Computational Drug Repositioning: From Data to Therapeutics

MR Hurle[1], L Yang[1], Q Xie[1], DK Rajpal[2], P Sanseau[3] and P Agarwal[1]

Traditionally, most drugs have been discovered using phenotypic or target-based screens. Subsequently, their indications are often expanded on the basis of clinical observations, providing additional benefit to patients. This review highlights computational techniques for systematic analysis of transcriptomics (Connectivity Map, CMap), side effects, and genetics (genome-wide association study, GWAS) data to generate new hypotheses for additional indications. We also discuss data domains such as electronic health records (EHRs) and phenotypic screening that we consider promising for novel computational repositioning methods.

Since the advent of the genomic era, most of the searches for new drugs have begun with the concept of a single target that acts through a specific mechanism. In some cases, the target is genetically linked to the disease; in others, mechanistic hypotheses lead to a biochemical assay screen of the target, and the resulting tool compound is further evaluated in relevant model systems. Other searches begin with a phenotypic screen in which the model system itself is screened for efficacious compounds. In all these cases, developers then optimize the lead compound, hoping to avoid side effects due to either off-target binding or unanticipated physiologic roles of the intended target.[1] Between 1999 and 2008, this process resulted in only 50 first-in-class small-molecule agents being approved by the US Food and Drug Administration. Of these, 17 were identified as arising from target-based discovery methods and 28 from phenotypic discovery methods.[2]

Drug repositioning (also referred to as repurposing) has long been a necessary strategy of drug development[3–5] because it can renew a failed drug or expand the number of indications for a successful one.[6] **Figure 1** highlights the differences in the time lines of repositioning as compared with those of traditional drug discovery methods. Potentially, repositioning can reduce the traditional time line of 10–17 years and make drugs available for use in patients in 3–12 years.[3] Computational repositioning is the process of designing and validating automated workflows that can generate hypotheses for new indications for a drug candidate. The potential for computational repositioning is high, given that a systematic process can incorporate prioritization information that can accelerate time lines even further. Indications that lend themselves to a quick proof-of-concept or experimental-medicine study can be taken into account, and small clinical studies can be rapidly initiated for compounds for which safety data in human patients are already available.

It is prudent to examine all drug candidates—both abandoned and active—for repositioning, with particular focus on those that have passed human-safety hurdles because those would have the shortest path to patient benefit. The benefit–risk profile may vary depending on the unmet medical need. Side effects that might make a drug unacceptable for use in a chronic condition such as diabetes or asthma might be more acceptable in life-threatening medical conditions. For example, through repositioning, thalidomide has been approved for use in multiple myeloma and in erythema nodosum in leprosy, despite the serious teratogenicity associated with the drug.[7] Indeed, even drugs that have failed clinical trials for safety reasons might be repositioned. Regulatory considerations may, however, necessitate additional safety studies, depending on the new indication and the length of the trial proposed. Although regulatory exclusivity can provide a short period of market exclusivity for the new indication, reformulation is often considered necessary before generic compounds can acquire sufficient intellectual property value.[5]

Typically, repositioning has been accomplished by using the mechanistic knowledge of the target to infer a new disease indication or by observing new clinical phenotypes. In some recent instances, serendipity has played a part in the identification of new indications leading to repositioning[8]—most famously, the repositioning of sildenafil citrate from an anti-angina drug to

[1]Computational Biology, GlaxoSmithKline R&D, King of Prussia, Pennsylvania, USA; [2]Computational Biology, GlaxoSmithKline R&D, Research Triangle Park, North Carolina, USA; [3]Computational Biology, GlaxoSmithKline R&D, Stevenage, UK. Correspondence: P Agarwal (pankaj.agarwal@gsk.com)

Received 1 November 2012; accepted 29 December 2012; advance online publication 15 January 2013. doi:10.1038/clpt.2013.1

