

# Accelerating Precision Drug Development and Drug Repurposing by Leveraging Human Genetics

Jill M. Pulley,<sup>1</sup> Jana K. Shirey-Rice,<sup>1</sup> Robert R. Lavieri,<sup>1</sup> Rebecca N. Jerome,<sup>1</sup> Nicole M. Zaleski,<sup>1</sup> David M. Aronoff,<sup>2,3</sup> Lisa Bastarache,<sup>4</sup> Xinnan Niu,<sup>4</sup> Kenneth J. Holroyd,<sup>1,5</sup> Dan M. Roden,<sup>6</sup> Eric P. Skaar,<sup>7</sup> Colleen M. Niswender,<sup>8,9</sup> Lawrence J. Marnett,<sup>10,11</sup> Craig W. Lindsley,<sup>10,12</sup> Leeland B. Ekstrom,<sup>1,13</sup> Alan R. Bentley,<sup>5</sup> Gordon R. Bernard,<sup>1,6</sup> Charles C. Hong,<sup>14,15</sup> and Joshua C. Denny<sup>4</sup>

<sup>1</sup>Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee.

<sup>2</sup>Department of Medicine, Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, Tennessee.

<sup>3</sup>Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee.

<sup>4</sup>Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee.

<sup>5</sup>Center for Technology Transfer and Commercialization, Vanderbilt University, Nashville, Tennessee.

<sup>6</sup>Office of Research, Vanderbilt University Medical Center, Nashville, Tennessee.

<sup>7</sup>Department of Pathology, Microbiology, and Immunology, Division of Molecular Pathogenesis, Vanderbilt University School of Medicine, Nashville, Tennessee.

<sup>8</sup>Department of Pharmacology, Vanderbilt Center for Neuroscience Drug Discovery, Vanderbilt University Medical Center, Nashville, Tennessee.

<sup>9</sup>Vanderbilt Kennedy University Center for Excellence in Development Disabilities, Nashville, Tennessee.

<sup>10</sup>Department of Chemistry, Vanderbilt University, Nashville, Tennessee.

<sup>11</sup>Department of Biochemistry, Vanderbilt University Medical School, Nashville Tennessee.

<sup>12</sup>Department of Medicinal Chemistry, Vanderbilt Center for Neuroscience Drug Discovery, Vanderbilt University Medical Center, Nashville, Tennessee.

<sup>13</sup>Nashville Biosciences, Nashville, Tennessee.

<sup>14</sup>Department of Medicine, Division of Cardiovascular Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee.

<sup>15</sup>Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, Tennessee.

## ABSTRACT

The potential impact of using human genetic data linked to longitudinal electronic medical records on drug development is extraordinary; however, the practical application of these data necessitates some organizational innovations. Vanderbilt has created resources such as an easily queried database of >2.6 million de-identified electronic health records linked to BioVU, which is a DNA biobank with more than 230,000 unique samples. To ensure these data are used to maximally benefit and accelerate both de novo drug discovery and drug repurposing efforts, we created the Accelerating Drug Development and Repurposing Incubator, a multidisciplinary think tank of experts in various therapeutic areas within both basic and clinical science as well as experts in legal, business, and other operational domains. The Incubator supports a diverse pipeline of drug indication finding projects, leveraging the natural experiment of human genetics.

**Keywords:** PheWAS, BioVU, genetics, translational research, repurposing

## INTRODUCTION

### We Desperately Need More Therapeutic Options

Although science has uncovered the molecular basis for ~5,700 human medical conditions, patients can only benefit from approved therapies for about 500 of these conditions.<sup>1</sup> Even many approved, available therapies have suboptimal efficacy or unacceptable toxicity in many patient populations. This problem is not limited to ultra-rare diseases. Diseases such as Parkinson's, Alzheimer's, addictions, autoimmune diseases, and many different cancers still have few or no effective treatment options available. Translational research seeks to bridge the "valley of death" between basic scientists and clinical researchers to generate the breakthrough therapies that are needed to improve human health. Pharmaceutical companies used to be responsible for carrying discoveries across this divide, but over the past few decades it has become increasingly difficult for them to keep

up with the sheer volume of drug targets and other findings coming out of the biomedical research enterprise.<sup>2</sup>

