



Human and Machine Intelligence Together Drive Drug Repurposing in Rare Diseases

Anup P. Challa^{1,2†}, Nicole M. Zaleski^{1†}, Rebecca N. Jerome¹, Robert R. Lavieri¹, Jana K. Shirey-Rice¹, April Barnado³, Christopher J. Lindsell⁴, David M. Aronoff^{5,6,7}, Leslie J. Crofford³, Raymond C. Harris⁸, T. Alp Ikizler⁸, Ingrid A. Mayer⁹, Kenneth J. Holroyd¹⁰ and Jill M. Pulley^{1*}

¹ Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, United States, ² Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN, United States, ³ Division of Rheumatology and Immunology, Department of Medicine, Vanderbilt Medical Center, Nashville, TN, United States, ⁴ Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, United States, ⁵ Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States, ⁶ Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, United States, ⁷ Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, United States, ⁸ Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States, ⁹ Division of Hematology/Oncology, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, United States, ¹⁰ Center for Technology Transfer and Commercialization, Vanderbilt University, Nashville, TN, United States

OPEN ACCESS

Edited by:

Subodh Kumar Mishra,
University at Albany, United States

Reviewed by:

Khanh N. Q. Le,
Taipei Medical University, Taiwan
Fuhai Li,
Washington University in St. Louis,
United States

*Correspondence:

Jill M. Pulley
jill.pulley@vumc.org

†These authors share first authorship

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 10 May 2021

Accepted: 06 July 2021

Published: 28 July 2021

Citation:

Challa AP, Zaleski NM, Jerome RN, Lavieri RR, Shirey-Rice JK, Barnado A, Lindsell CJ, Aronoff DM, Crofford LJ, Harris RC, Alp Ikizler T, Mayer IA, Holroyd KJ and Pulley JM (2021) Human and Machine Intelligence Together Drive Drug Repurposing in Rare Diseases. *Front. Genet.* 12:707836. doi: 10.3389/fgene.2021.707836

Repurposing is an increasingly attractive method within the field of drug development for its efficiency at identifying new therapeutic opportunities among approved drugs at greatly reduced cost and time of more traditional methods. Repurposing has generated significant interest in the realm of rare disease treatment as an innovative strategy for finding ways to manage these complex conditions. The selection of which agents should be tested in which conditions is currently informed by both human and machine discovery, yet the appropriate balance between these approaches, including the role of artificial intelligence (AI), remains a significant topic of discussion in drug discovery for rare diseases and other conditions. Our drug repurposing team at Vanderbilt University Medical Center synergizes machine learning techniques like phenome-wide association study—a powerful regression method for generating hypotheses about new indications for an approved drug—with the knowledge and creativity of scientific, legal, and clinical domain experts. While our computational approaches generate drug repurposing hits with a high probability of success in a clinical trial, human knowledge remains essential for the hypothesis creation, interpretation, “go-no go” decisions with which machines continue to struggle. Here, we reflect on our experience synergizing AI and human knowledge toward realizable patient outcomes, providing case studies from our portfolio that inform how we balance human knowledge and machine intelligence for drug repurposing in rare disease.

Keywords: drug repurposing, evidence synthesis, rare diseases, machine learning, phenome wide association studies, precision medicine

INTRODUCTION

In today's age of big healthcare data, how should artificial intelligence (AI) and human critical thinking and creativity coexist within the drug development process? The long-range effects of integrating AI tools such as machine learning (ML) into drug development remain to be determined, particularly for complex problems such as rare disease strategies. Within the burgeoning field of drug repurposing—which seeks to facilitate and accelerate drug development by discovering new use cases for existing drugs (Nosengo, 2016)—this discussion is especially timely. The oldest, most well-known cases of repurposing, such as sildenafil for erectile dysfunction (Roundtable on Translating Genomic-Based Research for Health et al., 2014), are often considered serendipity but were in reality unexpected effects astutely recognized by humans. As procedures and knowledge bases for the identification and validation of repurposing candidates continue to mature and grow (Challa et al., 2018; Xue et al., 2018; Pulley et al., 2020), the field is extremely fertile for the application of AI to identify patterns suggestive that a drug candidate is attractive for repurposing.

Repurposing of existing therapeutic agents has generated significant interest in the realm of rare diseases as an innovative strategy for finding new opportunities to manage these complex conditions (Delavan et al., 2018; Valencic et al., 2018; Scherman and Fetro, 2020; Roessler et al., 2021). As data resources that integrate longitudinal phenotype information with genetic data and patient demographics continue to expand and evolve, so does their utility for identifying new therapeutic insights for an array of rare diseases. Several studies have harnessed these data with ML to identify drug repurposing candidates, exploring connections among multi-omics data with potential therapeutic implications (Jarada et al., 2020; Chen et al., 2021). Drug discovery in rare disease has been a particularly innovative use case for this approach, allowing investigators to predict molecular etiologies for rare diseases from existing knowledge on illnesses with similar presentations. Subsequently, these predictions can inform a shortlist of repurposable therapies for such rare diseases. This is done by considering the therapeutic indices of agents indicated for those illnesses that ML identifies as having pathophysiology and symptomology similar to the rare disease of interest (Álvarez-Machancoses and Fernández-Martínez, 2019; Brasil et al., 2019; Lee et al., 2019). The power of ML in this context is in its capacity to identify patterns at scale. ML's automation of this discovery step expedites the pace of hit discovery and expands the landscape of potential candidates translatable to future stages of the development pipeline. This approach's efficiency is magnified when the probability of this kind of a hit is otherwise low, as it is for many rare disease therapeutic campaigns (Ekins et al., 2019).