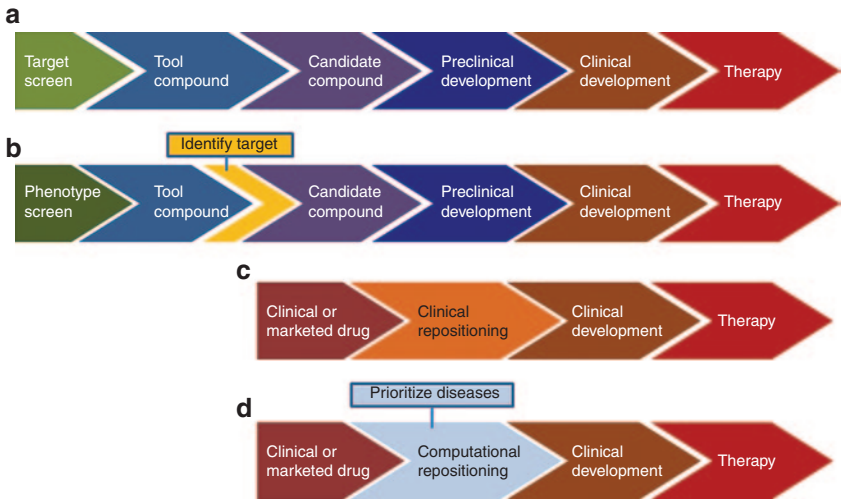

Figure 1 Four schematic pipelines to develop new therapies. (**a**) Target screening may result in a compound that undergoes development, followed by preclinical and clinical evaluation, ultimately becoming a therapy. (**b**) Phenotypic screening may identify a compound that follows a similar path as that of target screening; however, there is often a subsequent step to identify the molecular target of the compound. (**c**) Clinical observations suggest a repositioning hypothesis that may result in a clinical development plan for a new indication. (**d**) Computational methodologies systematically generate repurposing hypotheses for candidate molecules, possibly resulting in validation and clinical development to reposition the molecules.

a therapy for erectile dysfunction.[9] Systematic analysis of data in the literature has also led to new therapies. For example, Swanson used multiple lines of evidence to propose the use of fish oil to treat Raynaud's syndrome, a hypothesis that was later validated in a clinical trial.[10,11]

Computational analyses amplify these traditional approaches because they allow the researcher to generate, evaluate, and prioritize data for several drugs and diseases simultaneously. As a measure of progress, a semiautomated literature method recently recovered 14 of Swanson's 19 associations and computationally evaluated these for appropriate directionality.[12] Furthermore, with systematic efforts being continually bolstered by newly available platform data types, the rapidly expanding database in the literature, and improvements in analytical methodology, these methods can be expected to increase in value. It is important to note that the utility of these computational techniques extends beyond drug repositioning; they can also be used to find the initial indications for a drug.

COMPUTATIONAL REPOSITIONING METHODS

Excellent reviews and special issues on computational drug repositioning are available.[13,14] However, because this is a rapidly developing field, we highlight (i) some of the newer methods; (ii) advances in some of the key methods that we believe are the closest to delivering benefit to patients, generally based on the extent of experimental validation and support; and (iii)methods that are deemed promising for the future. There have been developments in each of the discussed methodologies in recent years, including some impressive *in vivo* validation of the transcriptional methods based on Connectivity Map (CMap). We also briefly discuss other methods that combine data across platforms and additional data sources such as electronic health records (EHRs) and screening data that are extremely promising but have not yet been applied directly for repositioning.

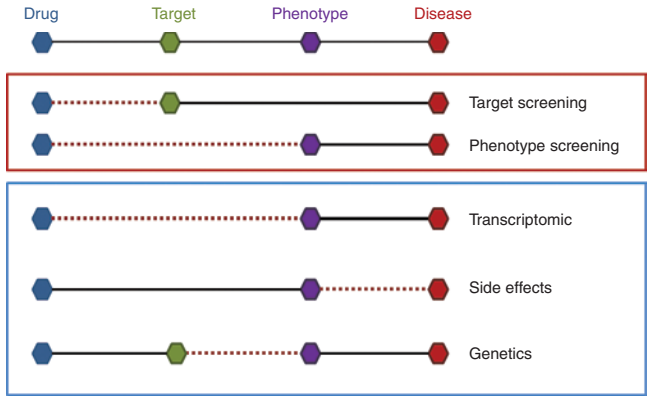

Figure 2 Repositioning methods (blue box) share some common features with standard drug discovery methods (red box). Dotted red lines represent the links generated by the approach. Black lines represent hypothesized links generated from domain knowledge. Target screening tries to identify a compound by screening against a target that is known to be associated with a specific disease. Phenotypic screening involves screening for a phenotype associated with the disease and may fill in the target later. Transcriptomic approaches can link a drug with an expression-based phenotype. Side-effect methods link a known drug phenotype to a new disease. Genetics-based methods can link targets with a phenotype that is associated with the disease (in many cases, the disease can be directly measured and therefore acts as the phenotype).

Repositioning methods have features in common with those of standard drug discovery; they identify new links between entities and complete the drug–disease connection using domain expertise (Figure 2). Traditional target-based drug discovery connects a target to a disease, and a screening process is used to find drug-like molecules that bind the target. Phenotypic screening skips the target and makes a direct attempt to find a drug-like molecule that will exhibit the correct phenotypic behavior. Most repositioning

methods use similar approaches but connect different concepts. We discuss three of these methods in more detail:

1. Transcriptomic methods such as CMap associate the drug with a specific aspect of a disease (e.g., a new phenotype based on genome-wide gene expression) and then extrapolate that knowledge to the disease itself.
2. Side-effect methods associate drugs with new disease indications by representing a disease on the basis of a set of side effects associated with its treatment and then identifying other drugs with similar side effects.
3. Genetics-based methods connect the target to a genetic phenotype, which may be the disease itself or a phenotype related to the disease.