The systemic, chronic challenges faced by the pharmaceutical and biotech industries are well documented.<sup>3,4</sup> The advent of new technologies, such as high-throughput screening, induced pluripotent stem cells, clustered regularly interspaced short palindromic repeats (CRISPR) gene editing, proteomics, and next-generation sequencing, appears to have not had, as of yet, a significant impact on drug development. However, as detailed later, many major pharmaceutical companies have incorporated genomics into their development pipelines, and it is too early to evaluate the impact of genomics on many current clinical development programs. More specifically, there has been little effect on the number of drugs entering clinical development or, more importantly, on the success rate of drugs with novel mechanisms of action making it to the market, which still hovers around 6% overall for new chemical entities.<sup>5</sup> A fundamental problem in this paradigm is that pharmaceutical companies and others pursuing development and marketing approval face hundreds of decisions that must be made with severely limited information. The amount of time and money spent (and risk taken before any return on investment is ever seen) during both preclinical drug discovery and clinical drug development has increased in recent years.<sup>6</sup> These decisions remain difficult and costs remain high because there is often a fundamental lack of understanding of the human biology underpinning a disease process. This often results in an inability to efficiently target pathways that will lead to safe and effective therapies.

## MATERIALS AND METHODS

### Expanding the Therapeutic Armamentarium

The concept of leveraging human genetics to support target validation in drug discovery and drug development has permeated the pharmaceutical industry.<sup>5</sup> A recent review by Nelson *et al.* suggests that pursuing targets that are supported by human genetics could double the success rate in clinical development.<sup>7</sup> A number of pharmaceutical companies have made substantial investments in human genetics as a way to enhance their drug development efforts: Amgen acquired DeCODE genetics for \$415 million in 2012,<sup>8</sup> and Regeneron's Genetics Center was announced in January 2014.<sup>9</sup> One specific example of human genetics supporting target validation that ultimately led to FDA approval of agents targeting the gene of interest is the well-known story of *PCSK9*, in which it was discovered that individuals with very low cholesterol levels carried nonsense mutations in the *PCSK9* gene.<sup>10,11</sup>

The use of de-identified human genetic data tied to robust, de-identified electronic medical records, described later, has

powerful implications for predicting the potential effects of a pharmacological intervention in humans, before any clinical trial (or preclinical discovery program) is ever initiated. For a thorough overview of this topic, the proceedings from a workshop held by the National Academies of Sciences, Engineering, and Medicine in March 2016 are available online.<sup>12</sup> The core asset we are leveraging for hypothesis generation in our approach is BioVU, a large repository of de-identified DNA samples that Vanderbilt University Medical Center has assembled over >10 years from (what otherwise would have been) excess, discarded patient blood samples collected during routine clinical testing. This biobank is a centralized resource for investigating genotype-phenotype associations that has enabled large-scale, innovative research. Biospecimens within BioVU are linked to corresponding clinical and demographic data derived from the Synthetic Derivative, our de-identified database of electronic health records (EHRs) for research purposes.<sup>13-16</sup>

More recently, the phenome-wide association study (PheWAS) has been introduced as a systematic and efficient approach to elucidate novel disease-variant associations and pleiotropy by using BioVU.<sup>17</sup> In contrast to traditional methods that can be used to identify genes that are associated with specific diseases, PheWAS can be used to identify the diseases that are associated with a specific gene product (protein, which is the potential drug target).<sup>17-20</sup> It is the comprehensive and diverse nature of the diagnostic information within EHRs that enables PheWAS.

