Mendelian disease comorbidity as a resource for novel therapeutic targets

Abstract

Drugs targeting disease causal genes are more likely to succeed for that disease. But complex disease causal genes are not always clear. In contrast, Mendelian disease causal genes are well-known and druggable. In previous work, we examined 90 Mendelian and 65 complex diseases, finding 2,908 pairs of clinically associated (comorbid) diseases in which the Mendelian disease causal genes may have a pleiotropic effect on the complex disease. Here, we hypothesize that Mendelian disease causal genes can be potential therapeutic targets for the comorbid complex disease due to their pleiotropic effects. We match each complex disease to a set of Mendelian disease causal genes based on its comorbidity and suggest the drugs targeting these genes as potential drug candidates for the complex disease. We demonstrate that the candidate drugs are enriched among currently investigated or indicated drugs for the relevant complex diseases (OR=1.8, p=1.79e-21). We also present evidence that certain Mendelian diseases have a better potential for future drug discovery and drug repurposing applications. In conclusion, our findings suggest a novel way to take advantage of rich knowledge about Mendelian disease biology to improve treatments of common diseases.

Introduction

Traditional drug development pipeline is costly and slow. It is estimated that around $2.6 billion is spent and approximately 12-15 years are required for just one new drug to reach the market[1,2]. Additionally, clinical trial success rates remain low[3]. This can be attributed to the inability of tested drugs to demonstrate safety or efficacy[4], mainly due to unfavorable drug properties (e.g. drug pharmacokinetics) and inadequate target identification in the preclinical studies[5]. Therefore, new approaches are needed to predict which drugs will succeed. Some of the tested drugs may already be in common use for other disease and finding new uses for these drugs (known as drug repurposing) can significantly reduce the time and cost associated with the traditional drug discovery and accelerate clinical advances[7].

During the past years, genetics has significantly contributed to predicting drug success. Genome-wide association studies (GWAS) have identified plenty of genetic variants associated with complex diseases that were tested as therapeutic drug targets[8]. For instance, mutations in the *IL23R* locus have been associated with Crohn's disease[9]. Ustekinumab, a monoclonal antibody originally approved for the treatment of psoriasis, was known to target the *IL23* p-40 subunit[10]. Hence, ustekinumab was tested in clinical trials and it is now approved for the treatment of moderate to severe Cronh's disease[11–14]. This example highlights the importance of genetics both in target prioritization and drug discovery.

Nelson et al. analyzed historical data of clinical trials and found that drug-target pairs supported by human genetic evidence for a relevant disease are twice as likely to succeed in clinical trials and get approved compared to those without[15]. Four years later, King et al.

validated these findings by analyzing data not available at the time of the Nelson et al. study, further emphasizing the importance of genetics in the drug development process[16]. Moreover, King et al. found that the success rate of a drug is higher when the disease causal gene is clearly identified, as in the case of monogenic (Mendelian) diseases. However, identifying the disease causal genes from GWAS can be challenging as the majority of the GWAS hits are in non-coding regions[17]. In contrast, Mendelian disease causal genes are both well-known and druggable[18]. For example, Androgen Insensitivity Syndrome is known to be caused by mutations in the androgen receptor gene (*AR*) and a total of 83 drugs are known to target *AR*[19,20]. Then, developing a way to translate knowledge about Mendelian disease biology to complex diseases could significantly impact drug discovery.

Blair et al. analyzed health records of patients with Mendelian diseases to identify to which complex diseases they are predisposed[21], finding clinically associated (comorbidity) pairs of 95 Mendelian and 65 complex diseases[21]. That study also showed evidence that comorbidity can be tied to a common genetic basis[21]. In a follow-up study focusing on cancers, Melamed et al. showed that Mendelian disease causal genes are likely to be frequently mutated in comorbid cancers[22]. For instance, patients with Rubinstein Taybi syndrome, a Mendelian disease caused by mutations in *CREBBP*[23], are predisposed to lymphoma[24]. CREBBP is one of the most frequently inactivated genes in lymphoma[25], indicating that the comorbidity can be attributed to pleiotropic effect of CREBBP mutation causing Rubinstein Taybi and contributing to the development of lymphoma. Both above studies systematically demonstrate that Mendelian disease comorbidity can suggest a role of Mendelian causal genes on complex diseases.

Building on these findings, we hypothesize that if the Mendelian disease causal genes contribute to the development of a comorbid complex disease, then these genes can be novel therapeutic targets for that disease. To evaluate our hypothesis, we test the enrichment of drugs targeting these genes among currently investigated or indicated drugs for that disease (***Figure 1A***).

Results

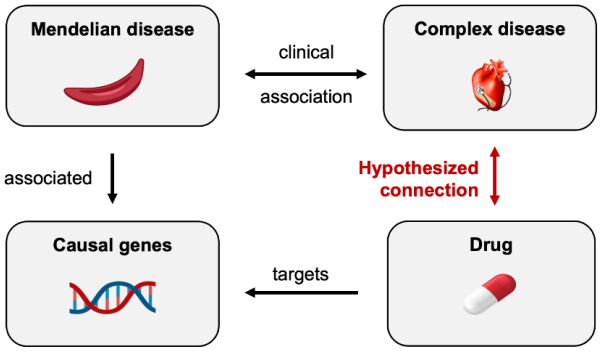
Integrating data to test Mendelian diseases as a resource for drug repurposing candidates

From Blair et al.[21], we use 2,908 Mendelian and complex disease pairs of Mendelian diseases clinically associated with increased risk of complex disease (referred to as *comorbidity pairs* from now on). The results include 90 Mendelian diseases with known causal genes and 65 complex diseases across six disease categories (cardiovascular, hormonal, immune, neoplasms, neurological, ophthalmological). ***Figure 1B*** shows the distribution of the number of complex disease comorbidities per Mendelian disease in our sample. Using drug-gene target information from DrugBank (https://go.drugbank.com/) and Mendelian causal genes from Online Mendelian Inheritance in Man (OMIM), we compile a list of 788 drugs that target the Mendelian disease causal genes as repurposing candidates for the relevant comorbid complex diseases. These are candidate effective drugs for each complex disease based on its comorbid Mendelian diseases (***Figure 1A***).

