Isaac Madan
March 14, 2014
BMI 258, Final project

The potential of genome-based computational methods for drug repositioning

Drug discovery and development are critical components in today's healthcare ecosystem and are also the lifeblood of the pharmaceutical industry. Globally, brand name drug manufacturing is a market that sees over $160 billion in annual revenue, placing thousands of different types of drugs in pharmacies and drug stores and evaluating exponentially many more as prospective therapies.[1,2]

Unfortunately, despite the clear importance and value of drugs from a biomedical standpoint, the process of bringing a drug to market is both lengthy and costly: it is posited that only 1 in 5,000 drugs will reach consumers, after roughly 15 years of development and evaluation, summing around $1 billion in expenses.[3] Thus, it has been tasked upon pharmaceutical companies and the broader scientific community to identify ways in which we can bring more drugs to market and treat a wider variety of ailments, while lowering the cost and time needed to do so.

Drug repositioning

One proven way to more effectively and cheaply develop novel therapeutics is through drug repositioning—the process of applying known drugs to new indications. In other words, rather than developing a drug from scratch, pharmaceutical companies can apply a drug they have already discovered—and often developed and brought to market—to a new disease, on which the drug has a therapeutic effect.

One seminal example of the value of drug repositioning is the case of sildenafil. Pfizer originally evaluated the drug as a therapeutic for angina—"chest pain or

discomfort that occurs if an area of your heart muscle [does not] get enough oxygen rich blood."[4] However, the drug proved a poor therapeutic for angina. After additional research, Pfizer concluded that sildenafil could in fact serve to treat erectile dysfunction. The branded drug, Viagra, is now widely known and generates over $2 billion in annual revenue for the company.[5] In this way, drug repositioning minimizes the attrition of discovered compounds and saves both time and money for innovators.

Why drug repositioning?

Repositioning is effective for a number of reasons. First, the process of drug development involves pre-clinical trials, clinical trials, and regulation—a lengthy and complicated series of events that evaluates the toxicity and efficacy of the drug in question. Vast majorities of failed therapeutics are unsuccessful during this development process, so through repositioning, drug developers can see greater success rates and approval speeds by considering drugs that have already been brought to market or clinically evaluated previously.[6]

Second, traditional drug development if often dependent on the discovery of both a target, on which a drug can act to ail disease, and a compound, which can actually act upon the target. Through drug repositioning, we effectively limit ourselves to one half of this problem—we are working with compounds that have already been identified. This approach minimizes time spent discovering novel compounds—which often times, is a fruitless endeavor—enabling us to reuse ones we already have available to us.

Moreover, it is important to note that many rare diseases currently do not have suitable drug treatments, largely because the process of discovering and developing drugs for these diseases is often not economically viable for pharmaceutical companies.

Through repositioning, drug developers can significantly reduce costs, and in turn, treat a broader array of diseases than previously possible.

Repositioning techniques

Traditional approaches to drug repositioning include *in vitro* and *in vivo* methods that leverage cell culture and animal models. Alternatively, with the rapid acceleration in the field of bioinformatics, accompanied by a deluge of drug and disease data, much of it publicly available, computational techniques are growing in scope and popularity, so we will focus on these throughout the remainder of this paper.

Computational drug repositioning strategies are numerous, and Dudley *et al.* suggest that strategies can be broken down and categorized across two axes: drug-based methods, on the one hand, and disease-based methods on the other.[7] Drug-based methods involve examining pre-existing drugs at chemical and molecular levels and determining similarities in these profiles. For example, a drug that is similar to another drug in the way it molecularly binds to a protein implicated in a disease, presumably may be effectively applied to other diseases that involve that protein. Disease-based methods involve examining the indications that known drugs treat. For example, if Disease A exhibits the same symptoms as Disease B, or their molecular pathologies are similar, a drug known to treat Disease A may also serve as a promising therapeutic for Disease B.