### A New Approach That Is Simultaneously *In Vivo* and *In Silico*

It is widely known that many, but not all, existing *in vitro* and preclinical animal models of disease are extremely poor predictors of human biology.<sup>21</sup> A better way to predict drug effects in humans, in terms of both efficacy and potential adverse effects, is clearly needed. In the method described here, we rely on natural human genetic variation as a proxy for—and method of more accurately predicting—the physiologic effects of therapies in humans. Briefly, we identify variants in drug target genes that recapitulate drug effects and then execute PheWAS to find all phenotypes (diseases) across the human phenome (based on PheWAS codes derived from ICD-9/10 codes and manually grouped to define clinical phenotypes) that are associated with carrying at least one copy of a minor allele.<sup>17,22</sup>

When a single nucleotide polymorphism (SNP) associated with a specific disease causes an amino acid change in an expressed protein and the mutant protein has been characterized either *in vitro* or *in vivo*, the interpretation of the data is relatively straightforward; however, there are many cases in which an SNP is either uncharacterized and/or synonymous, meaning a base pair change results in the presence of a different

codon that codes for the same amino acid. It would be altogether incorrect to dismiss synonymous SNPs as “uninteresting,” as their effects on protein expression levels, structure, and, ultimately, function can cause disease. Sauna and Kimchi-Sarfaty have reviewed the molecular mechanisms by which synonymous SNPs can cause disease and have provided a number of excellent examples.<sup>23</sup> Unlike computer modeling, our work uses actual human diseases and markers of their pathophysiology as diagnosed and recorded directly within the clinical setting, and, in a sense, is a human library of drug targets. These methods can be extended to predict phenotypic manifestations of the pharmacological targeting of a given protein in humans.

In addition, results obtained from a PheWAS analysis may also suggest deleterious effects, thus allowing for early insight into potential (on-target) adverse events or patient populations that should later be excluded from a clinical trial. Finally, this approach can be leveraged to identify potential indications or serious adverse events for yet-to-be discovered drug candidates with defined “druggable” targets, or for new compounds with opposite effects of known drugs. Establishing contributory human biology with regard to a discrete molecular target is a critical component of this work.<sup>24</sup> We propose that these methods will accelerate the pace of drug development through more precise and rapid indication identification.

### The Application of Biomedical Informatics to Drug Repurposing

It has been postulated and stands to reason that drug repurposing can increase efficiency by offering time and cost benefits in clinical development compared with traditional methods, because repurposing/repositioning candidates have often been through several stages of clinical development and have established safety and pharmacokinetic profiles. In this perspective, we define repurposing generally as any of the following:

- different indications for products (vs. those originally selected) that have been shelved due to late-stage efficacy failures, sometimes called rescuing
- secondary indications for generic products
- expanded indications for already marketed proprietary products
- more precise and targeted indications than would otherwise have been selected for therapies in clinical testing stages
- changing or adding indications to drugs already in pre-clinical or clinical development; and
- use of naturally occurring substances which are safe for human use

Drug repurposing is not a new activity<sup>25</sup>; however, the use of modern bioinformatics coupled with human genetics to

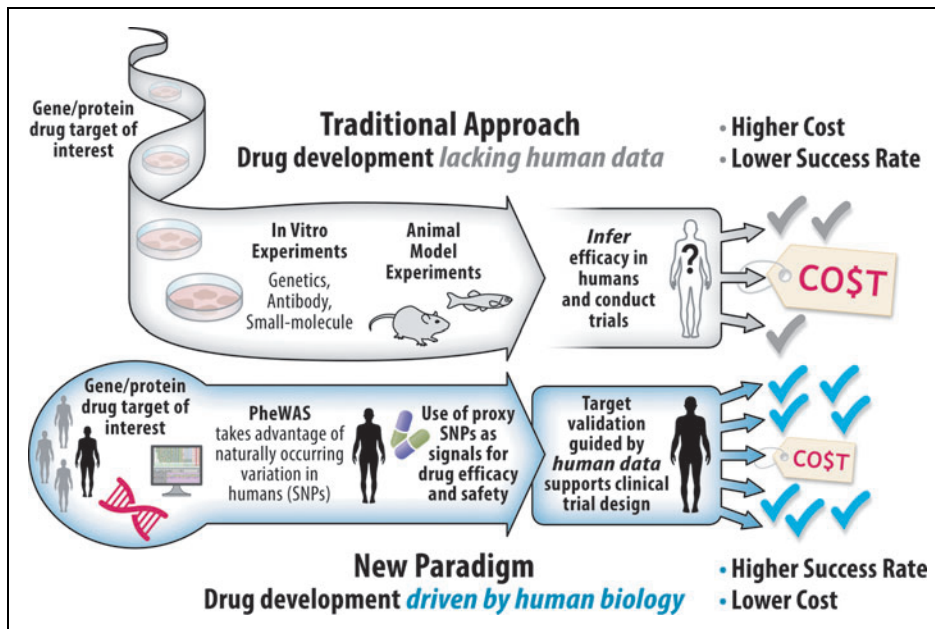
systematically guide and inform decision making in drug repurposing is a relatively new approach. We propose that with methods described here, there is an important place for repurposing efforts within a company’s portfolio, and the risk-mitigation afforded by repurposing could help address some of the economic challenges facing big pharma.<sup>3</sup> In addition, we believe that a focus on repurposing efforts will lead to more patients being treated with the drugs that they need significantly faster than more traditional approaches.