In each case, it is the systematic combination of experimental data and preexisting knowledge that completes the connection.

Transcriptomics

Transcriptomic data can provide a list of over- and underexpressed genes in an experimental condition such as disease vs. normal or drug-treatment group vs. control. These gene lists can then be interrogated to evaluate pathways or networks that might be dysregulated. The CMap approach was one of the first attempts to take a more holistic view of these transcriptomic data and use them to connect expression profiles across conditions.[15] In particular, it suggested that drugs could potentially have a therapeutic effect on the disease condition if there is a strong negative correlation between the disease signatures and the drug expression profiles. CMap may be construed as a phenotypic screen for the drug, in which the expression disease signature in the context of the whole genome represents the phenotype. *In vivo* models have been used to validate predictions resulting from this approach, including ursolic acid for muscle weight loss[16] and the anticonvulsant drug topiramate for inflammatory bowel disease.[17]

The current version (build02) of the CMap database contains gene expression profiles generated by dosing of more than 1,300 compounds into one or more cell lines (http://www.broadinstitute.org/cmap). It can be probed with a variety of gene signatures representing disease states and other phenotypic models, using a method of similarity based on the Kolmogorov–Smirnov statistic used in gene set enrichment analysis.[18]

Although changes in expression in cell lines may not be perfectly translatable to those in human diseased tissues, and despite the fact that the dose is not optimized for each compound, several cases of experimentally validated results have recently been published using this database and methodology. These have been summarized elsewhere.[19] Nevertheless, it is worthwhile to highlight a few impressive examples.

In some instances, an interesting compound was initially identified because the researchers were motivated to identify a compound for the treatment of their disease of interest. They first generated a "disease signature," i.e., a list of genes that they believed showed increased or decreased expression in that particular disease; next, from the CMap database they identified potential compounds with strong negative scores to the signature. In the final step, they tested the compounds in an available model. Phenoxybenzamine was identified using a disease signature derived from a rat model of osteoarthritic pain and was then tested in the same model.[20] Ursolic acid, the major waxy component of apple peels, was identified using disease signatures generated from the muscles of fasting humans, fasting mice, and humans with spinal cord injury. Ursolic acid was then added to the diets of mice and found to increase muscle weight by 7%, an effect accompanied by consistent changes in other metabolic parameters.[16]

Aiding the effort to systematically exploit these data are gene expression repositories such as the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) and Array Express (http://www.ebi.ac.uk/arrayexpress). These contain raw data and processed results from thousands of individual experiments, making it possible to probe the CMap database with collections of disease signatures[21,22] produced by other researchers. Sirota *et al.*[22] created a set of 100 disease signatures from the Gene Expression Omnibus and identified many novel drug–disease pairs. Experimental validation for one of their predictions—cimetidine against lung carcinoma—included evidence of reduction in the growth of the lung adenocarcinoma cell line A549, both *in vitro* and in a mouse xenograft model. In another effort, the anticonvulsant drug topiramate was identified using a signature derived from an inflammatory bowel disease data set in the Gene Expression Omnibus and confirmed in a rat model.[17]

The CMap database can also be used to identify connections between the drugs themselves. These connections can be used to suggest that the indication for one drug could be an additional indication for another drug. The 1,309 molecules in CMap were connected into a network in MANTRA (http://mantra.tigem.it) on the basis of the similarities in their expression profiles.[23] The authors identified groups of tightly connected drugs with shared modes of action as defined by the manual curation of data including Anatomical Therapeutic Chemical codes. For example, fasudil (a rho-kinase inhibitor) was identified as being similar to deoxy-D-glucose. Given that deoxy-D-glucose is known to induce autophagy, it was proposed that fasudil might also induce autophagy. This was experimentally confirmed using an *in vitro* assay.[23]

A growing body of literature supports the use of CMap for repositioning; however, there are significant challenges as well. Quantitative estimates of its accuracy remain difficult. It is incumbent on the informatics community to evaluate relevant methods[24] and reduce the number of false positives, given that clinical verifications can prove expensive. A prediction can often be strengthened using independent disease signatures. If multiple drugs against the same target hit the same disease signature, it can suggest a target-based effect.

Efforts are under way to increase the number of cell lines represented in CMap and to decrease the resources needed to generate these profiles. One promising approach is to reduce the cost of the platform itself, i.e., to use only genes that are necessary for distinguishing patterns from one another rather than the whole genome itself (http://www.broadinstitute.org/LINCS). Such an advance would enable the use of CMap in higher-throughput

processes, perhaps eventually including high-throughput screening of millions of compounds to identify the ones that *ab initio* show the most promise for use in a particular disease condition.

