ANNUAL REVIEWS

Annual Review of Pharmacology and Toxicology

Harnessing the Power of Electronic Health Records and Genomics for Drug Discovery

Kristi Krebs and Lili Milani

Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia;
email: lili.milani@ut.ee

Annu. Rev. Pharmacol. Toxicol. 2023. 63:65–76

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

https://doi.org/10.1146/annurev-pharmtox-051421-111324

Keywords

electronic health records, genomics, GWAS, PheWAS, Mendelian randomization, drug targets

Abstract

A long-standing recognition that information from human genetics studies has the potential to accelerate drug discovery has led to decades of research on how to leverage genetic and phenotypic information for drug discovery. Established simple and advanced statistical methods that allow the simultaneous analysis of genotype and clinical phenotype data by genome- and phenome-wide analyses, colocalization analyses with quantitative trait loci data from transcriptomics and proteomics data sets from different tissues, and Mendelian randomization are essential tools for drug development in the postgenomic era. Numerous studies have demonstrated how genomic data provide opportunities for the identification of new drug targets, the repurposing of drugs, and drug safety analyses. With an increase in the number of biobanks that enable linking in-depth omics data with rich repositories of phenotypic traits via electronic health records, more powerful ways for the evaluation and validation of drug targets will continue to expand across different disciplines of clinical research.

INTRODUCTION

Comprehensive knowledge of therapeutic targets is essential for advancing drug discovery and developing better and safer therapeutic agents. Close to 10% of drugs from Phase I clinical trials end up being approved (1). Most fail in Phase II clinical trials, and approximately 75% of the failures are attributable to safety concerns or lack of proof of efficacy. Therapeutic agents approved between 2009 and 2018 had an estimated median cost of $985.3 million for bringing a drug to the market (2); thus, there is an obvious need for methods to increase the likelihood of success. Retrospective studies have indicated that pursuing drug targets with support from human genetics increases the likelihood of successful drug discovery at least twofold (3, 4).

Electronic health records (EHRs) have been increasingly utilized in genomic studies of diseases and have provided several opportunities for cost-effective and extensive research (5). During recent decades there has been an increase in the development and launch of biobanks, and the number of individuals genotyped or sequenced is growing rapidly. Linking biobanks to EHRs enables researchers to analyze thousands of samples and associate genetic profiles with several health outcomes. Furthermore, for the discovery of disease mechanisms and novel drug targets, an instrumental factor is meta-analysis across cohorts, which is facilitated by the large number of accessible biobanks all over the world and global initiatives that bring this information together (6).

Here we describe the major approaches for leveraging EHRs and genomics for a more in-depth process of drug discovery based on data from human studies. We describe how genomics and EHRs have been used for the discovery of new drug targets, for the repurposing of drugs, and for drug safety analysis.

METHODS FOR DRUG TARGET STUDIES BASED ON GENOMICS

Many different approaches and methods have been applied for leveraging both genomics and data in EHRs for drug discovery and safety studies (**Figure 1**).

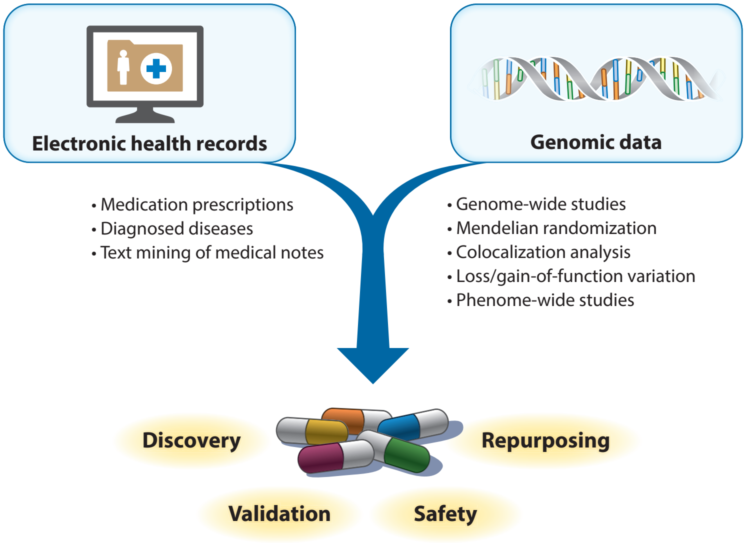
Electronic health records
Genomic data
• Medication prescriptions
• Diagnosed diseases
• Text mining of medical notes
• Genome-wide studies
• Mendelian randomization
• Colocalization analysis
• Loss/gain-of-function variation
• Phenome-wide studies
Discovery
Repurposing
Validation
Safety

Figure 1

Different methods for the analysis of data from electronic health records and genomics for drug research.