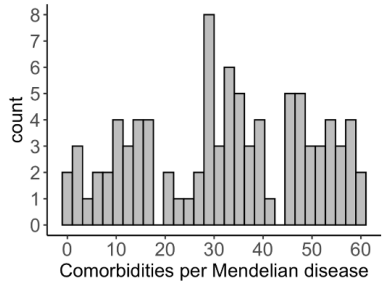
To test our hypothesis, we compare candidate drugs against currently used or investigated drugs. We curate information for 29,758 clinical trials that investigate 1,795 drugs for 64 complex diseases (median of 110 investigated drugs per complex disease). In addition to the investigated drugs, we compile a data set of 1,373 approved drugs indicated for 64 complex diseases (median of 42 indicated drugs per complex disease) (***Figure 1C***).

Figure 1. Outline of the approach

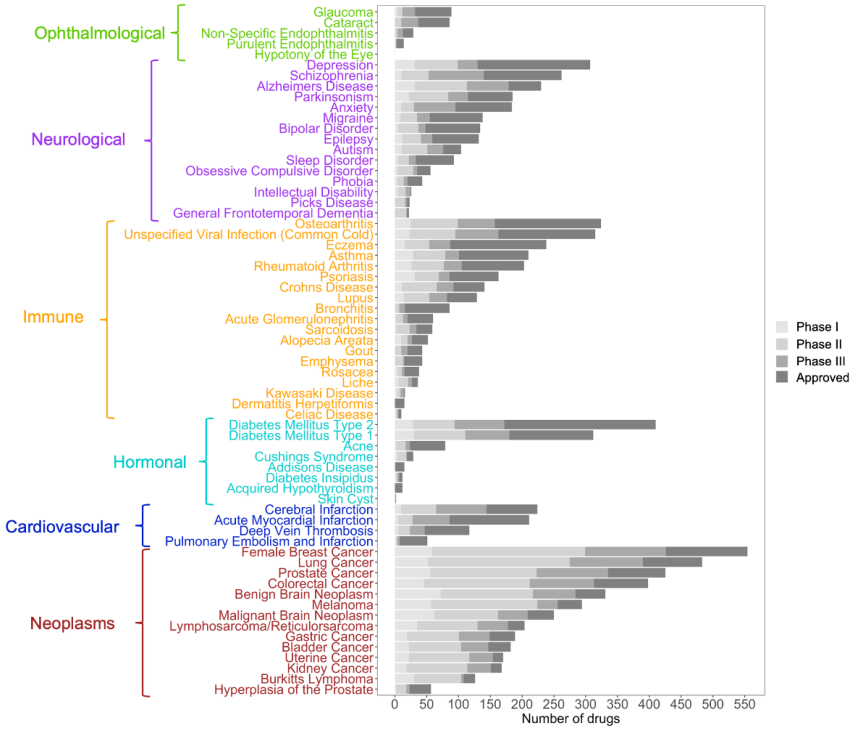
A. Proposed methodology.



B. Distribution of the number of complex disease comorbidities per Mendelian disease



C. Number of investigated (per clinical trial phase) and indicated drugs per complex disease.



Mendelian disease comorbidity identifies drugs under current investigation or indication

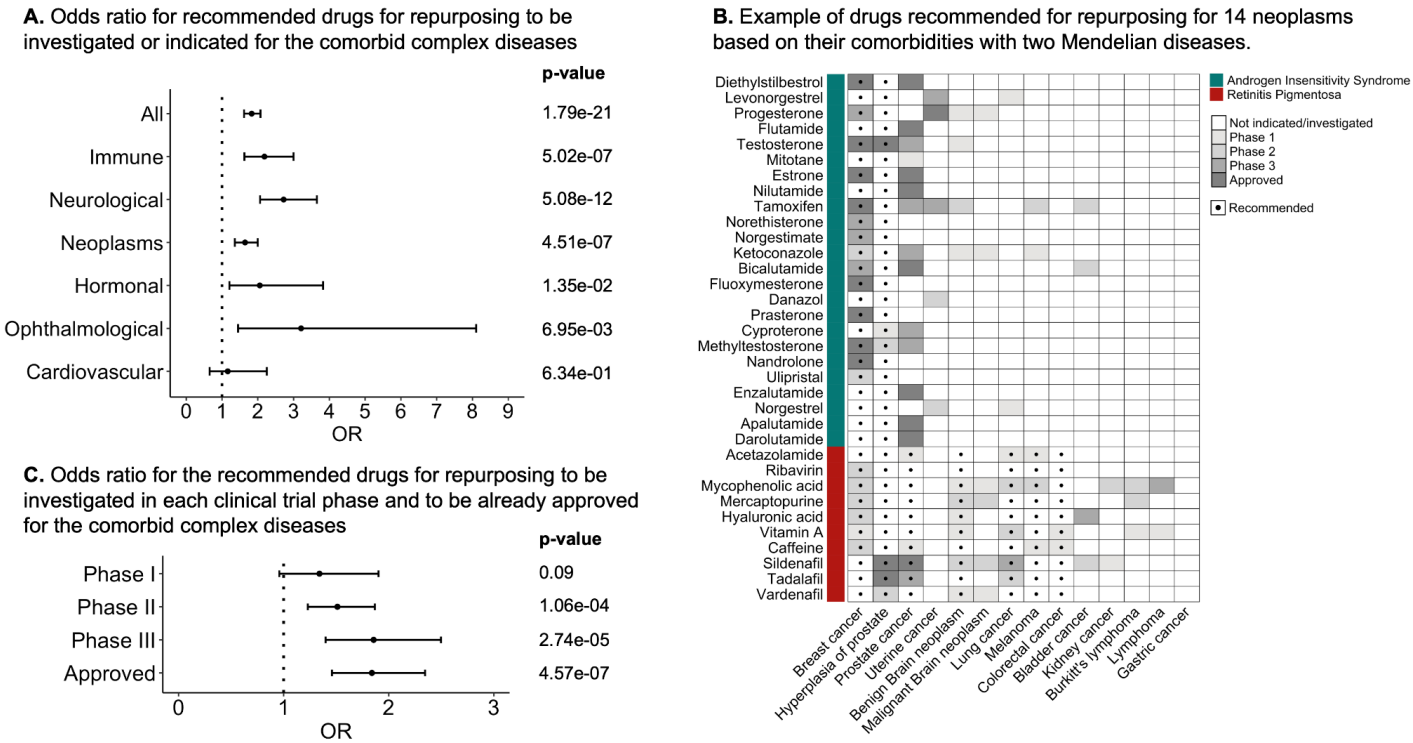
First, we assess whether the candidate drugs are enriched for those currently investigated or indicated for the comorbid complex diseases. Accounting for the number of gene targets per drug, and the category of disease, we find that the candidate drugs are significantly enriched for drugs currently investigated or indicated (OR=1.8, p = 1.79e-21). Next, we seek to exclude artifactual explanations for this signal. Because some Mendelian diseases are more rare than others, there could be a spurious association between the ability to discover disease genes and the ability to discover clinical associations. Therefore, we randomly permute which complex diseases each Mendelian disease is comorbid with. A random permutation preserves the characteristics of each Mendelian disease, such as prevalence and number of complex disease comorbidities, but not the list of candidate drugs for each complex disease. After 1,000 permutations, we find that the observed association is significantly stronger than expected by chance (p_permutation<0.001) (***Supplementary S1A***).