Side effects

A more traditional source of repositioning hypotheses is the clinical setting. Several success stories highlight the role of clinical observations. An example is sildenafil citrate; while evaluating data for the primary therapeutic indication of angina, a careful study of side effects showed the drug's potential for therapeutic use in a new indication: erectile dysfunction.[9] Another example is exenatide; statistically significant weight loss was observed as a side effect in a clinical trial of the drug.[25] This led to a clinical trial to test its therapeutic effect in non-diabetic obese subjects (http://www.clinicaltrials.gov/ct2/show/NCT00856609). Similarity of side effects has been used to infer molecular targets of drugs; indeed, this work has provided the best source of side-effects-related data from drug labels.[26] Recent computational methods have leveraged these data and automated the approaches to predict new therapies.

Although it may seem counterintuitive to use side effects to suggest new therapies, therapeutic effects and side effects are both measurable changes resulting from a drug intervention. Therefore, side effects can potentially serve as "phenotypic biomarkers" for disease treatments. Consider therapeutics for the same disease that work through different mechanisms. Our hypothesis is that if they share the same uncommon side effects, there might be an underlying mechanism of action that links the side effect and the disease treatment. Computational methods have produced lists that relate diseases to specific side effects, generating hypotheses for new indications.[27,28] It has been estimated that, with respect to 27% of these newly predicted relationships, there had already been a published clinical trial that referenced both the drug and the newly predicted association with the disease.[27,28] For instance, the labels of many drugs indicated for control of transplant rejection report increased cytomegalovirus infections as a side effect. This is expected, given that these drugs are immune suppressants and would therefore be associated with a higher rate of infections. The information can be inverted to postulate that drugs that list increased cytomegalovirus infections as a possible side effect, e.g., methotrexate, may have value as potential treatments for transplant rejection.[28] The use of methotrexate for transplant rejection has been reported.[29] One of the main advantages of this computational strategy is that there are no translational issues because both the physiologic therapeutic modulation and the side effects are observed in human subjects rather than in animal models.

Given that the average number of side effects reported in a drug label is close to 70,[30] an analysis of the data poses a challenge. Although the high number of side effects does allow computation of the specificity in terms of relationship to diseases, it also introduces a source of noise because the rate of side effects reported may not be significantly higher than that in the placebo control groups. Unfortunately, such quantitative data are not widely available. It would also be advisable to take explicit account of mechanism of action, dosage, exposure, and genetic variation. The careful combination of side-effect data with other human pharmacologic data, perhaps from postmarketing studies, offers some potential to reposition drugs while avoiding the issue of translating from model systems.

Genetics

Human genetics studies offer one of the strongest lines of evidence to connect specific genes to specific human diseases. A drug can be evaluated for repositioning to treatment of another disease if its known protein target is genetically associated with a disease that is not among those for which the drug is indicated.

Recent technological advances such as next-generation sequencing[31] have accelerated the discovery of gene–disease associations by drastically reducing the cost of whole-exome sequencing.[32] One example of the utility of this method in drug repositioning is the identification of *NTE5*, a gene encoding an enzyme involved in adenosine metabolism, as being linked to a dominant and adult-onset form of arterial calcification.[33] In this case, it was proposed that an antithrombotic drug, dipyridamole, could potentially be helpful because it inhibits reuptake of adenosine. For many monogenic rare disorders, the mutations in the causal genes are often disruptive of the protein product, and it is therefore more likely that an activator will be needed if the target is the disease gene.

Genome-wide association studies (GWAS) have shown the association between genetic variants and polygenic diseases, resulting in the identification of genes proximal to these variants as being linked to numerous complex diseases.[34] Francis Collins, in a commentary announcing the establishment of the US National Center for Advancing Translational Sciences, observed that 6 of the 44 GWAS loci for type 2 diabetes were associated with marketed drugs. This observation was then expanded to suggest that GWAS-identified genes (rather than a random set of genes) were more likely to be tractable by small molecules and biologicals.[35] These studies suggest that the convergence of human genetics and pharmacology might highlight some of the better drug targets. A total of 155 genes were identified as being targeted by at least one asset already on the market or in development.[35] For 92 of these target genes, the drug indication was different from the disease trait identified by GWAS. This suggests that drugs that target these 92 gene products should be evaluated for the new disease traits indicated by the GWAS.