Despite the promise, current ML architectures struggle to parse the diverse, semi-structured feature sets inherent to the available data. Thus, we propose that ML can be a supplement—but not a replacement—for the perspectives of domain experts in drug repurposing. The complementary and interconnected nature of human intelligence and AI is becoming apparent (Kim et al., 2021). We assert that in its current state ML is

best leveraged in drug repurposing efforts to inform human “go/no-go” decision-making. Relying on rigid rules-based criteria typically required by ML overlooks that important information remains insufficiently codified to even be able to apply a rule.

Our drug repurposing team at Vanderbilt University Medical Center (VUMC) synergizes computational techniques like the phenome-wide association study (PheWAS) (Denny et al., 2010, 2013)—an ML method for generating hypotheses about new indications for an approved drug—with data mining from public databases (Challa et al., 2019). As we describe throughout the case studies we present in this manuscript, PheWAS inputs phenotypes encoded as administrative billing codes [e.g., the International Classification of Diseases ontology (World Health Organization, 1993; Centers for Disease Control and Prevention, 2021)] and tests the strength of associations between the presence of these codes in patients' electronic health records (EHRs) and the incidence of single nucleotide mutations recorded for these patients within clinical genomics repositories that accompany their EHRs. The associations are formalized through a logistic classification algorithm, which provides *p*-values (correctable for multiple testing) and odds ratios to quantitatively assess the strength of each predicted mutation-disease pair (Denny et al., 2010, 2013). These data are then overlaid with the knowledge of scientific and clinical domain experts, which facilitates review of true causality, identification of drugs associated with the targets carrying implicated mutations on their encoding genes, as well as the likelihood of successfully repurposing these drugs for the diseases indicated by PheWAS. In this article, we reflect on the processes involved in our program, with a view to areas of synergy for ML approaches like PheWAS and human knowledge. We share illustrative examples from our portfolio that support our beliefs in the unique balance of human and machine intelligence required for drug repurposing in rare disease.

OVERVIEW OF THE DRUG REPURPOSING PLATFORM AT VANDERBILT UNIVERSITY MEDICAL CENTER

Our drug repurposing platform at VUMC (Pulley et al., 2017, 2018a,b; Goldstein et al., 2018; Jerome et al., 2018; Choby et al., 2019) leverages natural human genetic variation as a proxy for, and method for more accurately predicting, the physiologic effects of therapies in humans. The key resource enabling this work is BioVU, a repository of 245,000 unique, de-identified DNA samples derived from discarded blood collected during routine clinical testing (Roden et al., 2008). BioVU, combined with extensive and granular phenotype data from the EHR, serves as a centralized resource for conducting largescale disease-agnostic research on fundamental questions of how genetic variation corresponds to variations in observable attributes, like the PheWAS analyses described above. The open-source codebase enabling execution of PheWAS and replication of previously published PheWAS signals (Pulley et al., 2017; Jerome et al., 2020) and a catalog of single nucleotide polymorphism

(SNP)/phenotype findings are publicly available online (Denny et al., 2013; PheWAS, 2021).

With a sufficiently large dataset, SNPs with low minor allele frequencies can allow for assessment of therapeutic targets among rare diseases, given the size and breadth of data yielded by large, diverse healthcare centers. As above, our methods (**Figure 1**) identify variants in drug target genes (Wishart et al., 2018) and then execute PheWAS (Denny et al., 2010) to find the diseases associated with these SNPs. When a known indication of a drug results from this analysis (e.g., when a SNP that is known to lower gastric acid secretion has a statistically significant association with reduced risk of heartburn), we can reasonably conclude that the SNP recapitulates the effects of a comparative drug (e.g., a proton pump inhibitor). Building upon these validation signals, we assess associations between the SNP and potential novel indications for the associated drug, as represented in phenotypes yielded by PheWAS. Our method does not solely rely on the classifier; it also requires the integration of conceptual knowledge on the clinical manifestations of human disease and markers of disease pathophysiology. To facilitate this conceptual synthesis of existing knowledge, we have developed a plausibility assessment approach with extensive evidence review using comprehensive searches of a wide range of information science resources. These methods are applied to make selections for clinical development. Using this procedure, we have established a suite of 13 drug-indication pairings, and we have generated confirmatory data for five (out of six) programs to date. **Table 1** provides several illustrative examples from our program's slate of projects.