Genome-Wide Association Studies

Genome-wide association studies (GWASs) have added extensive knowledge to our understanding of drug targets and effects by enabling hypothesis-free and systematic analysis of variants across the entire genome. Further longitudinal collections of phenotypic information coupled with medication exposure data from EHRs create an important platform for the analysis of drug effects. A study exploring the results of 361 GWASs revealed that 62 of them pointed at a drug target developed for the same disease, and, furthermore, 92 genes were mapped to a GWAS trait different from their drug indications, indicating potential for therapeutic repurposing, including denosumab for the treatment of Crohn's disease (7). A recent article further demonstrated a genomics-driven drug-discovery framework using cross-population meta-analysis where the pipeline identified 144 drug/compound-disease pairs and demonstrated a catalog of candidate drugs for repositioning (8). For example, Namba et al. (8) highlighted possible drug candidates targeting coagulation-related genes for venous thromboembolism (VTE), of which PROC, F2, and F10 are already targeted by approved drugs, and KLKB1 and F11 are targeted by drugs that are in clinical trials for VTE.

Mendelian Randomization: Nature's Randomized Trials

Several studies have indicated that Mendelian randomization (MR) is a very promising tool for drug target studies. In MR, genetic variants are used as a proxy for exposure to assess the causal effect of this event on an outcome phenotype. It relies on the fact that genetic variants are randomly distributed to offspring, and therefore, evaluating the effects of genetic variants can function as nature's version of randomized controlled clinical trials. MR methods have been proven to be valuable in drug discovery, safety, and repurposing studies. MR methods have been used to investigate drug safety and risk for on-target adverse effects; for example, a recent study indicated that using genetic variants as a proxy of ACE inhibition was associated with an increased risk of colorectal cancer (9). One example of repurposing is a massive study where MR was used to study genetic variants in the IL-6 receptor (IL-6R) gene with the aim to assess the effect of IL-6R inhibition for primary prevention of coronary heart disease (10). Swerdlow et al. (10) found that a variant (rs7529229) in *IL-6R* was associated with increased circulating IL-6 concentration (an increase per allele of 9.45%) and reduced C-reactive protein and fibrinogen levels. In the analysis of 25,458 coronary heart disease cases they further demonstrated that the same variant was associated with decreased odds of coronary heart disease. Another study recently explored over 3,000 genes encoding druggable proteins to predict their potential as targets for Parkinson's disease and proposed 23 drug-targeting mechanisms with four possible drug-repurposing opportunities for metformin, rocuronium, roledumab, and warfarin (11).

Recent years have paved the way for studies of circulating proteins that are also attractive as potential drug targets. Using GWASs for the detection of protein quantitative trait loci (pQTL) can expose disease-association pathways and reveal novel drug targets (12). Furthermore, using MR to evaluate causality between protein levels and disease risk can support drug target validation and provide explanations for the mode of action of a drug. One important study demonstrating the power of pQTL studies in drug discovery was a genome-wide meta-analysis across 14 studies of 90 proteins associated with cardiovascular traits. Folkersen et al. (13) used MR for the identification of potential causal disease pathways and found 14 proteins that are already included in clinical drug development programs and 11 proteins that have not yet been targeted in clinical trials but are promising target candidates. For example, strong evidence of causality was identified for epidermal growth factor for the treatment of schizophrenia [Beta with 95% confidence interval (CI):

-0.31 (-0.50, -0.13); p-value 9.9×10^{-4}], pappalysin 1 (PAPPA) for type 2 diabetes [Beta with 95% CI: -0.27 (-0.42, -0.11); p-value 8.6×10^{-4}], and spondin 1 (SPON1) for atrial fibrillation [Beta with 95% CI: 0.14 (0.06, 0.20); p-value 1.2×10^{-6}]. SPON1 and PAPPA did not show any strong evidence of inverse causality with other phenotypes, suggesting high specificity for the intended indication.