Drug-based methods

More specifically, drug-based methods evaluate therapeutic compounds on the basis of their chemistry, molecular action, and structure. First, similarity in isolated chemical structure can help explain the biological activity of a compound or the types of

targets it may affect. Thus, compounds similar in chemical structure may exhibit similar therapeutic effects.

Second, considering compounds in the context of a biological environment allows us to better understand the biological changes a drug enables. These changes summarize a compound's "mechanism of action," and in other words, help us understand, at a biological level, how a drug affects disease pathology. Thus, drugs with similar mechanisms of action may affect a disease in a similar way.

Third, we can analyze how chemical activity and mechanism of action come together in the way in which compounds physically bind to and interact with targets. Compounds provide therapeutic effects by binding to sites that are implicated in disease pathogenesis, so similar binding profiles may suggest similar therapies. In all, similarity across any of these three aforementioned characteristics may yield relationships between drugs and bring candidate repositioning compounds to light.

Disease-based methods

Disease-based methods involve evaluating diseases on the basis of their similarity, their molecular activity or behavior, and their phenotypes in relation to drug side effects.

First, disease similarity can be distinguished and defined in a number of ways. Most obvious would be to compare diseases on the basis of their phenotypes, while another option is to examine the spectrum of drugs that service the disease and compare that to the drug spectrum of other diseases. As an example, if multiple drugs are effective across two diseases, it may be that these diseases can be deemed similar,

and thus additional drugs verified to treat one disease may successfully be repositioned to treat the other.

Second, disease pathogenesis can often be determined down to the molecular level. In other words, there may be some molecular activity or dysfunction from which a disease results. If we can successfully identify overlaps in these molecular activity profiles across diseases, we may be able to conclude that two diseases are in fact similar and may be affected similarly by the same therapeutics at a molecular level, even if their phenotypes are wildly different.

Third, drugs often exhibit side effects, and in some cases, these side effects may be congruent with the symptoms of certain diseases. If this is the case, perhaps there is similarity in the molecular activity or pathology across this subset of drugs and diseases, potentially exposing repositioning candidates, or at the very least, enhancing our understanding either of a drug's mechanism of action or a disease's molecular pathology.

In some sense, this side-effect driven approach can also be considered a drug-based method. For example, in 2012, Bisgen *et al.* hypothesized that drugs that exhibit "similar side effect profiles are likely to be effective for the same disease." In their study, they made use of public FDA drug labels—which house side effect information for every FDA-approved drug on the market—categorizing information into topics, through a process known as topic modeling. Their results suggest that they were generally able to use topic modeling of side effect information to predict drugs that share an indication, at a degree significantly better than predicting by chance.[8]

Leveraging networks and databases

Realistically, a drug repositioning strategy can employ both drug-based and disease-based methods to successfully identify repositioned drugs candidates. Thus, we are now seeing the rise of network-based techniques and both public and private databases that enable data access and help illuminate relationships across drugs, indications, and diseases.

As an example, in 2012, Daminelli *et al.* demonstrated that a network relating prominent drugs, drug targets, and diseases helps reveal novel repositioning candidates. In their case, they focused on a specific network motif known as a bi-clique—"a subnetwork in which every drug is linked to every target and disease." By identifying incomplete bi-cliques, the researchers were able to predict how a drug may be repositioned towards indications it does not already relate to within the network. Likewise, this approach is equally useful in identifying drug targets that may be therapeutically affected by existing drugs within the network.[9] In this way, this network enables researchers to draw direct ties between drugs and diseases.

The PROMISCUOUS database project developed by von Eichborn *et al.* in 2011, demonstrates another network-based approaches that synergizes both drug and disease characteristics. The database houses information on many thousands of drugs, each entry annotated with known proteomic interactions in which the drug is involved—better outlining each drug's molecular mechanism of action. The intention is that this database can then be enriched or compared alongside other publicly available drug metadata, such as the FDA's side effect profiles. In this way, PROMISCUOUS may effectively serve as a base platform from which we can begin a network-based drug

repositioning approach.[10] More specifically, the platform enables researchers to more easily compare molecular activity across drugs and also correlate these activities with those of molecular disease pathologies by merging with external datasets.

