subsequently, and following Choobdar et al.⁵⁵, disease modules with fewer than three genes were

removed. The final set contained 503 modules for 157 diseases. The set of disease modules and their corresponding genes is summarized in Sup. Table 6 online.

Disease module-based drug ranking. Disease module-based drug rankings were built for the direct targets and extended targets drug sets. In both cases drugs were mapped to disease modules with which they had at least one overlapping gene. Then, the drugs were sorted based on the number of disease modules they overlap with.

Disease module representations. Cytoscape v3.2.1⁵⁶ and Inkscape v1.0.2 (<https://inkscape.org/>) were used to visualize disease modules and the bipartite drug–disease module networks.

Data availability

All the data that support the findings of this study are available at <https://bitbucket.org/sonnhammergroup/unadrug>.

Code availability

The codes used in this project are available at <https://bitbucket.org/sonnhammergroup/unadrug>.

Received: 9 March 2021; Accepted: 30 September 2021

Published online: 19 October 2021

References

- Deininger, M. W. N. & Druker, B. J. Specific targeted therapy of chronic myelogenous leukemia with imatinib. *Pharmacol. Rev.* **55**, 401–423 (2003).
- Yella, J., Yaddanapudi, S., Wang, Y. & Jegga, A. Changing trends in computational drug repositioning. *Pharmaceuticals* **11**, 57 (2018).
- Nabirotkin, S. *et al.* Next-generation drug repurposing using human genetics and network biology. *Curr. Opin. Pharmacol.* **11**, 1–15 (2019).
- Hopkins, A. L. Network pharmacology. *Nat. Biotechnol.* **25**, 1110–1111 (2007).
- Sexton, P. M. & Christopoulos, A. To bind or not to bind: Unravelling GPCR polypharmacology. *Cell* **172**, 636–638 (2018).
- Pushpakom, S. *et al.* Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2018).
- European Centre for Disease Prevention and Control. COVID-19 situation update worldwide, as of 21 August 2021. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
- Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 1–18 (2020).
- Beigel, J. H. *et al.* Remdesivir for the treatment of Covid-19—final report. *N. Engl. J. Med.* **383**, 1813–1826 (2020).
- Hotez, P. J., Corry, D. B., Strych, U. & Bottazzi, M. E. COVID-19 vaccines: Neutralizing antibodies and the alum advantage. *Nat. Rev. Immunol.* **20**, 399–400 (2020).
- Gysi, D. M. *et al.* Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025581118 (2021).
- FDA. Coronavirus Treatment Acceleration Program (CTAP) | FDA. <https://www.fda.gov/drugs/coronavirus-covid-19-drugs/coronavirus-treatment-acceleration-program-ctap>
- Alexeyenko, A. & Sonnhammer, E. L. L. Global networks of functional coupling in eukaryotes from comprehensive data integration. *Genome Res.* **19**, 1107–1116 (2009).
- Alberghina, L., Höfer, T. & Vanoni, M. Molecular networks and system-level properties. *J. Biotechnol.* **144**, 224–233 (2009).
- Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
- Kitsak, M. *et al.* Tissue specificity of human disease module. *Sci. Rep.* **6**, 1–12 (2016).
- Goh, K. I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8685–8690 (2007).
- Menche, J. *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
- Yue, Z. *et al.* Repositioning drugs by targeting network modules: A Parkinson's disease case study. *BMC Bioinform.* **18**, 532 (2017).
- Guo, X. *et al.* A network pharmacology approach to explore the potential targets underlying the effect of sinomenine on rheumatoid arthritis. *Int. Immunopharmacol.* **80**, 106201 (2020).
- Farha, M. A. & Brown, E. D. Drug repurposing for antimicrobial discovery. *Nat. Microbiol.* **4**, 565–577 (2019).
- Dotolo, S., Marabotti, A., Facchiano, A. & Tagliaferri, R. A review on drug repurposing applicable to COVID-19. *Brief. Bioinform.* **2020**, 1–16 (2020).
- Cheng, F. *et al.* Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* **9**, 1–12 (2018).
- Geahlen, R. L. Getting Syk: Spleen tyrosine kinase as a therapeutic target. *Trends Pharmacol. Sci.* **35**, 414–422 (2014).
- Katsyuba, E. & Auwerx, J. Modulating NAD⁺ metabolism, from bench to bedside. *EMBO J.* **36**, 2670–2683 (2017).
- Zhao, S., Mysler, E. & Moots, R. J. Etanercept for the treatment of rheumatoid arthritis. *Immunotherapy* **10**, 433–445 (2018).
- Inman, B. A., Longo, T. A., Ramalingam, S. & Harrison, M. R. Atezolizumab: A PD-L1-blocking antibody for bladder cancer. *Clin. Cancer Res.* **23**, 1886–1890 (2017).
- Orchard, S. *et al.* The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2013).
- Kost-Alimova, M. *et al.* A high-content screen for mucin-1-reducing compounds identifies fostamatinib as a candidate for rapid repurposing for acute lung injury. *Cell Rep. Med.* **1**, 100137 (2020).
- Lu, W. *et al.* Elevated MUC1 and MUC5AC mucin protein levels in airway mucus of critical ill COVID-19 patients. *J. Med. Virol.* **93**, 582–584 (2021).
- Strich, J. R. *et al.* Fostamatinib inhibits neutrophils extracellular traps induced by COVID-19 patient plasma: A potential therapeutic. *J. Infect. Dis.* **223**, 981–984 (2021).
- Bye, A. P. *et al.* Aberrant glycosylation of anti-SARS-CoV-2 IgG is a pro-thrombotic stimulus for platelets. *Blood* <https://doi.org/10.1101/2021.03.26.437014> (2021).
- Vergis, N. *et al.* Multi-arm Trial of Inflammatory Signal Inhibitors (MATIS) for hospitalised patients with mild or moderate COVID-19 pneumonia: A structured summary of a study protocol for a randomised controlled trial. *Trials* **22**, 1–4 (2021).
- Medina-Franco, J. L., Giulianotti, M. A., Welmaker, G. S. & Houghten, R. A. Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov. Today* **18**, 495–501 (2013).
- Hughes, J. P. *et al.* Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).