As a specific example, a GWAS meta-analysis identified a variant near *TNFSF11* as being associated with Crohn's disease ($P = 4.9 \times 10^{-10}$).[35] TNFSF11, commonly known as receptor activator of nuclear factor-κB ligand, together with tumor necrosis factor-α, activates the nuclear factor-κB pathway. There are antibodies for the treatment of osteoporosis on the market as well as in development, that target the receptor activator of nuclear factor-κB ligand by preventing activation.[36] The GWAS data suggest that these antibodies may be potentially useful for treating Crohn's disease. In addition, allelic expression studies have demonstrated that the Crohn's risk allele increases the expression of the receptor activator of nuclear factor-κB ligand, indicating that repression of its activity could be an appropriate therapy.[36]

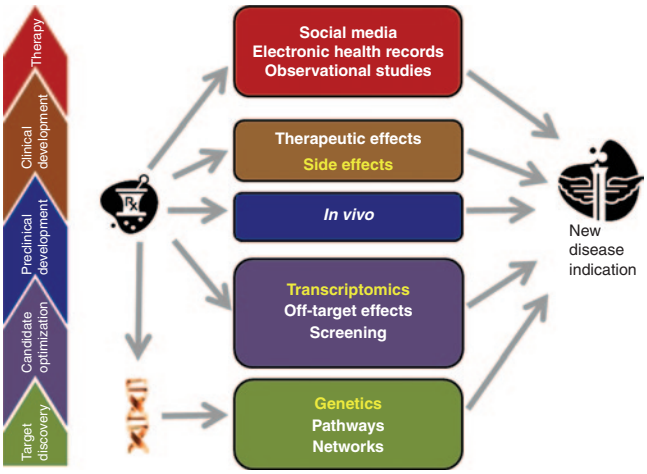

Figure 3 Data are produced during the various stages of the drug discovery and development process, serving as a critical resource for developing repositioning methodologies. Major sources of data highlighted in the text are indicated in yellow.

It is evident that the gene–disease associations identified by GWAS inform possible drug repositioning opportunities. However, there are still some challenges in using GWAS data.[34] The lack of a direction of the required therapeutic effect makes it difficult to ascertain, from GWAS information alone, whether an activator or an inhibitor is required. Moreover, monogenic disorders are mostly caused by a loss of function, and this matches poorly with the fact that most drugs are antagonists. Repositioning methods should therefore focus on drugs that either target negative regulators of the identified gene or offset the mechanistic effect of the loss of function as in the dipyridamole example discussed earlier.

Other methods

In addition to the three methods discussed above, the various stages of drug discovery produce additional data relevant to repurposing (**Figure 3**). Significant effort has been expended in improving the mechanistic understanding and classification of diseases so as to identify potential drug targets. The target discovery field is therefore rich in data from genetics, signaling pathways, and protein–protein interactions, much of which has been processed into networks. New indications have been predicted from these data, often by evaluating similarities between diseases.[37–39] Screening data[40] and off-target effects offer tantalizing opportunities to reposition drugs. The availability of a compound also enables the use of transcriptomic techniques such as CMap. *In vivo* models, including mouse knockout models, are a potential source of repositioning data, given that international consortia can generate high-throughput mammalian phenotypic data in standard formats.[41] Surprisingly, little or no use of knockout data has yet been made to directly and systematically reposition drugs. We have already discussed the use of pharmacologic data in terms of side effects, but increased capture and analysis of efficacy data and other data from clinical trials might lead to additional opportunities. Observational studies, ranging from case reports to statistically powered studies based on health-care claims data and EHRs, have proven powerful in detecting adverse-event signals, but a compelling case for their use in repositioning drugs still needs to be made.

Pathways and networks (disease similarity). Diseases can be connected to one another in a network on the basis of shared features. For example, one study was able to link diseases with protein-interaction modules identified as being transcriptionally regulated[39] and compare diseases with one another on the basis of these shared modules. Disease similarity can then be used to reposition drugs from one disease to another using either these modules[39] or canonical pathways.[38] Chiang and Butte[42] used another disease similarity approach that they termed "guilt by association"; this approach suggested that if two diseases shared many therapies then the drugs approved for one of the diseases should also be evaluated for efficacy against the other disease. The researchers showed that suggestions based on this method were 12 times more likely to lead to a clinical trial associated with the hypotheses. Integrative methods (such as PREDICT) suggest that if the drug X and the disease Y are similar to the drug and disease, respectively, in a known drug–disease relationship, then X should be evaluated for treating Y.[43] PREDICT achieved a high statistical measure of accuracy using cross-validation. However, the black-box nature of these predictions can present a challenge for experimental verification. Regardless, if the precision of these techniques is proven to be high, they will gain more widespread acceptance and will be experimentally validated.

Screening. High-throughput screening results may be a rich source of repositioning hypotheses. A public resource that facilitates such screening is PubChem,[44] which contains results from thousands of screens, many of them being phenotypic ones. If a drug or a closely related molecule is active in a phenotypic screen, it presents the simplest of cases for repositioning, and it should be tested in the disease associated with that phenotype.[40] PubChem bioassay data have already been used to predict adverse drug reactions manifested in specific system organ classes.[45] It is conceivable that similar approaches may be used to predict disease indications as well.