We repeat the analysis at a disease category level, and we find significant results for neurological, immune, neoplasms, ophthalmological, and hormonal diseases (***Figure 2A***).

However, after permutation analyses, only immune, neurological, and neoplasm remain significant (p_permutation < 0.05) (***Supplementary S1A***). This might be explained by the fact that in our sample, cardiovascular, ophthalmological, and hormonal disease categories contain smaller number of complex diseases which might reduce statistical power. For instance, only 4 complex diseases belon to cardiovascular diseases while 19 to immune diseases. ***Figure 2B*** shows an example of drugs recommended for repurposing for 14 neoplasms based on their comorbidities with two Mendelian diseases, illustrating that the candidate drugs extensively overlap with drugs currently investigated or indicated for these neoplasms.

Next, we ask if the candidate drugs are more likely to be in more advanced drug development phases for the relevant complex diseases. First, we find that the candidate drugs are enriched for drugs under investigation in any clinical trial phase (OR=1.58, p=4.8e-09; p_permutation<0.001). Using information about drugs under investigation in either Phase I, II or III (we exclude drugs in clinical trials with unknown phase), we find that increasing clinical trial phase is associated with increasing predictive ability (p_permutation<0.05) (***Figure 2C***, ***Supplementary S1B***). Using information about only indicated drugs for the complex diseases, we find that candidate drugs are enriched for already approved drugs (OR=1.84, p=4.57e-07; p_permutation=0.002) (***Figure 2C***, ***Supplementary S1B***).

Figure 2. Mendelian and complex disease comorbidity can be used for drug repurposing



Mendelian diseases targeted by high number of drugs are a better resource for drug repurposing candidates

Mendelian disease causal genes are known to be good drug targets[18]. However, the total number of drugs targeting each Mendelian disease might differ. In our sample, a median of 4 genes are causally associated per Mendelian disease (min=1; max=51) (***Supplementary S2A***). Additionally, each Mendelian disease is targeted by a median of 5 drugs (min:0;

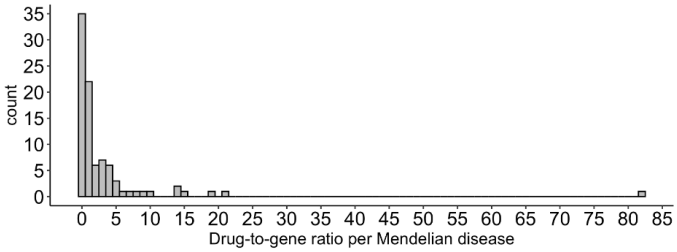
max=117) with a median drug-to-gene ratio of 1 (min=0; max=82) (***Figure 3A, Supplementary S2B***). This observation motivated us to delve deeper into understanding the characteristics of Mendelian diseases that identify currently investigated or indicated drugs and thus, be more informative for future drug discovery efforts. More specifically, we examine the total number of drugs targeting each Mendelian disease genes and the number of comorbidities per Mendelian disease.

First, we repeat our analysis above for each Mendelian disease targeted by at least one drug. That is, we test whether the specific drug candidates recommended by each Mendelian disease are enriched for currently indicated or investigated drugs for its comorbid complex diseases. We find 9 significant Mendelian diseases at a nominal level ($p<0.05$) (***Supplementary S3A***). This was expected as most Mendelian diseases are rare and, undiscovered causal genes and clinical associations may exist which lower their power to predict currently investigated or indicated drugs for their comorbid complex diseases. The significant Mendelian diseases are targeted by a median of 47 drugs per Mendelian disease compared to a median of 8 drugs per non-significant Mendelian disease (p=2e-05, Wilcoxon rank-sum test) (***Figure 3B***). However, two Mendelian diseases targeted by similar number of drugs might be associated with different number of genes. For instance, Long QT syndrome is associated with 8 causal genes targeted by 117 drugs and Androgen Insensitivity Syndrome is associated with 1 causal gene targeted by 82 drugs. For this reason, we also calculate and compare the drug-to-gene ratio between the significant and non-significant Mendelian diseases. Significant Mendelian diseases have a median drug-to-gene ratio of 3.12 (min=0.97, max=82) and non-significant Mendelian diseases have a median drug-to-gene ratio of 1.39 (min=0.14, max=19). This different is also found to be statistically significant (***Supplementary S3B***). Additionally, we test if there is a difference in the number of complex disease comorbidities between the significant and non-significant Mendelian diseases, founding no statistically significant association (***Supplementary S3C***).