In 2008, Gunther *et al.* paid special attention to an aforementioned drug-based approach—examining the "molecular basis of drug action." To aggregate pertinent information about drug's molecular action, the researchers combed through biological literature and enriched findings with drug-related metadata, such as related Gene Ontology terms, side effects, and molecular pathways. In doing so, the team released SuperTarget and Matador, two resources that sought to enable better analysis of the molecular actions of therapeutics. Such databases that aggregate diverse sources of information will become increasingly essential as we deal with growing amounts of disparate data.[11]

Leveraging computational techniques

Developing novel methods for drug repositioning requires not only access to data via networks and databases, but also the ability to leverage computational bioinformatics methods and integrate these methods together to garner new insight.

First, novel high throughput computational techniques allow for the rapid analysis of compounds and targets at a molecular level. Namely, one recent study leveraged public microarray data—measuring expression of thousands of genes—pertaining to 100 diseases and 164 drug compounds.[12] They determined up and down regulated genes—collectively called a gene expression signature—in each of these datasets by employing a statistical technique known as statistical analysis of microarrays (SAM) developed by Tusher *et al.* in 2001.[13] The researchers were then able to compare drug-

based signatures and disease-based signatures by leveraging the Broad Institute's Connectivity Map, which "links gene patterns associated with disease to corresponding patterns produced by drug candidates and a variety of genetic manipulations."[14] They then performed comparisons of gene expression profiles for every drug-disease combination and surfaced the cases in which there was most overlap in up-regulated disease genes and down-regulated drug genes, or vice versa. They validated that one of their predictions involving cimetidine, an antiulcer drug, was effective as a "therapeutic in the treatment of lung adenocarcinoma" in animal models, suggesting this repositioning technique is effective.[15]

In another highly similar study, Wu *et al.* enriched up- and down-regulated genes with gene set information that provided insight on biological function. In this way, researchers were able to generate "biological process perturbation profiles," which described how a drug might act at a molecular level, and associate these findings with the same analysis of disease profiles. In particular, this method is an effective means of not only identifying drug repositioning candidates, but also better understanding a drug's mechanism of action, which is unknown or poorly understood in many cases.[16]

Second, another genetic approach that enables a higher level of granularity that may prove promising in drug repositioning is the genome-wide association study (GWAS). These GWA studies have grown in popularity over the past six years, amassing millions of data points corresponding to individual single-nucleotide polymorphisms (SNPs). In 2012, Sanseau *et al.* sought to apply GWAS to drug repositioning by first aggregating all the SNPs associated with disease traits, and filtering this list down to roughly 155 genes, all of which had also been associated with

corresponding drug projects. Of these 155 genes, the researchers determined that 63 of the genes were associated with traits that matched its corresponding drug indication in other studies, therefore verifying the relationship between drug and indication. Next, 92 genes had GWAS traits that differed from its corresponding drug indication, suggesting that this trait may represent a novel indication for the drug—thus potentially serving as a valid repositioning. The remaining genes were associated with GWAS traits that were not disease indications.[17] So, this study serves to demonstrate that GWAS studies in which we can understand disease on a per-SNP basis can yield novel indications for drugs.

Third, novel algorithms development will enable us to more effectively identify similarities in drug compounds and disease targets. Specifically, an area of great interest is binding site homology. In other words, drug compounds are responsible for binding to particular targets—if these targets have structural similarity across diseases, they may be responsive to the same types of drugs. An initial analysis of binding site homology involves a geometrical comparison, observing the actual 3D physical structure of the sites in question. Beyond this, local structural alignment algorithms can be employed to iteratively—at the most basic level—search for local homologies within binding site structures. Alignment algorithms are a subject of great complexity and rapid advancement and thus are beyond the scope of this paper, though their development is imperative to progress in this aspect of drug repositioning. Alongside algorithms are scoring functions that determine the quality of an alignment—modifying the scoring parameters affects the quality of a match and is thus an imperfect and iterative procedure that will help lead us to more accurate results.