Off-target effects. Once a compound has been identified as binding to a target, the related targets and compounds can be predicted using structural algorithms. Potential targets (i.e., off-targets) can be identified via similarities of their ligand-binding pockets. Using such an approach, the enoyl-acyl carrier protein reductase of *Mycobacterium tuberculosis* was determined to have a structure similar to that of rat Catechol-o-methyltransferase, the target of the Parkinson's disease drug entacapone. The compound was found to inhibit both the activity of enoyl-acyl carrier protein reductase and the growth of the pathogen.[46] Because not all protein structures have been resolved, methods utilizing the structural or physiochemical properties of the ligands were developed,[47,48] impressively validating 23 of 30 novel drug-target predictions.[49] An approach that uses knowledge of both compound and protein structure is molecular docking,[50] which estimates the physicochemical strength of ligand–protein interactions. This approach has successfully

identified the psychiatric drug haloperidol as a lead anti-HIV compound.[51] Because the compound will also bind to its originally identified target under most conditions, repositioning between anti-infective drugs and human targets is likely to be the most efficacious application of off-target methods.

In vivo. *In vivo* phenotypic screening using targeted mutations has been widely used to associate genes with phenotypes. The Mouse Phenome Database (http://phenome.jax.org) contains about 1,400 phenotypic measurements related to human diseases, including cancer susceptibility, aging, obesity, infections, atherosclerosis, blood disorders, and neurosensory disorders.[52] These phenotypes associate a gene with a disease and, by inference, a drug that targets the gene product with the disease. The emergence of international consortia to generate mammalian phenotypic data in standard formats bodes well for the utility of these data.[41] We believe such data will play an increasing role in future repositioning efforts, although the translation of findings from *in vivo* models to humans may prove challenging.

Therapeutic effects. Although success stories have been reported on repositioning of drugs on the basis of clinical observations of pharmacologic effects in human patients, there is a paucity of published literature on the systematic use of repositioning. This is perhaps a consequence of the lack of publicly available clinical trial data; however, such clinical trial data are now becoming more readily available.[53]

Observational studies, EHRs, and social media. As drugs progress into the clinic, they generate clinical data that can be analyzed for repositioning opportunities. These data may be dwarfed by real-world observational data that become available once the drug is approved. The almost incomprehensible volume and complexity of EHRs present opportunities that could well be the source of the next great advances in drug repositioning.[54] Online self-reported patient data have been analyzed to test the effect of lithium on the progression of amyotrophic lateral sclerosis.[55] Adverse drug reactions may be mentioned on message boards, and methods that might detect signals are emerging.[56] These methods may eventually be extended to detecting beneficial effects as well as to other social media. Methods developed and evaluated as part of the Observational Medical Outcomes Partnership, primarily to detect early signals of adverse events, may also one day be extrapolated to the discovery of beneficial drug effects.[57] A landmark study (using spontaneous reporting data from the Food and Drug Administration's Adverse Event Reporting System) systematically predicted associations between adverse events and drugs and then verified some of the predictions using EHR data.[58] These efforts relating to mining of observed clinical side effects also offer a great opportunity to reposition drugs.

Experimental validation and independent verification add validity to computational hypotheses. Efforts at mining data from the published literature can often identify case reports or experimental medicine studies that support a hypothesis. In other instances, internal preclinical or clinical observations can strengthen or even confirm hypotheses. Without a computational hypothesis, a preclinical or clinical hypothesis may not be actionable; however, once the two are put together, they can accelerate the path toward a clinical trial.

Conclusion

Phenotypic and target-based discovery are the cornerstones of the drug discovery process. It is also indisputable that repositioning is a key value driver. We have discussed approaches to computational repositioning, and, because of the long lead times, most of the published examples of repositioning are currently supported by preclinical validations. Nevertheless, new methods that leverage challengingly large data sets (such as those relating to transcriptomics, side effects, and genetics) are becoming established. The increasing availability of clinical trial data may present an opportunity in this regard as well.[53] All these efforts suggest the immense potential of computationally detecting repositioning opportunities from clinical phenotypic data.

Both the Medical Research Council in the United Kingdom (http://www.mrc.ac.uk/Fundingopportunities/Calls/MoD/MRC008389) and the National Institutes of Health in the United States[59] have announced public–private partnerships to repurpose failed pipeline drugs. Eventually, some of the funded proposals may be based on computational hypotheses such as CMap-based profiles (http://www.genometry.com/open-innovation) or could use innovative ideas emerging from screening the US National Center for Advancing Translational Sciences collection in the Lilly PD^2 panel.[60] The National Institutes of Health's National Center for Advancing Translational Sciences has a collection of 3,800 approved and investigational compounds.[61] This collection could be an excellent resource for repositioning, by integrating existing knowledge and generating new experimental data around these compounds.