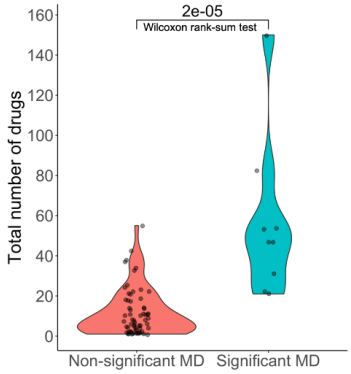
These results imply that Mendelian diseases targeted by a high number of drugs are a good source for novel therapeutic targets for the comorbid complex diseases, but we exclude other obvious explanations for this signal. One obvious explanation is that the high number of drugs, rather than the information about comorbidity, increases the power to detect enrichment of the candidate drugs among currently investigated or indicated drugs for a complex disease. To exclude this possibility, we randomly permute the drugs targeting each Mendelian disease. In this case, a random permutation would change which drugs target each Mendelian disease but keep unchanged the total number of drugs targeting each Mendelian disease. After 1,000 permutations, the observed enrichment was found to be highly statistically significant (p_permutation=0.002) (***Figure 3C***).

Figure 3. Mendelian diseases targeted by high number of drugs are be more useful for drug repurposing opportunities

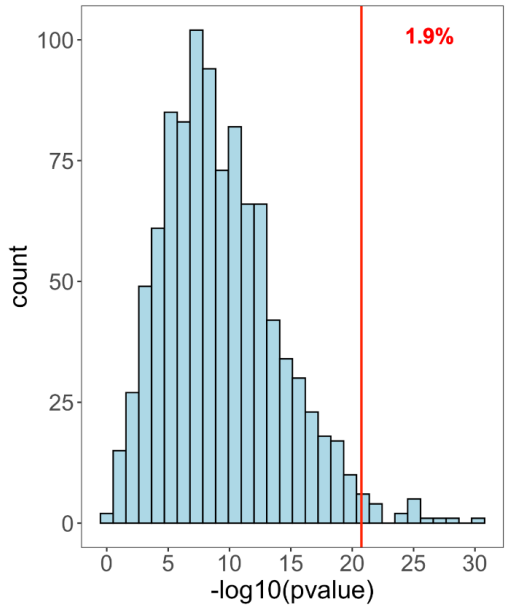
A. Histogram of drug-to-gene ratio per Mendelian disease showing that Mendelian diseases genes are targeted by a wide range of number of drugs.



B. Mendelian diseases that are useful for drug repurposing are targeted by statistically significantly higher number of drugs



C. Random permutation of the drugs targeting each Mendelian disease genes shows that our results are not expected by chance



Comorbidity pairs linked to pleiotropic disease genes enhance identification of drugs under current investigation or indication

Comorbid Mendelian and complex diseases are often attributable to shared genetic mechanisms, as in the example of Rubinstein Taybi syndrome and lymphoma mentioned in the introduction. But, in some cases, a Mendelian disease might predispose patients to a complex disease by indirectly interacting with its associated risk factors. For instance, cystic fibrosis is caused by mutations in the *CFTR* gene[26], resulting in mucus accumulation within the airways[27]. This, in turn, traps germs and increases the risk of patients with cystic fibrosis to lung-related diseases such as bronchitis[27]. These two examples demonstrate how a Mendelian disease can predispose patients to a complex disease either due to direct pleiotropic effects of disease genes or by affecting risk factors associated with the complex disease. Under our hypothesis, comorbidity relationships that also have evidence of genetic similarity are more likely to be enriched for success.

To test this, we focus on neoplasms, one of the disease categories with the strongest association in our analysis (***Figure 2A***). We build on the work of Melamed et al[22] who investigated the extent of genetic overlap (genetic similarity) of 314 comorbidity pairs of 60 Mendelian diseases and 10 cancers. First, we find that drugs targeting the Mendelian disease causal genes are enriched among currently investigated or indicated drugs for the comorbid cancers (OR=1.68, p=9.19e-06; p_permutation=0.004). Having established this association, we focus on the subset of 131 Mendelian and cancer comorbidity pairs with genetic similarity and find that the significance increases (OR: 1.97; p=2.63e-09;

p_permutation=0.006) (***Supplementary S4***). These findings indicate that comorbidity pairs known to be linked to pleiotropic genes are a better resource of drug repurposing candidates.

Discussion

We have shown that drugs targeting Mendelian disease causal genes are enriched for currently investigated or indicated drugs for its comorbid complex diseases. We believe that this finding has two interpretations. First, it suggests that Mendelian disease causal genes can serve as a valuable resource for identifying potential drug repurposing candidates for comorbid complex diseases, due to the pleiotropic effects of the Mendelian genes. Second, it suggests that future drug design efforts for a complex disease can benefit from considering its comorbid Mendelian disease genes as plausible novel therapeutic targets. It is known that incorrect identification of drug targets in preclinical studies is a major reason for the high attrition rates observed in clinical trials[5]. However, Mendelian diseases are a valuable source of well-known and well characterized genes, which makes them promising drug target candidates.

We have also found that Mendelian diseases targeted by a high number of drugs have greater power to predict currently investigated or indicated drugs for its comorbid complex diseases. We suggest that these Mendelian diseases might be useful for increasing the success of future drug discovery and drug repurposing efforts for its comorbid complex diseases. We believe that this happens because the genes associated with them are better studied and characterized, something that will increase the confidence of target identification and validation in preclinical studies. Moreover, these genes are known to be targeted by existing drugs, meaning that they have been successfully used in the past for drug discovery. However, future studies are needed to further understand the observed pattern.

In addition, we have found that comorbidity pairs of Mendelian diseases and cancers with genetic similarity have enhanced ability to identify currently investigated or indicated drugs. This further supports our hypothesis that Mendelian disease genes that contribute to the development of a comorbid complex disease, as in the example of Rubinstein Taybi Syndrome, can be plausible drug targets for that complex disease. Although we have shown that comorbidity is already a useful signal for both novel target identification and drug repurposing, we suggest that future drug design efforts will be benefited from compiling genetics from diseases besides cancer to identify comorbidity pairs with genetic similarity that might further enhance opportunities of success.