Each project in our repurposing pipeline represents a precise disease or, when the disease is not a classical rare disease, a rare disease endotype, defined by a discrete pathophysiologic mechanism and cluster of characteristics that is revealed by review and synthesis of narrative descriptions of disease presentations (e.g., as detailed in clinical notes) within candidate EHRs. It is well known that the presentations of rare diseases are often so complex or opaque that even skilled clinicians require extensive workups and multidisciplinary consults to unearth a sufficiently discriminatory differential. ML's goal of pattern recognition could allow for more efficient identification of these rare disease differential diagnoses, but to make accurate predictions of which drugs might work for which rare disease, the ML algorithms require knowledge of presence, absence, and corresponding significance of features to facilitate accurate rule-in/rule-out modeling. We propose that these methods do not yet have the sophistication to mirror or replace the expert domain knowledge applicable to finding new rare disease indications for existing drugs, as we have found in our repurposing experiences.

LIMITATIONS OF THE UTILITY OF AI BASED ON OUR DRUG REPURPOSING EXPERIENCES

Researchers are developing approaches to systematically support and expand repurposing efforts, some of which use ML technologies (Chen et al., 2021; Gupta et al., 2021;

Issa et al., 2021). Promising techniques include quantitative structure-activity relationship modeling (Ekins et al., 2019; Challa et al., 2020), *in silico* docking experiments on druggable targets, virtual high-throughput screening, adverse event matching, and applying advanced statistical approaches to big clinical, genomic, pathway, and gene regulation data to discover new relationships within this information toward personalized medicine (Challener, 2018; Le et al., 2019, 2020). Despite the advantages of an AI approach like unsupervised ML to discover previously unrecognized patterns, as well as the large extent to which AI allows for pliable model development, we have found several challenges hindering the potential of AI to be fully realized in the context of our platform. While we believe that PheWAS offers an advance over other functionally similar ML approaches in its ability to work robustly across the healthcare data of diverse enterprises (Hermann et al., 2021; Salvatore et al., 2021; Schneider et al., 2021) and its empowerment of holistic, high-throughput discovery through minimal model pre-conditioning, its results require manual interpretation.

A key first step in our hypothesis generation workflow is identifying, *via* PheWAS, SNPs within a drug target's genes that replicate known therapeutic information about the drug (i.e., known indications or side effects of the drug). Given that the indications or side effects on a drug's label may not directly correlate to patient experiences in the real world, identification of the "controls" that underlie our models requires leveraging deep conceptual knowledge from the literature and from clinical collaborators. For example, we used "aphasia" (odds ratio = 5.05, $p = 0.007$) as an anchoring phenotype in analyzing PheWAS results for *GRIN2A*, which encodes the protein target of memantine. Memantine is approved for treatment of Alzheimer's disease but has limited efficacy for that indication as a standalone therapy. Literature evidence indicates stronger therapeutic efficacy for aphasia (Berthier et al., 2011), which is among the top associations in our PheWAS dataset and is also a feature of neuropsychiatric lupus, which we further identified as a pathogenicity of SNPs linked to *GRIN2A*. Cognitive dysfunction in patients with systemic lupus erythematosus patients then became our repurposed indication, which we are currently interrogating through clinical testing (NCT03527472). If we had only relied on ML to construct our model, would a rules-based classifier have recognized a theme of aphasia, or the "rare" subtype of lupus, when these diseases are written in clinical narratives without formal diagnosis?

Another example of ways in which manual patient chart reviews can contribute to our process involves our use of electrolyte imbalance as the anchoring phenotype for *TACR1*, which encodes the protein target of aprepitant. Aprepitant is approved for chemotherapy-induced nausea and vomiting; review of patient charts confirmed that vomiting explained our controlling set of electrolyte imbalance PheWAS results in patients with *TACR1* mutations. Could ML stably and efficiently integrate such relationship-driven disease classification knowledge into a larger inference model, in a way that would prevent calling this true-positive result a false-negative?

Establishing parity between a drug's mechanism of action and SNP directionality enables us to confirm known indications,