Adoption of computational methodologies to reposition drugs is expanding and becoming mainstream. There have been several recent examples demonstrating validation using *in vivo* models, and this trend is likely to improve with further collaboration between the clinical pharmacology and bioinformatics communities.[62] The rise of translational bioinformatics as a discipline suggests that the field of computational biology is maturing and is perhaps more ready than ever to develop solutions that will have a beneficial impact on human health in the near term.[63,64]

ACKNOWLEDGMENTS

The authors thank Vinod Kumar, Aaron Spivak, and David Soong for their critical reading of the manuscript.

CONFLICT OF INTEREST

The authors are employed by GlaxoSmithKline.

1. Achenbach, J., Tiikkainen, P., Franke, L. & Proschak, E. Computational tools for polypharmacology and repurposing. *Future Med. Chem.* **3**, 961–968 (2011).
2. Swinney, D.C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507–519 (2011).
3. Ashburn, T.T. & Thor, K.B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).
4. Aronson, J.K. Old drugs–new uses. *Br. J. Clin. Pharmacol.* **64**, 563–565 (2007).

5. Barratt, M.J., & Frail, D.E. *Drug Repositioning: Bringing New Life to Shelved Assets and Existing Drugs.* (John Wiley & Sons, Hoboken, NJ, 2012).
6. Gelijns, A.C., Rosenberg, N. & Moskowitz, A.J. Capturing the unexpected benefits of medical research. *N. Engl. J. Med*. **339**, 693–698 (1998).
7. Ladizinski, B., Shannon, E.J., Sanchez, M.R. & Levis, W.R. Thalidomide and analogues: potential for immunomodulation of inflammatory and neoplastic dermatologic disorders. *J. Drugs Dermatol*. **9**, 814–826 (2010).
8. Chong, C.R. & Sullivan, D.J. Jr. New uses for old drugs. *Nature* **448**, 645–646 (2007).
9. Ghofrani, H.A., Osterloh, I.H. & Grimminger, F. Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat. Rev. Drug Discov*. **5**, 689–702 (2006).
10. Swanson, D.R. Medical literature as a potential source of new knowledge. *Bull. Med. Libr. Assoc*. **78**, 29–37 (1990).
11. Swanson, D.R. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect. Biol. Med*. **30**, 7–18 (1986).
12. Cameron, D. *et al*. A graph-based recovery and decomposition of Swanson's hypothesis using semantic predications. *J. Biomed. Inform*. (2012); e-pub ahead of print 28 September 2012.
13. Sanseau, P. & Koehler, J. Editorial: computational methods for drug repurposing. *Brief. Bioinformatics* **12**, 301–302 (2011).
14. Dudley, J.T., Deshpande, T. & Butte, A.J. Exploiting drug-disease relationships for computational drug repositioning. *Brief. Bioinformatics* **12**, 303–311 (2011).
15. Lamb, J. *et al*. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
16. Kunkel, S.D. *et al*. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. *Cell Metab*. **13**, 627–638 (2011).
17. Dudley, J.T. *et al*. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med*. **3**, 96ra76 (2011).
18. Subramanian, A. *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
19. Qu, X.A. & Rajpal, D.K. Applications of connectivity map in drug discovery and development. *Drug Discov. Today* **17**, 1289–1298 (2012).
20. Chang, M., Smith, S., Thorpe, A., Barratt, M.J. & Karim, F. Evaluation of phenoxybenzamine in the CFA model of pain following gene expression studies and connectivity mapping. *Mol. Pain* **6**, 56 (2010).
21. Hu, G. & Agarwal, P. Human disease-drug network based on genomic expression profiles. *PLoS ONE* **4**, e6536 (2009).
22. Sirota, M. *et al*. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med*. **3**, 96ra77 (2011).
23. Iorio, F. *et al*. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. USA* **107**, 14621–14626 (2010).
24. Cheng, J., Xie, Q., Hurle, M., Freudenberg, J.M., Yang, L. & Agarwal, P. Evaluation of analytical methods for Connectivity map data. *Pac. Symp. Biocomput*. **18**, 5–16 (2013).
25. Buse, J.B., Henry, R.R., Han, J., Kim, D.D., Fineman, M.S. & Baron, A.D.; Exenatide-113 Clinical Study Group. Effects of exenatide (exendin-4) on glycemic control over 30 weeks in sulfonylurea-treated patients with type 2 diabetes. *Diabetes Care* **27**, 2628–2635 (2004).
26. Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J. & Bork, P. Drug target identification using side-effect similarity. *Science* **321**, 263–266 (2008).
27. Duran-Frigola, M. & Aloy, P. Recycling side-effects into clinical markers for drug repositioning. *Genome Med*. **4**, 3 (2012).
28. Yang, L. & Agarwal, P. Systematic drug repositioning based on clinical side-effects. *PLoS ONE* **6**, e28025 (2011).