Our work has some limitations. First, we were not able to capture all the clinical trials that tested drugs for the complex diseases in our sample. An example is the drug darolutamide, which is known to be tested in clinical trials for female breast cancer (ClinicalTrials.gov identifier: NCT03004534). This happened because conditions and interventions of many clinical trials did not have associated MeSH terms. Nevertheless, we analyzed a high number of clinical trials which allowed the statistical evaluation of our hypothesis. Second, we were not able to group all forms of the same ingredient to one drug (e.g., testosterone sulfate and testosterone acetate as testosterone). This is because DrugBank assigns them separate IDs and does not provide a relational dataset. However, we do not believe this negatively affected our results as we expect both ingredients to have similar targets and properties and, consequently, be under investigation or indicated for similar complex diseases. Third, we could only test well characterized drugs (with known

targets), meaning that newly investigated drugs with unknown drug targets were not considered. Fourth, although we carefully matched ICD-10 codes of complex diseases to MeSH codes, there is always a possibility that we did not include relevant MeSH codes.

To our knowledge, this is the first study that suggests Mendelian disease comorbidity as a novel way to improve treatments of common diseases. We trust that our findings can be a foundation for future, evidence-based drug design and drug repurposing efforts.

Materials and methods

Data sources

Mendelian and complex disease comorbidities

We download 3,106 Mendelian and complex disease comorbidity pairs between 95 Mendelian and 65 complex diseases from the Blair et al. supplementary "Table S4" (available online)[21]. Our goal is to find drugs targeting Mendelian disease causal genes and recommend them as repurposing candidates for the comorbid complex diseases. Therefore, we remove 5 Mendelian diseases and their comorbidities that are due to chromosomal abnormalities and the causal gene is not obvious (Down Syndrome, Edward Syndrome, Klinefelter Syndrome, Patau Syndrome, Turner Syndrome). We use the remaining 2,908 comorbidity pairs between 90 Mendelian diseases and 65 complex diseases in all downstream analyses, unless mentioned otherwise. Additionally, we group the 65 complex diseases into 6 major disease categories: cardiovascular (4), hormonal (8), immune (19), neoplasms (14), neurological (15), ophthalmological (5).

We also annotate all the 65 complex diseases with relevant MeSH codes. To do this, we first download the supplementary "Table S2" from Blair et al. (available online) which contains the ICD-10 billings codes the authors used to identify the complex diseases. We then manually match them to MeSH codes. Eventually, we have 230 unique MeSH codes for 65 complex diseases (median of 3 MeSH codes per complex disease; range: 1 to16).

Mendelian and cancer comorbidities with genetic similarity

We download 135 Mendelian and cancer comorbidity pairs between 36 Mendelian diseases and 10 cancers from the Melamed et al. "Supplementary Data S3" (available online)[22]. For each comorbidity pair, this file contains the adjusted p-values (p_{adj}) for each metric of genetic similarity the authors tested. A genetic similarity metric for a comorbidity pair is considered to be statistically significant if the estimated $p_{adj} < 0.1$. We remove 2 Mendelian and cancer comorbidity pairs that are not found in the supplementary "Table S4" of Blair et al. (Retinitis Pigmentosa - Malignant Brain Neoplasm, Specified Anomalies of the Musculoskeletal System - Melanoma). Eventually, 131 Mendelian and cancer comorbidity pairs with at least one significant metric of genetic similarity remained. Next, we create a dataset that contains all the Mendelian disease comorbidities of the above 10 cancers (n=314), using the Blair et al. "Table S4". These two datasets are used in the "Comorbidity pairs linked to pleiotropic disease genes enhance identification of drugs under current investigation or indication" Results section.

Drug-gene target information

We download information for 5,800 drugs from DrugBank (version 5.19, date of download: June 16, 2022, https://www.drugbank.com/). For each drug, we keep its DrugBank ID and its gene targets (provided in HGNC symbol). Drugs contained in this file can be approved, investigational, small molecule, biotech, experimental, nutraceutical, illicit, or withdrawn.

Investigated drugs for the 65 complex diseases

To find drugs currently investigated for the 65 complex diseases in our sample, we download clinical trial data from the Aggregate Content of ClinicalTrials.gov (AACT) database in pipe-delimited format[28]. AACT (https://aact.ctti-clinicaltrials.org/) is a publicly available relational database that contains all information about every study registered in ClinicalTrials.gov. We obtain information for 432,597 clinical trials that were registered in ClinicalTrials.gov by the date of download (November 4, 2022; note that AACT is updated daily). For each clinical trial, we keep the clinical trial ID, clinical trial phase, conditions studied, and interventions tested for them. We filter out clinical trials tested behavioral, device, diagnostic test, dietary supplements, procedures, radiation, or other interventions and do not contain MeSH terms for both conditions and interventions. For the total of 109,430 clinical trials that remain, we match the associated MeSH terms of both conditions and interventions to MeSH codes using the Unified Medical Language System (UMLS) database (https://www.nlm.nih.gov/research/umls/index.html). We group clinical trial phases to Phase I (contains Phase I and Early Phase I), Phase II (contains Phase II and Phase I/Phase II), Phase III (contains Phase III and Phase II/Phase III) or unknown phase (no information provided). Phase IV studies are conducted after a drug gets approved to find long-term benefits and side-effects that could not be discovered in the duration of a clinical trial. Therefore, we consider drugs in Phase IV clinical trials as indicated drugs (see “Drugs indicated for the 65 complex diseases”). We then filter for clinical trials that tested drugs for the 65 complex diseases in our sample (31,053 clinical trials).

For the above clinical trials, we convert the MeSH codes of interventions to DrugBank IDs using the UMLS API (crosswalk function). However, DrugBank does not assign IDs to drug combinations. To include them in our analysis, we convert the MeSH codes that did not match directly to a DrugBank ID to RxNORM CUIs using the UMLS API (crosswalk function). Then, for each drug combination, we obtain the active pharmaceutical ingredients using the UMLS API (“Retrieving Source-Asserted Relations"; relation label=has_part). Although MeSH vocabulary assigns different codes to each form of an active pharmaceutical ingredient, DrugBank assigns IDs only to the general forms. For example, liposomal doxorubicin and doxorubicin are two separate entries in the MeSH vocabulary but in DrugBank they are one entry (doxorubicin). To deal with this discrepancy, we follow the steps described above and we obtain the active pharmaceutical ingredients using the UMLS API (relation label=form_of). Then, we convert the RxNORM CUIs to DrugBank IDs using the UMLS API (crosswalk function).