36. Ghiassian, S. D., Menche, J. & Barabási, A. L. 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**, e1004120 (2015).
37. Wang, R. S. & Loscalzo, J. Network-based disease module discovery by a novel seed connector algorithm with pathobiological implications. *J. Mol. Biol.* **430**, 2939–2950 (2018).
38. Piñero, J., Berenstein, A., Gonzalez-Perez, A., Chernomoretz, A. & Furlong, L. I. Uncovering disease mechanisms through network biology in the era of Next Generation Sequencing. *Sci. Rep.* **6**, 24570 (2016).
39. Liu, Y. *et al.* Network-assisted analysis of GWAS data identifies a functionally-relevant gene module for childhood-onset asthma. *Sci. Rep.* **7**, 1–10 (2017).
40. Sharma, A. *et al.* A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum. Mol. Genet.* **24**, 3005–3020 (2014).
41. Zaman, N. *et al.* Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep.* **5**, 216–223 (2013).
42. Guney, E., Menche, J., Vidal, M. & Barabási, A. L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 1–13 (2016).
43. Ogris, C., Guala, D., Kaduk, M. & Sonnhammer, E. L. L. FunCoup 4: New species, data, and visualization. *Nucleic Acids Res.* **46**, 601–607 (2017).
44. Sonnhammer, E. L. L. & Ostlund, G. InParanoid 8: Orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**, D234–D239 (2015).
45. Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
46. Wang, Y. *et al.* Therapeutic target database 2020: Enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* **48**, D1031–D1041 (2020).
47. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855 (2019).
48. Davis, A. P. *et al.* The comparative toxicogenomics database: Update 2019. *Nucleic Acids Res.* **47**, D948–D954 (2019).
49. Vijaymeena, M. K. & Kavitha, K. A survey on similarity measures in text mining. *Mach. Learn. Appl. Int. J.* **3**, 19–28 (2016).
50. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
51. Li, J. *et al.* Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med.* **2**, 99–112 (2021).
52. Guala, D., Sjölund, E. & Sonnhammer, E. L. L. MaxLink: Network-based prioritization of genes tightly linked to a disease seed set. *Bioinformatics* **30**, 2689–2690 (2014).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
54. Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *Eur. Phys. J. Spec. Top.* **178**, 13–23 (2009).
55. Choobdar, S. *et al.* Assessment of network module identification across complex diseases. *Nat. Methods* **16**, 843–852 (2019).
56. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Acknowledgements

The project that gave rise to these results received support for a fellowship from Fundación Margit y Folke Pehrzon.

Author contributions

E.S., D.G. and M.C. conceived of the presented idea and supervised the project. I.R. and L.G. retrieved and pre-processed the data. I.R. performed the data analysis. E.S., D.G., I.R. and M.C. wrote the manuscript. All authors read and approved the final manuscript.

Funding

Open access funding provided by Stockholm University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99721-y>.

Correspondence and requests for materials should be addressed to E.L.L.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021