Fourth, beyond these biology-specific approaches is one that is more logistical in nature. In the field of biomedicine, results are often scattered and embedded within scientific papers. This literature spans decades and often yields contradictory, and occasionally false, insights. Thus, leveraging this information in a systematic, high throughput way is often challenging and serves as a significant bottleneck to advances in drug repositioning. Thus, a computational technique of increasing importance is text mining and natural language processing. In doing so, we can begin to parse literature systematically, enabling us "to identify targets, to extract drug-disease, drug–target relationships and activity information of drugs." These computational techniques can be paired with biomedical ontologies—and be used to develop novel ontologies—such that we can define our findings in a way that is easily consumable and extensible now and in the future.[18]

Limitations

As we have seen, there exist a variety of novel computational methods and knowledge sharing mechanisms that will propel our ability to reposition drugs in the future. However, these methods are hardly perfect, and there are a number of current limitations that inhibit many techniques from going beyond proof-of-concept and into industry.

Namely, with respect to 3D modeling and chemical structure, current structures are erroneous or incomplete at times; this makes it difficult to rely on structure as a robust and reliable tool. Moreover, biological perturbations often occur *in vivo*, and thus examining a structure beyond the context of its activity in a cell or organism may not yield the results we would expect. And finally, our concrete understanding of a drug is

often limited to what we observe of it during pre-clinical or clinical trials. Getting a full snapshot of a drug's side effect profile requires understanding and synthesizing not only published medical literature, but also clinical data—such as electronic patient records that may detail a novel side effect or drug response, or even a valid repositioning opportunity.

Conclusion

Computational drug repositioning is novel means to rapidly bring therapeutics for a broader range of ailments to market. As we continue to explore and develop these techniques in academia, these strategies will continue to improve in efficacy and efficiency, enabling wider adoption in industry. While these techniques have the disruptive potential to reinvent the primary mechanism of innovation within the pharmaceutical industry, it is important to remember the ultimate goal—to improve healthcare, to treat more diseases, and to treat them better. It has yet to be determined whether computational approaches will align us with this goal, but through additional studies and review, such as this one, we can grow closer to finding out.

References

[1] http://clients1.ibisworld.com/reports/us/industry/ataglance.aspx?entid=487
[2] http://www.ncbi.nlm.nih.gov/pubmed/21690101
[3] http://www.medicinenet.com/script/main/art.asp?articlekey=9877
[4] https://www.nhlbi.nih.gov/health/health-topics/topics/angina/#
[5] http://www.theguardian.com/business/2012/nov/27/new-uses-old-drugs-business
[6] http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/6466/Cost.pdf%3Fsequence%3D1
[7] http://bib.oxfordjournals.org/content/12/4/303.full
[8] http://www.ncbi.nlm.nih.gov/pubmed/23046522
[9] http://www.ncbi.nlm.nih.gov/pubmed/22538435
[10] http://nar.oxfordjournals.org/content/39/suppl_1/D1060.long
[11] http://www.ncbi.nlm.nih.gov/pubmed/17942422
[12] http://stm.sciencemag.org/content/3/96/96ra77.long
[13] http://www.pnas.org/content/98/9/5116.abstract?ijkey=60f9b7b0bbcc8ebb87eb41b00328089c4d92e864&keytype2=tf_ipsecsha
[14] http://www.broadinstitute.org/genome_bio/connectivitymap.html
[15] http://stm.sciencemag.org/content/3/96/96ra77.long
[16] http://www.ncbi.nlm.nih.gov/pubmed/24067419
[17] http://www.nature.com/nbt/journal/v30/n4/abs/nbt.2151.html
[18] http://bib.oxfordjournals.org/content/12/4/312.long