In another important study where plasma protein levels were measured in 35,559 Icelanders, 938 genes encoding potential drug targets, with variants that influence levels of possible biomarkers for several diseases, were discovered (14). Further, a very recent study demonstrated the value of human metabolomic data in drug target studies by employing MR to evaluate genetically predicted effects of lipid-modifying therapies on the metabolome (15). These studies highlight the value of combining different layers of omics data, including proteomics, genomics, transcriptomics, and metabolomics, for enhanced drug discovery and development.

Loss-of-Function and Gain-of-Function Variants from Sequenced Genomes

The declining cost of next-generation sequencing technologies in recent decades has enabled the adoption of this method at a larger scale and has led to a rapid growth in the number of sequenced genomes. It is expected that a massive number of variants not previously described will be discovered. This paves the way for the discovery of rare loss-of-function (LoF) variants and protective alleles that can also be used for further drug target studies. One recent elegant study by Nielsen et al. (16) indicated that using different genomic approaches such as whole-genome sequencing, imputation, GWASs, and downstream analyses enables the construction of well-powered studies for the discovery of protein-altering variants that may point to promising pharmaceutical drug targets. In this study, Nielsen et al. observed that silencing *ZNF529* in human hepatoma cells resulted in increased low-density lipoprotein (LDL) uptake, thus suggesting that the inhibition of *ZNF529* or its gene product could be regarded as a novel and promising drug target candidate for the treatment of dyslipidemia. Furthermore, anticipating the need for large-scale human genomic data, the UK Biobank (UKB) plans to sequence the exomes of all ~500,000 UKB participants and recently released the results of the first 200,000 sequenced exomes (17). In this data set, approximately 10 million exonic variants and 1,492 genes with at least one homozygous LoF variant were detected. This tolerance to LoF variants, calculated as the ratio of the count of putative LoF variants observed in a population to the number expected based on mutation rates (observed/expected), also referred to as the constraint score, is particularly interesting for drug development and has also been surveyed in the Genome Aggregation Database (gnomAD) v2 data set of 141,456 individuals (18). Minikel et al. (18) calculated constraint scores across all protein-coding genes and made several important conclusions. For example, although known drug targets were slightly more constrained compared to all genes (44% versus 52%), their distribution of scores was qualitatively similar. They also observed a difference in constraint scores between categories of genes that were expected to vary in their degree of tolerance to inactivation, which has also been reported by others (19, 20). These findings validate the usefulness of constraint in evaluating drug targets and emphasize that even essential genes can be highly successful as targets of inhibitory drugs. Thus, to enable the full potential of GWASs for drug discovery, the resequencing of candidate drug target genes should be considered in the context of discovery of both LoF and gain-of-function alleles. Furthermore, linking these in-depth genomic data with rich repositories of phenotypic traits—information that is systematically stored in the EHRs or collected at focused clinical biobanks—will enable the thorough evaluation and validation of identified potential drug targets (17, 21).

Phenome-Wide Studies in Longitudinal Data from EHRs

EHRs yield longitudinal phenotypic information on medication exposures and diagnosed diseases, thus making them a powerful platform for different studies on both drug discovery and drug effects. Records of the health information of individuals are generated at the point of care by health-care providers, and systematic collection enables sharing and accessing across health-care systems to obtain more comprehensive clinical care (22). The information stored in the EHRs includes a combination of structured and unstructured data with information on demographics, medical and surgical history, allergies and medications, diagnoses and procedures, and reports from various clinical studies (22, 23). Thereby, EHRs enable us to construct a detailed clinical picture of each individual that can be used for research. Data for research studies are pseudonymized, meaning that any information that can identity a subject is replaced by pseudonyms as unique identifiers. Structured data use a uniform format for recording information and use controlled vocabularies such as International Classification of Disease (ICD) codes for patient diagnoses, procedures, and complications and the Anatomical Therapeutic Chemical classification system for classification of drugs. Although these codes were primarily created for billing purposes, they have become extremely valuable in medical research as well. The Observational Medical Outcomes Partnership (OMOP) common data model provided by the Observational Health Data Sciences and Informatics collaborative (24; **https://ohdsi.github.io/TheBookOfOhdsi/**) is an example of a global effort toward uniform standards and tools for harmonizing electronic health data for research. The concept behind the OMOP common data model is to transform data contained within disparate databases into a common format and representation (coding schemes, terminologies, vocabularies), which will then allow systematic analyses using a library of standard analytic tools that have been written based on the common format.