### The Role of the 505(b)(2) Regulatory Pathway in Drug Repurposing for New Indications

There are several regulatory pathways of relevance to repurposing: new chemical entity (NCE) route to FDA approval, the 505(b)(1) pathway, and the distinct 505(b)(2) route to FDA approval. The latter pathway is particularly attractive to organizations that are engaged in drug repurposing efforts for many reasons.<sup>26</sup> In the 505(b)(1) regulatory approval pathway for an NCE, the sponsor must conduct studies that allow for the preparation of full reports on safety and efficacy of the investigational new drug; this process typically takes as long as 15 years and can cost in excess of \$1 billion. By contrast, the 505(b)(2) regulatory approval pathway allows the sponsor to rely, in part, on existing data from approved drugs. This can eliminate the need for preclinical studies, shorten overall FDA approval time to less than 3 years, and reduce costs from >\$1 billion to tens of millions. In addition, unlike generic drugs approved under the 505(j) pathway (which only have 180 days of market exclusivity), products approved under the 505(b)(2) pathway can obtain a maximum of 7 years of market exclusivity.<sup>27,28</sup> This seems to be a healthy period of exclusivity relative to the substantially lower investment required to repurpose an approved drug. Moreover, since this approach can simultaneously identify patient subpopulations, perhaps those with rare genetic diseases who are more likely to respond favorably to a given therapeutic intervention, this approach can expedite targeted clinical trials for Fast Track review of drugs for serious unmet medical needs.

### A Start-Up Initiative to Accelerate Repurposing

*A promising disease agnostic infrastructure: an engine for generation of new testable hypotheses.* Drug repurposing and indication finding, based on human data, aims at better targeting or expanding the use of drugs that are currently on the market or that have been abandoned due to failed efficacy, by identifying new therapeutic indications. Because repurposing builds on previous research and development efforts, candidate therapies can be quickly prepared for clinical trials, greatly accelerating their development timeline and eventual integration into new treatment paradigms. Furthermore, because new therapeutic



**Fig. 1.** Comparison of drug development approaches. Drug development is currently an extraordinarily costly endeavor with a very low success rate, much of which can be attributed to incorrect therapeutic hypotheses based on preclinical models that do not accurately predict human biology.<sup>21</sup> The ADDRI's work fits within the new paradigm of drug development, which holds the promise of improved time and cost efficiency compared with the traditional approach. Human data drive the entire process in the new paradigm, and this should lead to a higher likelihood of success. ADDRI, Accelerating Drug Development and Repurposing Incubator.

indications are based on real human data, this approach should improve the probability of success in late-stage clinical trials (Fig. 1). However, even if successful, it is not known how much more efficient the drug repurposing process is from start to finish. Repurposing could still require substantive, time-consuming safety testing if (e.g.); the new use of a drug necessitates that it be given to patients far outside the existing/approved patient population. As part of our program, we will conduct an assessment of the repurposing program's effect on the pace of achieving downstream regulatory and commercialization milestones compared with traditional methods (which are well documented to require a 10- to 20-year time period).