29. Cahill, B.C. *et al*. Methotrexate for lung transplant recipients with steroid-resistant acute rejection. *J. Heart Lung Transplant*. **15**, 1130–1137 (1996).
30. Duke, J., Friedlin, J. & Ryan, P. A quantitative analysis of adverse events and "overwarning" in drug labeling. *Arch. Intern. Med*. **171**, 944–946 (2011).
31. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet*. **24**, 133–141 (2008).
32. Bamshad, M.J. *et al*. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet*. **12**, 745–755 (2011).
33. St Hilaire, C. *et al*. NT5E mutations and arterial calcifications. *N. Engl. J. Med*. **364**, 432–442 (2011).
34. Hindorff, L.A. *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
35. Franke, A. *et al*. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet*. **42**, 1118–1125 (2010).
36. Sanseau, P. *et al*. Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol*. **30**, 317–320 (2012).
37. Butte, A.J. & Kohane, I.S. Creation and implications of a phenome-genome network. *Nat. Biotechnol*. **24**, 55–62 (2006).
38. Li, Y. & Agarwal, P. A pathway-based view of human diseases and disease relationships. *PLoS ONE* **4**, e4346 (2009).
39. Suthram, S., Dudley, J.T., Chiang, A.P., Chen, R., Hastie, T.J. & Butte, A.J. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput. Biol*. **6**, e1000662 (2010).
40. Swamidass, S.J. Mining small-molecule screens to repurpose drugs. *Brief. Bioinformatics* **12**, 327–335 (2011).
41. Skarnes, W.C. *et al*. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
42. Chiang, A.P. & Butte, A.J. Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. *Clin. Pharmacol. Ther*. **86**, 507–510 (2009).
43. Gottlieb, A., Stein, G.Y., Ruppin, E. & Sharan, R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol*. **7**, 496 (2011).
44. Wang, Y. *et al*. PubChem's BioAssay Database. *Nucleic Acids Res*. **40**, D400–D412 (2012).
45. Pouliot, Y., Chiang, A.P. & Butte, A.J. Predicting adverse drug reactions using publicly available PubChem BioAssay data. *Clin. Pharmacol. Ther*. **90**, 90–99 (2011).
46. Kinnings, S.L., Liu, N., Buchmeier, N., Tonge, P.J., Xie, L. & Bourne, P.E. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol*. **5**, e1000423 (2009).
47. Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J. & Shoichet, B.K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol*. **25**, 197–206 (2007).
48. Lounkine, E. *et al*. Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486**, 361–367 (2012).
49. Keiser, M.J. *et al*. Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009).
50. Kitchen, D.B., Decornez, H., Furr, J.R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov*. **3**, 935–949 (2004).
51. DesJarlais, R.L. *et al*. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. USA* **87**, 6644–6648 (1990).
52. Grubb, S.C., Maddatu, T.P., Bult, C.J. & Bogue, M.A. Mouse phenome database. *Nucleic Acids Res*. **37**, D720–D730 (2009).
53. Coombes, R. GlaxoSmithKline grants researchers access to clinical trial data. *BMJ* **345**, e6909 (2012).
54. Jensen, P.B., Jensen, L.J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet*. **13**, 395–405 (2012).
55. Wicks, P., Vaughan, T.E., Massagli, M.P. & Heywood, J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat. Biotechnol*. **29**, 411–414 (2011).
56. Benton, A. *et al*. Identifying potential adverse effects using the web: a new approach to medical hypothesis generation. *J. Biomed. Inform*. **44**, 989–996 (2011).
57. Harpaz, R., DuMouchel, W., Shah, N.H., Madigan, D., Ryan, P. & Friedman, C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin. Pharmacol. Ther*. **91**, 1010–1021 (2012).
58. Tatonetti, N.P., Ye, P.P., Daneshjou, R. & Altman, R.B. Data-driven prediction of drug effects and interactions. *Sci. Transl. Med*. **4**, 125ra31 (2012).
59. Allison, M. NCATS launches drug repurposing program. *Nat. Biotechnol*. **30**, 571–572 (2012).
60. Lee, J.A. *et al*. Open innovation for phenotypic drug discovery: The PD2 assay panel. *J. Biomol. Screen*. **16**, 588–602 (2011).
61. Huang, R. *et al*. The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. *Sci. Transl. Med*. **3**, 80ps16 (2011).
62. Lesko, L.J. Drug research and translational bioinformatics. *Clin. Pharmacol. Ther*. **91**, 960–962 (2012).
63. Butte, A.J. Translational bioinformatics: coming of age. *J. Am. Med. Inform. Assoc*. **15**, 709–714 (2008).
64. Altman, R.B. Translational bioinformatics: linking the molecular world to the clinical world. *Clin. Pharmacol. Ther*. **91**, 994–1000 (2012).