Biobanks with the possibility of linking to EHRs have enabled the emergence of new approaches for leveraging all this information for research. A reverse GWAS approach, termed phenome-wide association study (PheWAS), was developed where a genetic variant is analyzed against multiple phenotypes to explore disease-gene associations (25). The first PheWAS in 2010 enabled successful replication of four known single-nucleotide polymorphism (SNP)-disease associations for multiple sclerosis, Crohn's disease, coronary artery disease (CAD), and rheumatoid arthritis (26) that resulted in the increasing popularity of using this method for research. Studies have also shown the potential of PheWASs for drug repurposing. By linking drug-targeted genes in the DrugBank database to the gene-phenotype associations in published PheWAS results (27), Rastegar-Mojarad et al. (28) validated the disease indications of drugs in 127 cases, and they further identified 2,583 cases that had strong potential for novel drug-disease associations. They highlighted the *LDLR* gene as an example of the repurposing potential of porfimer for the treatment of hypercholesterolemia and targeting the *TERT* gene as a repurposing opportunity for zidovudine in the treatment of diabetes. Another study demonstrating PheWASs as a powerful addition to current approaches for drug discovery interrogated 25 SNPs that were previously linked to 19 candidate drug targets through GWASs, analyzed their association with 1,683 clinical binary end points in four large cohorts, and meta-analyzed 145 end points to provide between-cohort comparisons of the results (29). They managed to replicate 75% of known GWAS associations and identified nine study-wide significant novel associations. These study results supported the inhibited targeting of PNPLA3 for the treatment of liver disease. However, the results also included associations with multiple other end points such as acne and high cholesterol levels, which indicates potential for relevant on-target adverse events. Discovery of potential drug side effects is clearly one of the important strengths of PheWASs in drug target evaluation. Another study specifically applied PheWASs for the detection of potential adverse drug effects and validated findings for 13 of 16 gene–drug class pairs (30). The authors also showed that PheWASs can replicate published safety information across multiple drug classes. PheWASs also have the potential to predict or validate negative findings from randomized clinical trials. A study where ICD-10 coding was used to define clinical end points validated findings from previous randomized controlled trials by determining that a LoF variant in *PLA2G7* had no associations with the improvement of vascular diseases such as stroke and coronary events (31).

Extracting Free Text from EHRs by Natural Language Processing

While structured data in EHRs are consistent and can be readily applied for research, unstructured data do not follow a particular format by allowing health-care providers to enter free text without constraints. Therefore, the analysis of unstructured data requires specific text-mining tools like natural language processing (NLP) for the extraction of relevant information (23). This approach enables researchers to further uncover phenotypic information embedded in the free-text documents and use it for research, including studies for drug discovery.

Applying NLP methods to extract adverse drug events from EHRs has been demonstrated in different countries (32–34). This approach was also successful in identifying individuals with penicillin allergy in the Estonian Biobank and Vanderbilt University's BioVU, where GWASs and fine mapping of human leukocyte antigen alleles were possible due to linked genotype data, thereby enabling the discovery of new insights into the genetics behind penicillin allergy (35). Other innovative methods include translating clinical text data into a text-based phenome that can be used for PheWASs (36), thus creating the potential to add further information and reveal unknown associations. A more broadly used approach has been aggregating one or more diagnosis codes (e.g., ICD-9 and ICD-10 codes) into so-called phecodes, corresponding to distinct diseases or traits (27, 37–39), which have subsequently been used in GWASs and PheWASs.

HARNESSING GENOMIC AND EHR DATA

For Discovery of Associations

There is now a long-standing recognition that using knowledge from studies of human genetics has a strong potential to accelerate drug discovery. It enables us to first discover genes relevant for human disease and then turn to model organisms for studies of underlying mechanisms. Currently, the most prominent example of genetic association studies leading to the identification of potential drug targets is the *PCSK9* story. The first studies made evident that gain-of-function variants in this gene were associated with elevated LDL levels and that this gene may have a causal role in CAD (40). This finding was followed by the assumption that LoF alleles of *PCSK9* may have functionally opposite effects and thus result in reducing the risk of CAD. Further, GWASs and resequencing and epidemiological studies all confirmed that LoF variants reduce LDL levels and thereby lifetime risk for CAD (41–43). This observation resulted in the development of two monoclonal antibodies that inhibit PCSK9 and that were granted US Food and Drug Administration (FDA) approval in 2015 (44). Thus, human genetic studies accelerated the drug development process from discovery to an approved drug within 12 years.