### Program Description

This program will extend our capacity to effectively direct a diverse pipeline of repurposing candidates, ensure the scientific validity and commercial viability of targets, and provide a mechanism for efficiently applying intellectual and other institutional resources toward drug repurposing. Several questions guide the evaluation of potential projects: (1) Stage: How advanced is the science (i.e., *in vivo* vs. *in vitro*, or concept stage)?; (2) Scientific validity: Is the concept valid based on an assessment by scientists/clinicians?; (3) Feasibility: Are there

suitable animal models, cell lines, or computational models available that can be used to assess efficacy?; (4) Safety: Are all early signs of safety clear? As previously alluded to, PheWAS data provide signals for the potentially deleterious consequences of chronic targeting of a gene product, which will be factored into our holistic evaluation of potential products. An important part of our new infrastructure is the validation of candidate genotype-phenotype associations that are identified based on known biology, literature associations, and clinical attractiveness. Staff with relevant qualifications and content-specific knowledge perform evidence reviews and curate the results for key associations. We execute comprehensive searches of a wide range of databases and other resources. Key facets of this targeted exploration include:

- identifying existing published results related to the gene, protein, SNP, and phenotype, including GWAS data and established resources for gene/phenotype associations (e.g., OMIM, ClinVar, MalaCards)
- analyzing data on the known and predicted functional impact of prioritized SNPs by using SIFT, PolyPhen, and Combined Annotation-Dependent Depletion
- evaluating the plausibility by analyzing the known and predicted involvement of the SNP, gene, and protein in the potential underlying biologic pathway of the phenotype
- assessing the peer-reviewed biomedical literature related to the SNP, gene, protein, and effects of modulation of the protein via genetic or pharmacologic methods to gauge and summarize the volume, strength, and nature of published evidence
- assessing the utility of known or predicted compounds for modulating the protein of interest; and
- gauging the potential clinical impact and treatment efficacy for the condition.

### Review Criteria

Additional considerations include other key characteristics related to the attractiveness of each repurposing opportunity, such as institutional bandwidth (e.g., resources and expertise) and whether there is a local project champion. Specific questions

for evaluating the relative commercial viability of repurposing concepts that we explore are as follows: (1) Patentability: Is the compound patentable? If patented, is there sufficient patent life left (*e.g.*, 7–10 years) or potential for a “use” patent? (2) Unmet medical need: Is there opportunity for a transformative impact on health? That is, are there sizeable gaps in the current landscape of relevant therapeutic strategies for some patient populations (*e.g.*, noticeable lack of efficacy or development of resistance over time, adverse effects precluding ongoing use)?; (3) Feasibility: How feasible are clinical trials (*e.g.*, Can a sufficient number of patients be recruited in a reasonable timeframe with well-defined end points)?; and (4) Burden of proof: Are there any approval challenges (*e.g.*, no drug previously approved for the specific condition/indication)? *A priori* termination points are identified in advance for situations when further pursuit of a repurposing candidate is unlikely to result in reasonable scientific or commercial gains. However, our process may also result in scientific discoveries supporting off-label use with limited commercial value, but with high potential patient impact.

### Illumina Infinium Exomechip Cohort Data Available Within BioVU

The Exomechip contains ~250,000 variants, which are largely across the protein coding region of the genome. The variants were discovered through exome and whole-genome sequencing in more than 12,000 individuals.<sup>29</sup> It was designed to serve as an intermediate step between current genotyping arrays, which are designed to study common variants, and exome sequencing, which can discover rare variants. Nearly all non-synonymous, splice, and stop-altering variants detected in an average genome through exome sequencing were included on the Exomechip. The Exomechip also includes SNPs with known disease associations as well as unpublished associations from consortia working on diabetes, blood lipids, blood pressure, lung function, myocardial infarction, anthropometric traits, psychiatric traits, Crohn’s disease, and age-related macular degeneration. Approximately 35,000 BioVU subjects have existent genotyping data on this platform. Now, queries can be done expeditiously, improving the cost efficiency of obtaining human information. In addition, genotyping efforts are currently underway to expand the number of genotyped individuals to >100,000 BioVU subjects on a new platform with >2 million SNPs per sample.