Eventually, we have 29,758 clinical trials that tested 1,795 drugs (unique DrugBank IDs) for 64 complex diseases in our sample. Note that the complex disease “dermatitis herpetiformis“ does not have any investigated drugs.

Indicated drugs for the 65 complex diseases

To find drugs that are currently indicated for the 65 complex diseases, we download 4,225 approved drugs from DrugBank (version 5.19, date of download: June 16, 2022). We get their indications by combining information from RxNORM (https://www.nlm.nih.gov/research/umls/rxnorm/index.html) and repoDB[29], as described below.

Using the RxNORM API (getClassByRxNormDrugName function), we obtain diseases with a relationship of "may_treat" or "may_prevent" with each approved drug. We then match the diseases to MeSH codes using the UMLS database.

repoDB is a publicly available database that contains drug repositioning successes and failures by integrating data from DrugCentral and ClinicalTrials.gov. We download the full database (last update: 2017) and, for each drug, we keep its DrugBank ID and approved indication(s) after excluding the ones with a note of suspended, terminated, or withdrawn. All indications are coded in UMLS CUIs, so we easily convert them to MeSH codes using the UMLS database.

After combining the data from RxNORM and repoDB, we have 939 unique drugs indicated for 58 complex diseases. We then add to this data set the drugs in clinical trials Phase IV (see "Investigated drugs for the 65 complex diseases") to get a total of 1,373 unique drugs indicated for 64 complex diseases. Note that the complex disease "hypotony of the eye" does not have any indicated drugs.

Suggesting drug repurposing candidates for a complex disease

To suggest drug repurposing candidates for a complex disease, we first find its comorbid Mendelian diseases. We obtain the genes causally associated with the Mendelian diseases from the Blair et al. supplementary "Table S3" derived from OMIM (available online). We then find drugs targeting the Mendelian disease causal genes and we suggest them as candidates for repurposing for the complex disease.

Statistical analysis

Logistic regression to evaluate the overlap of drug repurposing candidates with currently investigated or indicated drugs for a complex disease

We find 788 unique drugs that target the causal genes of the 90 Mendelian diseases. Using these drugs and the 65 complex diseases, we create a data frame where each row is a drug-complex disease pair. Therefore, the number of rows in this data frame is 51,220 (788 drugs multiplied by 65 complex diseases). Besides the drug and complex disease columns, this data frame additionally includes the following information: i) a nominal variable indicating the disease category of each complex disease; ii) a continuous variable of the number of gene targets per drug; iii) a binary indicator of whether a drug is recommended as a repurposing candidate for the complex disease and; iv) a binary indicator of whether a drug is currently investigated or indicated for the complex disease. We then use this data frame and a logistic regression approach to test the enrichment of the drug repurposing candidates for a complex disease for currently investigated or indicated drugs for that complex disease. We also account for the number of targets per drug (continuous) and the category of the complex disease (nominal). We use the following logistic regression model:

$$investigated/indicated \sim recommended + number\ of\ drug\ targets + disease\ category$$

The significance is assessed by the coefficient of the independent variable "recommended" (binary). We account for the number of targets per drug as drugs with a higher number of known targets are more likely to target a Mendelian disease gene. Additionally, we account for the disease category due to differences in the number of drugs investigated or indicated among the 6 disease categories tested in this study. For example, significantly more drugs are tested in clinical trials for neoplasms than ophthalmological (***Figure 1B***).

Permutation tests

To exclude artifactual explanations for the observed associations, we use permutation tests. A permutation test evaluates the statistical significance of an association while controlling for confounding factors and random chance. In our case, a random permutation changes which complex diseases are comorbid with each Mendelian disease, unless otherwise mentioned, by keeping unchanged all the intrinsic Mendelian disease characteristics, such as prevalence and number of comorbidities. This allows us to assess if the observed associations can be attributed to the Mendelian disease comorbidity or not. We perform 1,000 permutations and create a null distribution of p-values by running the logistic regression mentioned above. We then compare the observed p-value to this null distribution. We calculate the probability of observing a p-value at least as extreme as the original p-value by estimating the number of times a permuted p-value is lower or equal to the original p-value ($p_{obs} \geq p_{permutation}$). A result is considered significant if the calculated probability is less than 0.05 (p_permutation<50/1000).

References

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ*. 2016;47:20-33. doi:10.1016/j.jhealeco.2016.01.012
2. Mohs RC, Greig NH. Drug discovery and development: Role of basic biological research. *Alzheimers Dement Transl Res Clin Interv*. 2017;3(4):651-657. doi:10.1016/j.trci.2017.10.005
3. Wong CH, Siah KW, Lo AW. Corrigendum: Estimation of clinical trial success rates and related parameters. *Biostatistics*. 2019;20(2):366-366. doi:10.1093/biostatistics/kxy072
4. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun*. 2018;11:156-164. doi:10.1016/j.conctc.2018.08.001
5. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sin B*. 2022;12(7):3049-3062. doi:10.1016/j.apsb.2022.02.002
6. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. 2004;3(8):673-683. doi:10.1038/nrd1468
7. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18(1):41-58. doi:10.1038/nrd.2018.168
8. Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. *Nat Rev Genet*. 2021;22(10):658-671. doi:10.1038/s41576-021-00387-z
9. Duerr RH, Taylor KD, Brant SR, et al. A Genome-Wide Association Study Identifies *IL23R* as an Inflammatory Bowel Disease Gene. *Science*. 2006;314(5804):1461-1463. doi:10.1126/science.1135245