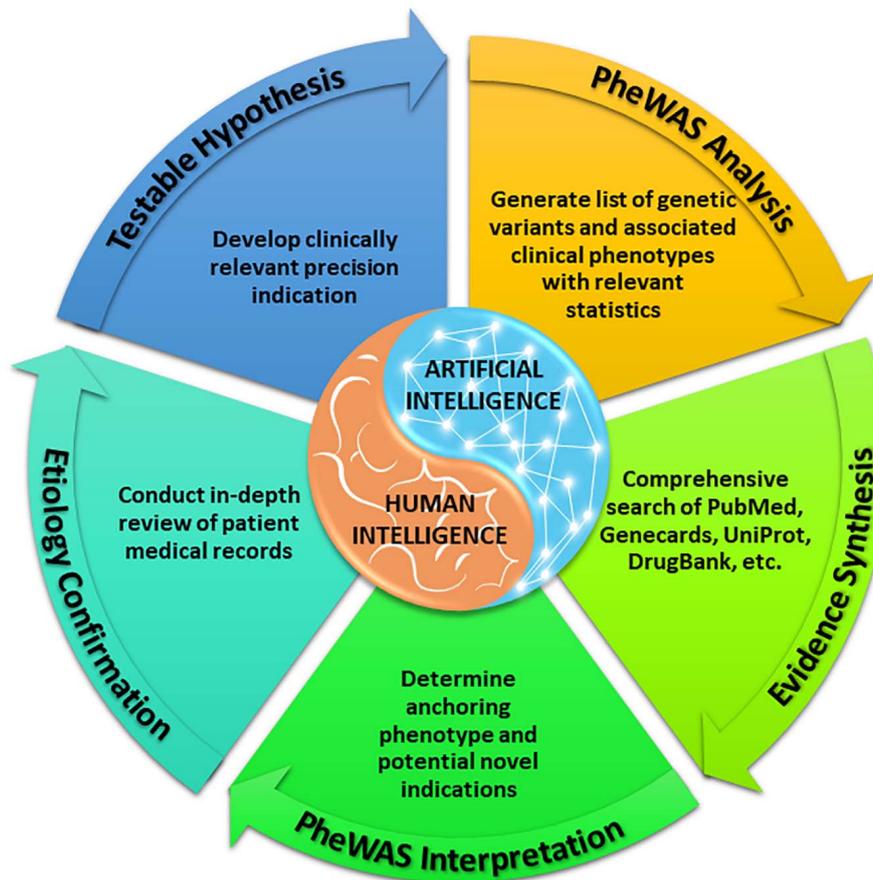


FIGURE 1 | VUMC Drug Repurposing Platform methods for establishing precision indications.

identify potentially novel indications, and pair drugs and SNPs with inference regarding mode of action for a SNP effect (e.g., inhibition, activation, agonism, or antagonism). However, SNP function, especially for rare variants, is often not known. Because ML model performance is poor in the presence of data missingness, automating this critical component of our workflow is not currently feasible. Our anchoring work thus plays a key role in allowing us to infer SNP function in the absence of previous evaluation of a variant's effect. Lastly, depending on the database or the literature source, SNP directionality is often inverted (i.e., which allele is considered “minor” or more rare as well as risk causing, which varies by population), and therefore can be misinterpreted in the exact opposite way it should be. Such catastrophic failure is unlikely to be identified by most ML frameworks.

Given that PheWAS results may contain false positives (i.e., statistically significant p -values that will not help the drug development process) and false negatives (i.e., weaker p -values that are truly meaningful to the process), automating the evaluation of PheWAS results based on p -values is unlikely to reliably identify the strongest repurposing indications, not to mention the multitude of cautionary tales and even the stance of the American Statistical Association against reliance on p -values

for decision making. One example of this problem is illustrated by our program to repurpose misoprostol for prevention of *Clostridioides difficile* recurrence, which has validating preclinical data (Zackular et al., 2019) and is currently in clinical testing (NCT03617172). Rather than arriving at our precision indication directly from a single PheWAS association, we worked from a cluster of individually weaker, but strongly related, associations with p -values that would not individually pass corrections for multiple testing. Combining data from across multiple analyses with awareness of diagnostic code overlap represents an area in need of methodological innovation (Strayer et al., 2020).

Phenome-wide association study relies heavily on diagnostic codes. Diagnostic codes, and thus their corresponding phenotypes, are not always self-explanatory, may vary in usage patterns based on local practice, may require assessment dependent on other ontologies, and often warrant consultation with clinicians for interpretation. For example, our program to repurpose ifetroban for prevention of metastases in multiple cancer types originated with a cluster of significant associations with several “secondary cancers.” Manual review of patient charts revealed that secondary cancer codes were being used to signify metastatic spread originating from various primary tumor sites, which was essential in developing our repurposing indication;

TABLE 1 | Example projects, Vanderbilt University Medical Center Drug Repurposing Platform.