## RESULTS

### Demonstration Projects Within a Balanced Portfolio

We hypothesize that the phenomic data—that is, diseases that humans actually have (*e.g.*, as a result of genetic inhibi-

tion or activation of important proteins or enzymes)—will transform how drugs find the right indications, patient populations, and endpoints. This is done efficiently, as cases and controls for one disease are reused for a different unrelated disease. This approach has been shown to improve time and cost efficiency. Programs are launched in a matter of weeks and not years. The timeline for completion of validation studies is 6 months, and go/no-go decision making is applied. Later, we show several of the current pipeline projects and the relevant PheWAS summary data for illustrative purposes (*Table 1*). Animal model validation studies are in progress for many of these targets. Model system validation studies are underway for ten more indication-drug pairs (not shown). A proof-of-concept clinical trial is actively being planned for one project, with others anticipated to begin in 2017.

## DISCUSSION

### The Data Can Suggest Vastly Different Indications in Distinct Therapeutic Areas

As others have discussed, this type of data is not meant to replace animal models of disease or, more specifically, utilization of *in vitro* and *in vivo* systems to understand the biology of a given disease process, but rather, these data are intended to simplify, focus, and accelerate the use of *in vitro* and *in vivo* systems to understand the human biology of a given disease process.<sup>30</sup> To that end, the use of *in vitro* and *in vivo* models of human diseases to validate PheWAS associations, which can include diseases in vastly different therapeutic areas for a single gene/drug target, will remain relevant going forward. In addition, methods for development work are currently underway to apply PheWAS to gene networks, which will be particularly relevant to understanding how to combat diseases that are driven by multiple pathways; cancer is an obvious example.

### A Genetic Mutation or Loss-of-Function Mutation is not the Same as an “Acute” Exposure to a Pharmacological Agent

An important, fundamental limitation to the use of genetic data to support drug development is that a loss-of-function mutation carried over a lifetime is not the same physiological event as an acute (or even chronic) exposure to a pharmacological intervention. In the case of a loss-of-function mutation or even a drug exposure, other signaling pathways can (and often do) compensate as a response. Similarly, even if a drug exactly mimics the biochemistry of a loss-of-function variant, the drug still may not produce a clinical benefit in a trial, because the duration of use is insufficient to observe an effect.

**Table 1. Selected Drug Repurposing Projects**

Gene	SNP and Basic SNP Effect	Mutation	Example Drug	Validates Expected Indication (P)	New Possible Indication (P)	Total Cases	Controls	Cases with Minor Allele	Comments
TNF	rs3093662 ( <i>in vivo</i> validated function, lowers TNF levels)	Intronic	Adalimumab	Rheumatoid arthritis ( $8.9 \times 10^{-5}$ )	Cervical cancer (0.004)	156	12,061	10	Available for study
PLA2G7	rs145315433 ( <i>in vivo</i> validated function, increased enzyme activity)	F51I	Darapladib	Diabetes type 2 with peripheral circulatory disorders ( $8.4 \times 10^{-4}$ )	Glomerulonephritis ( $2.8 \times 10^{-5}$ )	277	21,929	6	Available for study
CACNB2	rs149253719 (inferred from data as activating)	S214T	Verapamil	Angina (0.001)	Adrenal hypofunction ( $8.4 \times 10^{-5}$ )	362	25,592	5	Animal testing in planning
GLYT1	rs6429644 (inferred from data as activating)	Intronic	Bitopertin	Schizophrenia and other psychotic disorders (0.005)	Cerebral degeneration (0.002) Psychosis (0.025)	157	22,458	34	Available for study
TBXA2R	rs200445019 ( <i>in vitro</i> validated gain of function)	T399A	Ifetroban	Chronic venous hypertension ( $9.1 \times 10^{-6}$ )	Pulmonary heart disease (0.003)	1,512	24,713	18	Commercial partnership
PTGER2	rs139552094 (predicted loss of function)	C83G	Misoprostol	Ulcer of esophagus (0.01)	GI inflammatory diseases (0.02 multiple related phenotypes)	238	20,989	4	Animal validation testing complete, provisional patent filed
PCSK9	rs11591147 (known loss of function)	R46L	Alirocumab	Hypercholesterolemia (0.004)	Data do not suggest other viable indications	NA	NA	NA	NA