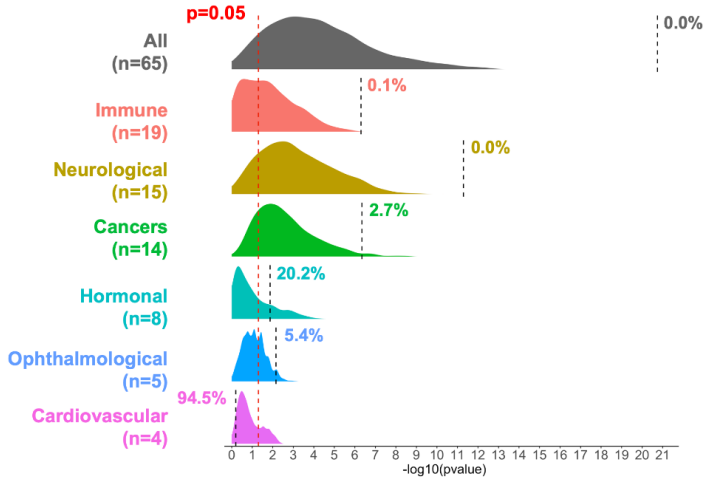
10. DrugBank. Ustekinumab: Usees, Interactions, Mechanism of Action. Accessed July 13, 2023. https://go.drugbank.com/drugs/DB05679
11. European Medicines Agency. Stelara - ustekinumab. Accessed July 13, 2023. https://www.ema.europa.eu/en/medicines/human/EPAR/stelara
12. Feagan BG, Sandborn WJ, Gasink C, et al. Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. *N Engl J Med*. 2016;375(20):1946-1960. doi:10.1056/NEJMoa1602773
13. Johnson-Johnson. FDA Approves STELARA® (Ustekinumab) for Treatment of Adults With Moderately to Severely Active Crohn's Disease. Accessed July 13, 2023. https://www.jnj.com/media-center/press-releases/fda-approves-stelara-ustekinumab-for-treatment-of-adults-with-moderately-to-severely-active-crohns-disease
14. Moschen AR, Tilg H, Raine T. IL-12, IL-23 and IL-17 in IBD: immunobiology and therapeutic targeting. *Nat Rev Gastroenterol Hepatol*. 2019;16(3):185-196. doi:10.1038/s41575-018-0084-8
15. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856-860. doi:10.1038/ng.3314
16. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. Marchini J, ed. *PLOS Genet*. 2019;15(12):e1008489. doi:10.1371/journal.pgen.1008489
17. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19(8):491-504. doi:10.1038/s41576-018-0016-z
18. Brinkman RR, Dubé MP, Rouleau GA, Orr AC, Samuels ME. Human monogenic disorders — a source of novel drug targets. *Nat Rev Genet*. 2006;7(4):249-260. doi:10.1038/nrg1828
19. DrugBank. Androgen receptor drugs. Accessed July 13, 2023. https://go.drugbank.com/bio_entities/BE0000132
20. Hornig NC, Holterhus PM. Molecular basis of androgen insensitivity syndromes. *Mol Cell Endocrinol*. 2021;523:111146. doi:10.1016/j.mce.2020.111146
21. Blair DR, Lyttle CS, Mortensen JM, et al. A Nondegenerate Code of Deleterious Variants in Mendelian Loci Contributes to Complex Disease Risk. *Cell*. 2013;155(1):70-80. doi:10.1016/j.cell.2013.08.030
22. Melamed RD, Emmett KJ, Madubata C, Rzhetsky A, Rabadan R. Genetic similarity between cancers and comorbid Mendelian diseases identifies candidate driver genes. *Nat Commun*. 2015;6(1):7033. doi:10.1038/ncomms8033
23. Bartsch O, Schmidt S, Richter M, et al. DNA sequencing of CREBBP demonstrates mutations in 56% of patients with Rubinstein–Taybi syndrome (RSTS) and in another patient with incomplete RSTS. *Hum Genet*. 2005;117(5):485-493. doi:10.1007/s00439-005-1331-y
24. Miller RW, Rubinstein JH. Tumors in Rubinstein-Taybi syndrome. *Am J Med Genet*. 1995;56(1):112-115. doi:10.1002/ajmg.1320560125
25. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. 2017;171(2):481-494.e15. doi:10.1016/j.cell.2017.09.027
26. Dean M, White MB, Amos J, et al. Multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients. *Cell*. 1990;61(5):863-870. doi:10.1016/0092-8674(90)90196-L
27. Vankeerberghen A, Cuppens H, Cassiman JJ. The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. *J Cyst Fibros*. 2002;1(1):13-29. doi:10.1016/S1569-1993(01)00003-0
28. Tasneem A, Aberle L, Ananth H, et al. The Database for Aggregate Analysis of ClinicalTrials.gov (AACT) and Subsequent Regrouping by Clinical Specialty. Gagnier JJ, ed. *PLoS ONE*. 2012;7(3):e33677. doi:10.1371/journal.pone.0033677
29. Brown AS, Patel CJ. A standard database for drug repositioning. *Sci Data*.

2017;4(1):170029. doi:10.1038/sdata.2017.29

Supplementary

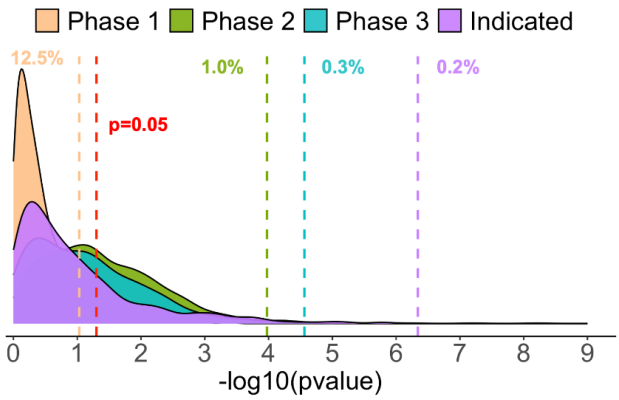
Supplementary S1. Permutation analysis for overall and per disease category.

A. Distribution of the probability the recommended drugs for a complex disease to be indicated or investigated for that complex disease after 1,000 permutations of each Mendelian disease comorbidities (colored area).