Drug	Gene SNP Mutation effect	Current use	Anchoring phenotype (p -value)	Novel phenotype (p -value)	Proposed FDA indication	Animal validation
Misoprostol	<i>PTGER2</i> rs139552094 C83G ↓ receptor signaling	Gastric ulcers; obstetrics	Ulcer of esophagus (1.1E-02)	Ulcerative colitis (2.4E-02)	Prevention of recurrent <i>C. difficile</i> colitis (oral)	Misoprostol increased survival in mice infected with <i>C. difficile</i>
Ifetroban	<i>TBX42R</i> rs200445019 T399A ↑ receptor activity	Cardiovascular	Chronic venous hypertension (9.1E-06)	Metastatic malignancies (3.8E-03)	Prevention of metastasis in cancers with high recurrence	Ifetroban decreased metastases in animal models of three different cancers (breast, pancreatic, and lung)
Fluoxetine	<i>SLC6A4</i> rs6355 Gly56Ala ↑ transporter activity	Psychiatric	Antisocial/borderline personality disorder (2.4E-02)	Non-autoimmune hemolytic anemias (7.7E-04)	Shiga toxin-producing <i>Escherichia coli</i> hemolytic uremic syndrome	Fluoxetine increases survival of mice with Shiga toxin-producing <i>E. coli</i>
Enoxaparin	<i>F10</i> rs3211783 Gly192Arg predicted ↑ F10 activity	Thrombo-embolism	Other diseases of blood and blood forming organs (1.2E-03)	Cluster of infection phenotypes: Mycoses (7.8E-07)	<i>Acinetobacter baumannii</i> infection	Enoxaparin reduces <i>A. baumannii</i> bacterial burdens in B6 mice

Bold text highlights key phenotypes.

this program also has validating preclinical data (Werfel et al., 2020) and is currently in clinical testing (NCT03694249). If we had relied only on ML to construct our indication, would a rules-based classifier have recognized this rare sub-population who appear predisposed to metastases, when these diseases are coded as “secondary cancers?”

An additional shortcoming of diagnostic code-based clinical datasets is that a single disease code may represent several disease endotypes with vastly different etiologies. This is the case for our chronic fatigue syndrome (CFS) repurposing program, in which our treatment strategy would correct a disruption in norepinephrine transport. This hypothesis was derived from a PheWAS signal in the gene that encodes the norepinephrine transporter, and data from our completed biomarker study (NCT03029377) that suggest that this disruption is present in a subset of CFS patients. If using automated approaches for hypothesis generation or evidence synthesis, it would be difficult to arrive at a repurposing indication that falls under an “umbrella disease” and delineate which evidence is relevant for the specific subset of the disease from both mechanism and symptomology perspectives.

Our evidence synthesis workflow relies on manual review of publicly available gene expression data. The data can be misleading and lack specificity regarding how expression may be altered in disease states, differ based on population demographics, and vary in organ systems or by gene subunit. In addition, some proteins have tissue- or cell-specific effects that can change from positive to negative by tissue or cell, whereas others do not. This complexity of directionality and context

is difficult to assess and integrate into a scalable and coherent repurposing hypothesis without significant human scientific expert interpretation. Perhaps ML methods can be trained for this purpose, but the scope of deducing such relational knowledge at industrial scale currently precludes the automation of this step in the drug discovery workflow.

Several other issues hinder automation of literature review. For example, biologic assay specificity, in terms of measurement and whether protein subunits can be detected, is sometimes not adequately described in papers; in several repurposing projects, limitations in available descriptions of enzyme-linked immunosorbent assay (ELISA) methods have required that we directly consult with ELISA vendors to assess results and relevance, in some instances uncovering notable flaws in the published literature related to human disease-related inferences. The relevance of published animal model data may also vary based on homology against a human comparator which itself varies based on the specific disease being studied. Given that some published literature contains weak study designs and/or incoherent findings, even drugs that proclaim a mechanism of effect may be found erroneous in subsequent investigations. In such situations, the inaccurate mechanism remains in the published literature and the chance of excluding it using automated methods is currently low (McEwen et al., 2010). Furthermore, a wealth of data relevant for drug repurposing is only minimally published. Additionally, there can be serious data quality issues where publicly available databases just contain completely incorrect information. A previous version of DrugBank listed ACE2 as a target of moexipril, a target of

interest in COVID-19; however, when we manually reviewed the references supporting this claim none of them contained any direct evidence that moexipril had any impact on ACE2; this has since been corrected. Finally, much drug development data are proprietary and can only be gleaned by talking directly with members of original clinical development teams; such information includes the full dose range clinically tested and comprehensive understanding of biological effects and safety profiles observed over that range. Because of this, some of the richest repurposing data are unstructured.

DISCUSSION

Our experience is that ML models are currently unable to handle the complexity inherent to key repurposing information sources. While advances in natural language processing have been helpful in extracting unstructured and semi structured data, the complexity of data structures and sources used in the repurposing pipeline do not lend themselves to such straightforward solutions. Sometimes the judgment call about a given compound's selectivity (or lack thereof) and how well suited the compound is for a new use balanced against other (known) pharmacodynamic, pharmacokinetics, and safety concerns is just not straightforward and requires critical thinking, intuition, and creativity to synthesize existing knowledge into a new hypothesis with a high probability of success and low likelihood of propagating errors that exist in the historical record. The deep utility and value of engaging humans in evaluating PheWAS results, placing them within the broader scientific context, weaving together various and disparate sources of data, and identifying opportunities for confirmatory scientific investigations should not be underestimated. While current ML and other AI techniques can substantially complement these efforts by helping to identify meaningful patterns for validation, it is premature to consider replacing our human-focused approach. Identifying those aspects of the pipeline that require human insights can identify those problems amenable to algorithmic innovation to accelerate the pace of novel drug discovery.