We have genetic data for at least one SNP in each of the seven genes listed in column 1 that are associated with known phenotypes, which validates the association data at the level of a given SNP. In addition, we have shown, in column 6, novel phenotypes that are associated with the gain- or loss-of-function variants in the relevant genes. The column to the far right shows the current state of development for various drugs (if any exist) with respect to the relevant discrete molecular target. Notably, animal model experiments performed by groups both not at our institution and totally independent of these analyses have corroborated some of the findings displayed earlier. We are currently pursuing animal model validation experiments and small, human proof-of-concept studies for some of the targets listed earlier.

SNP, single-nucleotide polymorphism.

**Despite Great Improvements in Efficiency in Preclinical Discovery, Clinical Development (as Governed by FDA Regulations), Timelines, and Costs Remain the Same and Very Significant**

The cost to obtain approval for an NCE is estimated to be in the hundreds of millions to billions of dollars; even a 505(b)(2) application can cost tens of millions of dollars, and with thousands of diseases for which no effective drugs have been identified, this is an untenable situation. It is not surprising that the economic incentives, as they stand today, mean that many companies simply cannot afford to pursue developing drugs (including generics) which may be based on rigorous scientific evidence and even show signs of efficacy in humans.

Although we are focused on utilizing these bioinformatics-based methods to de-risk drug development and drug repurposing by providing target validation in humans before the first dose is ever given to a person, we cannot ignore the practical policy constraints and barriers to helping patients. The new drugs created by the pharmaceutical industry in a given year address only 10–30 diseases at most and at a staggering cost.<sup>31</sup> Historically, companies have not leveraged human genetic data early in target validation efforts; however, this is currently changing across the industry. If we can remove some of these costs and time from the overall system, by utilizing better information about predicted drug effects *sooner*,

then it is our hope that efficacious drugs will be available to us as patients more efficiently.

### ACKNOWLEDGMENTS

The authors are grateful for the participation of all individuals contributing to this research. Specifically, they are grateful for exceptionally strong institutional support from Vanderbilt University Medical Center and Vanderbilt University leadership. The project described herein is supported by the CTSA award No. UL1TR000445 from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

### DISCLOSURE STATEMENT

No competing financial interests exist.

### REFERENCES

1. Collins FS: Seeking a cure for one of the rarest diseases: progeria. *Circulation* 2016;134:126–129.
2. Butler D: Translational research: crossing the valley of death. *Nature* 2008;453:840–842.
3. Nosengo N: Can you teach old drugs new tricks? *Nature* 2016;534:314–316.
4. Ascher J, Bogdan B, Dreszer J, Zhou G: Pharma's next challenge. [www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/pharmas-next-challenge](http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/pharmas-next-challenge). (Last accessed September 27, 2016).
5. Plenge RM, Scolnick EM, Altshuler D: Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* 2013;12:581–594.
6. Mullin R: Cost to develop new pharmaceutical drug now exceeds \$2.5B. [www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/](http://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/). (Last accessed September 27, 2016).
7. Nelson MR, Tipney H, Painter JL, et al.: The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;47:856–860.
8. Koss A, Gilbert E, Sood A: Amgen to acquire deCODE genetics, a global leader in human genetics. <http://investors.amgen.com/phoenix.zhtml?c=61656&p=irol-newsArticle&ID=1765710>. (Last accessed September 27, 2016).
9. Regeneron Genetics Center: [www.regeneron.com/Regeneron-Genetics-Center](http://www.regeneron.com/Regeneron-Genetics-Center). (Last accessed September 27, 2016).
10. Cohen JC, Boerwinkle E, Mosley TH Jr., Hobbs HH: Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006;354:1264–1272.
11. Hall SS: Genetics: a gene of rare effect. *Nature* 2013;496:152–155.
12. Addie S, Wagner Gee A, Olson S, Beachy H: *Deriving Drug Discovery Value from Large-Scale Genetic Bioresources*. National Academies Press, Washington, DC, 2016.
13. Roden DM, Pulley JM, Basford MD, et al.: Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;84:362–369.
14. National Institutes of Health: NIH Genomic Data Sharing Policy. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html>. (Last accessed September 27, 2016).
15. Bowton E, Field JR, Wang S, et al.: Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med* 2014;6:234cm3.
16. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR: Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Transl Sci* 2010;3:42–48.
17. Denny JC, Ritchie MD, Basford MA, et al.: PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010;26:1205–1210.