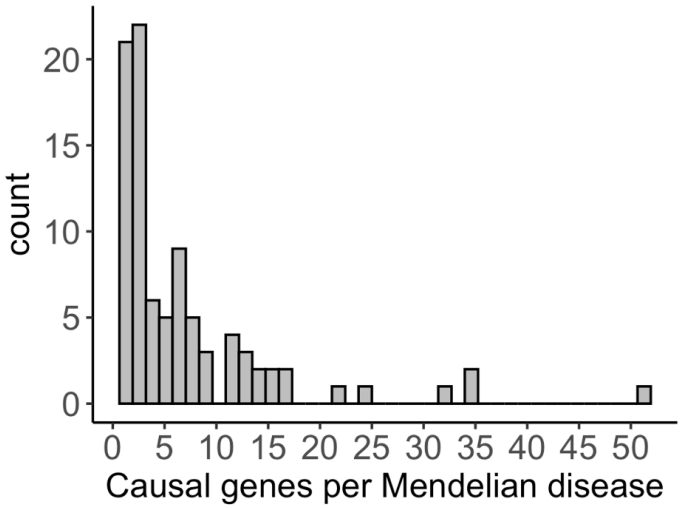
*red, dashed line: -log10(0.05); black, dashed line: -log10(observed_pvalue); n: number of complex diseases within each category

B. Drugs recommended for repurposing for a complex disease are more likely to be already investigated for a complex disease.

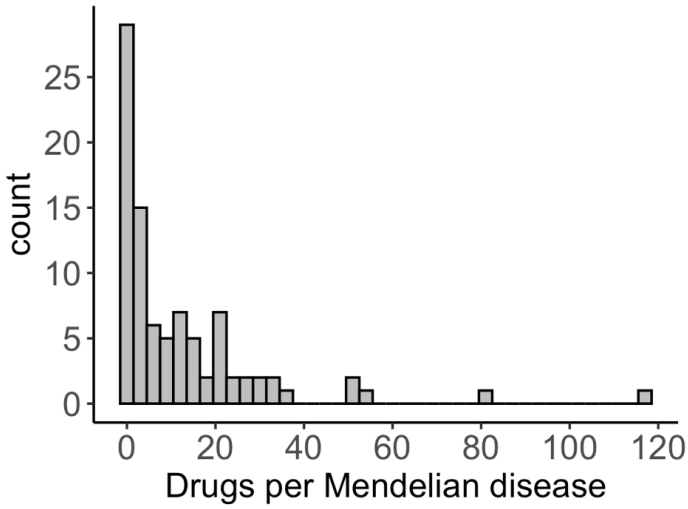


Supplementary S2. Number of causal genes and drugs targeting them per Mendelian disease

A. Distribution of the number of causal genes per Mendelian disease.

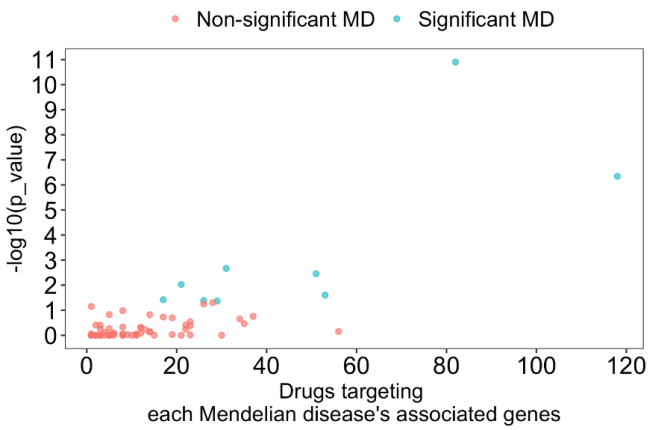


B. Distribution of the total number of drugs targeting each Mendelian disease associated genes.

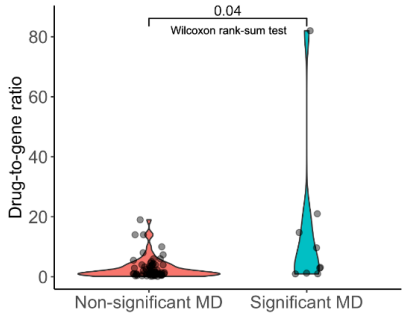


Supplementary S3.

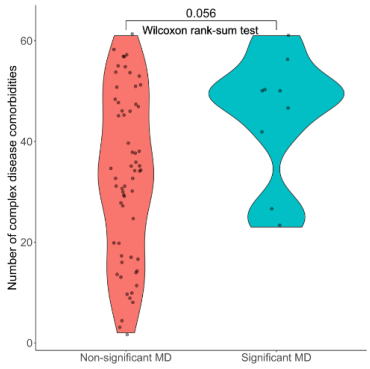
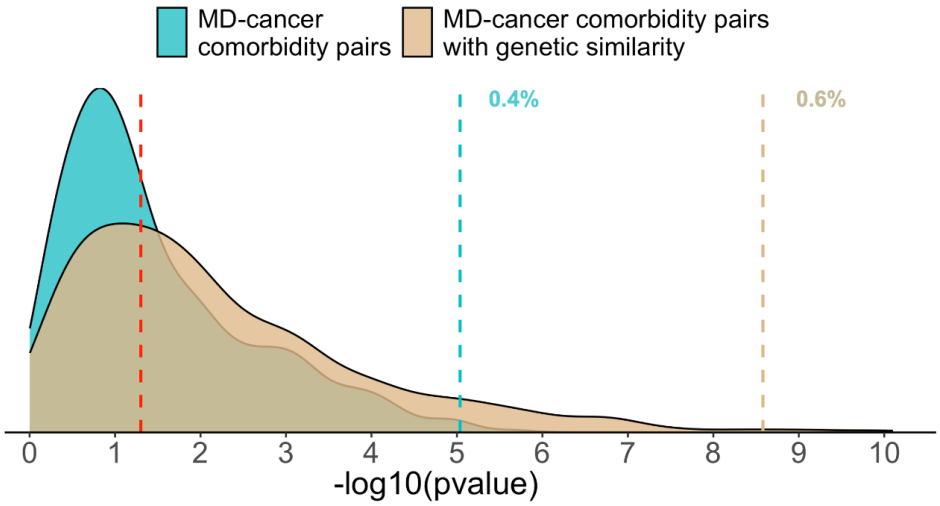
A. Number of drugs targeting each Mendelian disease's associated genes.



B. Mendelian diseases that are useful for drug repurposing have higher drug-to-gene ratio.



C. No statistically significant difference is observed between the number of complex disease comorbidities in the two groups.



Supplementary S4. Distribution of the probability the recommended drugs for a cancer to be indicated/investigated for that cancer after 1,000 permutations of each Mendelian disease comorbidities.