Repurposing of existing therapeutic agents has generated significant interest in the realm of rare disease treatment, as an innovative strategy for finding ways to manage these complex conditions. In conversations with pharmaceutical industry collaborators, we often receive questions about the optimal balance of machine and human decision-making necessary to support the pipeline. This question continues to intensify given the growing availability of largescale EHR data combined with genotyping that may be leveraged for drug development. By enabling precision phenotyping and connections with genomic variation, these large datasets present a particularly exciting opportunity for exploring repurposing opportunities for rare diseases. Given the importance of hypothesis generation and the dense, complex nature of critical pre-clinical data in drug development, human knowledge remains essential for interpretation and making “go-no go” decisions with which machines continue to struggle. Much of the technology necessary

to automate the complete hypothesis generation and evidence synthesis processes is not yet ready or created, but by paying careful attention to the problems that require human insight we have highlighted those areas where methodological innovation might have greatest impact on the pace of discovery.

Helping us identify previously unrecognized patterns that we can then rationalize through systematic thinking and generational of a testable hypothesis is the significant value add of ML in our current workflow. It is not a replacement for the perspectives of domain experts; instead it is best leveraged in drug repurposing efforts when it is informed by content experts and then considered synergistically with human decision-making to pursue top repurposing leads through randomized controlled trials.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors made substantive contributions to conception of this perspective and/or participated in the acquisition, analysis, and interpretation of data for the work. APC, NMZ, RNJ, RRL, JS-R, IAM, KJH, and JMP participated in drafting the work. APC, NMZ, RNJ, RRL, JS-R, AB, CJL, DMA, LJC, RCH, TAI, KJH, and JMP revised the perspective critically for key content. All authors approved the manuscript and agreed to be accountable for all aspects of the work.

FUNDING

The project described was supported by CTSA award Nos. UL1 TR002243 and U54TR02243-02 from the National Center for Advancing Translational Sciences. APC and DMA were also supported to study the role of PheWAS in drug development by award R21HD105304 from the National Institute of Child Health and Development. AB was supported by award K08AR072757-01 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The contents of this article were solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences, the National Institute of Child Health and Development, the National Institute of Arthritis and Musculoskeletal and Skin Diseases, or the National Institutes of Health.

ACKNOWLEDGMENTS

The authors extend their gratitude to Meghan Vance for her graphic design assistance in support of this manuscript.