18. Denny JC, Bastarache L, Ritchie MD, et al.: Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* 2013;31:1102–1110.
19. Shameer K, Denny JC, Ding K, et al.: A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum Genet* 2014;133:95–109.
20. Denny JC, Crawford DC, Ritchie MD, et al.: Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. *Am J Hum Genet* 2011;89:529–542.
21. Scannell JW, Blanckley A, Boldon H, Warrington B: Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* 2012;11:191–200.
22. Pendergrass SA, Brown-Gentry K, Dudek SM, et al.: The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet Epidemiol* 2011;35:410–422.
23. Sauna ZE, Kimchi-Sarfaty C: Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 2011;12:683–691.
24. Plenge RM: Disciplined approach to drug discovery and early development. *Sci Transl Med* 2016;8:349ps15.
25. Ashburn TT, Thor KB: Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;3:673–683.
26. Gaffney A: An increasing number of companies are using a once-obscure FDA Drug Approval pathway. [www.raps.org/Regulatory-Focus/News/2015/04/08/21933/An-Increasing-Number-of-Companies-Are-Using-a-Once-Obscure-FDA-Drug-Approval-Pathway/](http://www.raps.org/Regulatory-Focus/News/2015/04/08/21933/An-Increasing-Number-of-Companies-Are-Using-a-Once-Obscure-FDA-Drug-Approval-Pathway/). (Last accessed September 27, 2016).
27. Phelps KV: Taking the 505(b)(2) route. [www.dddmag.com/article/2012/08/taking-505b2-route](http://www.dddmag.com/article/2012/08/taking-505b2-route). (Last accessed September 27, 2016).
28. Food and Drug Administration: Guidance for Industry: Applications Covered by Section 505(b)(2). [www.fda.gov/downloads/Drugs/.../Guidances/ucm079345.pdf](http://www.fda.gov/downloads/Drugs/.../Guidances/ucm079345.pdf). (Last accessed September 27, 2016).
29. Infinium Exome-24 v1.0 BeadChip. [www.illumina.com/documents/products/datasheets/datasheet\\_humanexome\\_beadchips.pdf](http://www.illumina.com/documents/products/datasheets/datasheet_humanexome_beadchips.pdf). (Last accessed September 27, 2016).
30. Kent Lloyd KC, Robinson PN, MacRae CA: Animal-based studies will be essential for precision medicine. *Sci Transl Med* 2016;8:352ed12.
31. Bloom B: Repurposing social impact bonds for medicine. [https://ssir.org/articles/entry/repurposing\\_social\\_impact\\_bonds\\_for\\_medicine](https://ssir.org/articles/entry/repurposing_social_impact_bonds_for_medicine). (Last accessed September 27, 2016).

Address correspondence to:

Jill Pulley, MBA

Vanderbilt Institute for Clinical and Translational Research

Vanderbilt University Medical Center

2525 West End Avenue

Suite 600

Nashville, TN 37203

E-mail: [Jill.Pulley@Vanderbilt.edu](mailto:Jill.Pulley@Vanderbilt.edu)

### Abbreviations Used

BioVU	=	Vanderbilt's DNA biobank
EHRs	=	electronic health records
NCE	=	new chemical entity
PheWAS	=	phenome-wide association study
GWAS	=	genome-wide association studies
SIFT	=	scale-invariant feature transform
SNP	=	single-nucleotide polymorphism
OMIM	=	Online Mendelian Inheritance in Man
PolyPhen	=	Polymorphism Phenotyping