REFERENCES

- Álvarez-Machancoses, Ó, and Fernández-Martínez, J. L. (2019). Using artificial intelligence methods to speed up drug discovery. *Expert Opin. Drug Discov.* 14, 769–777. doi: 10.1080/17460441.2019.1621284
- Berthier, M. L., Pulvermüller, F., Dávila, G., Casares, N. G., and Gutiérrez, A. (2011). Drug therapy of post-stroke aphasia: a review of current evidence. *Neuropsychol. Rev.* 21, 302–317. doi: 10.1007/s11065-011-9177-7
- Brasil, S., Pascoal, C., Francisco, R., Dos Reis Ferreira, V., Videira, P. A., and Valadão, A. G. (2019). Artificial intelligence (AI) in rare diseases: is the future brighter? *Genes* 10:E978. doi: 10.3390/genes10120978
- Centers for Disease Control and Prevention (2021). *ICD - ICD-10-CM - International Classification of Diseases, Tenth Revision, Clinical Modification*. Available online at: <https://www.cdc.gov/nchs/icd/icd10cm.htm> (accessed June 22, 2021).
- Challa, A. P., Beam, A. L., Shen, M., Peryea, T., Lavieri, R. R., Lippmann, E. S., et al. (2020). Machine learning on drug-specific data to predict small molecule teratogenicity. *Reprod. Toxicol.* 95, 148–158. doi: 10.1016/j.reprotox.2020.05.004
- Challa, A. P., Lavieri, R. R., Lewis, J. T., Zaleski, N. M., Shirey-Rice, J. K., Harris, P. A., et al. (2018). “Systematically Prioritizing Targets in Genome-Based Drug Repurposing,” in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, (New York, NY: ACM), 543–543.
- Challa, A. P., Lavieri, R. R., Lewis, J. T., Zaleski, N. M., Shirey-Rice, J. K., Harris, P. A., et al. (2019). Systematically prioritizing candidates in genome-based drug repurposing. *Assay Drug Dev. Technol.* 17, 352–363. doi: 10.1089/adt.2019.950
- Challener, C. (2018). *Facilitating Drug Repositioning with Artificial Intelligence*. Available online at: <https://www.pharmasalmanac.com/articles/facilitating-drug-repositioning-with-artificial-intelligence> (accessed September 9, 2019).
- Chen, Z., Liu, X., Hogan, W., Shenkman, E., and Bian, J. (2021). Applications of artificial intelligence in drug development using real-world data. *Drug Discov. Today* 26, 1256–1264. doi: 10.1016/j.drudis.2020.12.013
- Choby, J. E., Monteith, A. J., Himmel, L. E., Margaritis, P., Shirey-Rice, J. K., Puijssers, A., et al. (2019). A phenome-wide association study uncovers a pathological role of coagulation factor X during *Acinetobacter baumannii* infection. *Infect. Immun.* 87:e00031-19. doi: 10.1128/IAI.00031-19
- Delavan, B., Roberts, R., Huang, R., Bao, W., Tong, W., and Liu, Z. (2018). Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discov. Today* 23, 382–394. doi: 10.1016/j.drudis.2017.10.009
- Denny, J. C., Bastarache, L., Ritchie, M. D., Carroll, R. J., Zink, R., Mosley, J. D., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31, 1102–1110. doi: 10.1038/nbt.2749
- Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., et al. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26, 1205–1210. doi: 10.1093/bioinformatics/btq126
- Ekins, S., Puhl, A. C., Zorn, K. M., Lane, T. R., Russo, D. P., Klein, J. J., et al. (2019). Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* 18, 435–441. doi: 10.1038/s41563-019-0338-z
- Goldstein, J. A., Bastarache, L. A., Denny, J. C., Roden, D. M., Pulley, J. M., and Aronoff, D. M. (2018). Calcium channel blockers as drug repurposing candidates for gestational diabetes: mining large scale genomic and electronic health records data to repurpose medications. *Pharmacol. Res.* 130, 44–51. doi: 10.1016/j.phrs.2018.02.013
- Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., and Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers.* [Epub ahead of print]. doi: 10.1007/s11030-021-10217-3
- Hermann, E. R., Chambers, E., Davis, D. N., Montgomery, M. R., Lin, D., and Chohanadisai, W. (2021). Brain magnetic resonance imaging phenome-wide association study with metal transporter gene SLC39A8. *Front. Genet.* 12:647946. doi: 10.3389/fgene.2021.647946
- Issa, N. T., Stathias, V., Schürer, S., and Dakshanamurthy, S. (2021). Machine and deep learning approaches for cancer drug repurposing. *Semin. Cancer Biol.* 68, 132–142. doi: 10.1016/j.semcancer.2019.12.011
- Jarada, T. N., Rokne, J. G., and Alhaji, R. (2020). A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J. Cheminform.* 12:46. doi: 10.1186/s13321-020-00450-7
- Jerome, R. N., Joly, M. M., Kennedy, N., Shirey-Rice, J. K., Roden, D. M., Bernard, G. R., et al. (2020). Leveraging human genetics to identify safety signals prior to drug marketing approval and clinical use. *Drug Saf.* 43, 567–582. doi: 10.1007/s40264-020-00915-6
- Jerome, R. N., Pulley, J. M., Roden, D. M., Shirey-Rice, J. K., Bastarache, L. A., R Bernard, G., et al. (2018). Using human “experiments of nature” to predict drug safety issues: an example with PCSK9 inhibitors. *Drug Saf.* 41, 303–311. doi: 10.1007/s40264-017-0616-0
- Kim, J., Kusko, R., Zeskind, B., Zhang, J., and Escalante-Chong, R. (2021). A primer on applying AI synergistically with domain expertise to oncology. *Biochim. Biophys. Acta Rev. Cancer* 1876:188548. doi: 10.1016/j.bbcan.2021.188548
- Le, N. Q. K., Do, D. T., Chiu, F.-Y., Yapp, E. K. Y., Yeh, H.-Y., and Chen, C.-Y. (2020). XGBoost improves classification of MGMT promoter methylation status in IDH1 wildtype glioblastoma. *J. Pers. Med.* 10:E128. doi: 10.3390/jpm10030128
- Le, N. Q. K., Yapp, E. K. Y., Nagasundaram, N., and Yeh, H.-Y. (2019). Classifying promoters by interpreting the hidden information of DNA sequences via deep learning and combination of continuous FastText N-grams. *Front. Bioeng. Biotechnol.* 7:305. doi: 10.3389/fbioe.2019.00305
- Lee, Y.-S., Krishnan, A., Oughtred, R., Rust, J., Chang, C. S., Ryu, J., et al. (2019). A computational framework for genome-wide characterization of the human disease landscape. *Cell Syst.* 8, 152.e6–162.e6. doi: 10.1016/j.cels.2018.12.010
- McEwen, B. S., Chattarji, S., Diamond, D. M., Jay, T., Reagan, L., Svenningsson, P., et al. (2010). The neurobiological properties of tianeptine (Stablon): from monoamine hypothesis to glutamatergic modulation. *Mol. Psychiatry* 15, 237–249. doi: 10.1038/mp.2009.80
- Nosengo, N. (2016). Can you teach old drugs new tricks? *Nature* 534, 314–316. doi: 10.1038/534314a
- PheWAS (2021). *PheWAS/PheWAS*. Available online at: <https://github.com/PheWAS/PheWAS> (accessed June 18, 2021).
- Pulley, J. M., Jerome, R. N., Ogletree, M. L., Bernard, G. R., Lavieri, R. R., Zaleski, N. M., et al. (2018a). Motivation for Launching a Cancer Metastasis Inhibition (CMI) Program. *Target Oncol.* 13, 61–68. doi: 10.1007/s11523-017-0542-1
- Pulley, J. M., Jerome, R. N., Shirey-Rice, J. K., Zaleski, N. M., Naylor, H. M., Puijssers, A. J., et al. (2018b). Advocating for mutually beneficial access to shelved compounds. *Future Med. Chem.* 10, 1395–1398. doi: 10.4155/fmc-2018-0090
- Pulley, J. M., Rhoads, J. P., Jerome, R. N., Challa, A. P., Erreger, K. B., Joly, M. M., et al. (2020). Using what we already have: uncovering new drug repurposing strategies in existing omics data. *Annu. Rev. Pharmacol. Toxicol.* 60, 333–352. doi: 10.1146/annurev-pharmtox-010919-023537
- Pulley, J. M., Shirey-Rice, J. K., Lavieri, R. R., Jerome, R. N., Zaleski, N. M., Aronoff, D. M., et al. (2017). Accelerating precision drug development and drug repurposing by leveraging human genetics. *Assay Drug Dev. Technol.* 15, 113–119. doi: 10.1089/adt.2016.772
- Roden, D. M., Pulley, J. M., Basford, M. A., Bernard, G. R., Clayton, E. W., Balser, J. R., et al. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* 84, 362–369. doi: 10.1038/clpt.2008.89
- Roessler, H. I., Knoers, N. V. A. M., van Haelst, M. M., and van Haaften, G. (2021). Drug repurposing for rare diseases. *Trends Pharmacol. Sci.* 42, 255–267. doi: 10.1016/j.tips.2021.01.003
- Roundtable on Translating Genomic-Based Research for Health, Board on Health Sciences Policy, Institute of Medicine (2014). *Drug Repurposing and Repositioning: Workshop Summary*. Washington, DC: National Academies Press.
- Salvatore, M., Gu, T., Mack, J. A., Prabhu Sankar, S., Patil, S., Valley, T. S., et al. (2021). A phenome-wide association Study (PheWAS) of COVID-19 outcomes by race using the electronic health records data in michigan medicine. *J. Clin. Med.* 10:1351. doi: 10.3390/jcm10071351

- Scherman, D., and Fetro, C. (2020). Drug repositioning for rare diseases: knowledge-based success stories. *Therapie* 75, 161–167. doi: 10.1016/j.therap.2020.02.007
- Schneider, C. V., Kleinjans, M., Fromme, M., Schneider, K. M., and Strnad, P. (2021). Phenome-wide association study in adult coeliac disease: role of HLA subtype. *Aliment Pharmacol. Ther.* 53, 510–518. doi: 10.1111/apt.16206
- Strayer, N., Shirey-Rice, J. K., Shyr, Y., Denny, J. C., Pulley, J. M., and Xu, Y. (2020). PheWAS-ME: a web-app for interactive exploration of multimorbidity patterns in PheWAS. *Bioinformatics* [Epub ahead of print]. doi: 10.1093/bioinformatics/btaa870
- Valencic, E., Smid, A., Jakopin, Z., Tommasini, A., and Mlinaric-Rascan, I. (2018). Repositioning drugs for rare immune diseases: hopes and challenges for a precision medicine. *Curr. Med. Chem.* 25, 2764–2782. doi: 10.2174/0929867324666170830101215
- Werfel, T. A., Hicks, D. J., Rahman, B., Bendeman, W. E., Duvernay, M. T., Maeng, J. G., et al. (2020). Repurposing of a thromboxane receptor inhibitor based on a novel role in metastasis identified by phenome-wide association study. *Mol. Cancer Ther.* 19, 2454–2464. doi: 10.1158/1535-7163.MCT-19-1106
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- World Health Organization (1993). *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. Geneva: World Health Organization.
- Xue, H., Li, J., Xie, H., and Wang, Y. (2018). Review of drug repositioning approaches and resources. *Int. J. Biol. Sci.* 14, 1232–1244. doi: 10.7150/ijbs.24612
- Zackular, J. P., Kirk, L., Trindade, B. C., Skaar, E. P., and Aronoff, D. M. (2019). Misoprostol protects mice against severe *Clostridium difficile* infection and promotes recovery of the gut microbiota after antibiotic perturbation. *Anaerobe* 58, 89–94. doi: 10.1016/j.anaerobe.2019.06.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Challa, Zaleski, Jerome, Lavieri, Shirey-Rice, Barnado, Lindsell, Aronoff, Crofford, Harris, Alp Ikizler, Mayer, Holroyd and Pulley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.