Another key example is the *SOST* gene, which encodes sclerostin, a protein that is secreted by osteocytes and negatively regulates bone formation. The discovery that a LoF mutation in *SOST* leads to high bone mass (45) formed the basis for the development of romosozumab, a humanized monoclonal antibody against sclerostin. It acts through sclerostin inhibition, which results in increasing bone mineral density, and is therefore warranted for the treatment of osteoporosis (46). In 2019, 18 years later, both the FDA and European Medicines Agency approved romosozumab for the treatment of osteoporosis in postmenopausal women at high risk of fractures (47, 48).

Furthermore, there are also a plethora of examples showing how findings from GWASs have supported ongoing drug discovery efforts by confirming the desired effect of a drug under development (reviewed in 49). In addition, GWASs have also retrospectively identified the genetic basis for drugs already in use. Statins have been used to treat hyperlipidemia for a long time by inhibiting 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, and a GWAS from 2008 confirmed that LDL levels are associated with variation in *HMGCR*, the gene that encodes HMG-CoA reductase (50).

For Drug Repurposing

One of the most important and promising angles of implementing human genetics in drug discovery is drug repurposing, a strategy to discover novel pharmacological effects for already-existing and approved drugs. Drug repurposing enables skipping some of the steps in drug development, thus potentially reducing the overall cost and time moving from a discovered target to an approved drug. The most remarkable example of drug repurposing initiated after discoveries made in the analysis of GWAS data is the repurposing of monoclonal antibodies modulating IL-23 for the treatment of Crohn's disease (51, 52). The monoclonal antibodies ustekinumab and risankizumab were first used for the treatment of psoriasis, and the former is now also officially approved for the treatment of Crohn's disease.

Though there are many studies demonstrating several repurposing candidates, it has been a challenge to determine which of the repurposing candidates have the highest likelihood of succeeding. To tackle this problem, a recent study described a proof-of-concept approach on how to identify and validate drug repurposing candidates using gene expression signatures, drug perturbation data, and clinical EHRs from BioVU (53). The approach was applied for two diseases, hyperlipidemia and hypertension, and the effects of 10 approved drugs were replicated. Furthermore, 25 drugs approved for other indications were identified, and for five of these, the therapeutic effects were further independently replicated in data from the All of Us Research Program. Thus, this or similar approaches (8) have good potential to be high-throughput options for identifying and prioritizing drug repurposing candidates.

For Evaluation of Side Effects

Genomics can also be applied to yield insights into potential adverse effects of drugs. Addressing naturally occurring variants in the human genome that alter the activity of a protein that is targeted by a particular drug may help to predict the on-target side effects of this drug. In the case of the osteoporosis drug romosozumab, the FDA included a Boxed Warning for the risk of cardiovascular events, which was supported and validated by meta-analysis of outcomes from several clinical trials and proof of LoF variants altering sclerostin function against phenotype data in the UKB and Estonian Biobank (54). Thus, with such genetic analyses, an increased risk of cardiovascular events could have been recognized before launching the clinical trials and perhaps enabled a better trial design. Similarly, a large meta-analysis of clinical trials of statin therapy indicated the known association of *HMGCR* with decreased LDL levels but also revealed a 9% increased risk for type 2 diabetes (55). This suggested an on-target effect of statin use requiring attention. A recent MR analysis of the genotype and EHR data of 53,385 individuals also replicated this link between genetic variants in *HMGCR* associated with lower LDL cholesterol and increased risk for type 2 diabetes (56). Another MR study identified potential protective effects of variants in *SCARA5* and *TNFSF12* on cardioembolic stroke. To further explore whether these potential drug targets of stroke affect other traits, a phenome-wide MR analysis was performed that confirmed *SCARA5* as a more promising target for the treatment of cardioembolic stroke, since *TNFSF12* indicated additional associations with increased risk for four circulatory system phenotypes (including intracerebral and subarachnoid hemorrhages), three digestive phenotypes, and one injuries and poisonings phenotype, which indicate anticipated side effects when drugging this protein (57). Furthermore, another recent study also indicated how human genetics data not only help in selecting effective drug targets for development but also aid the development of safer drugs by showing that phenotypes that have been associated with genes encoding drug targets can predict side effects in clinical trials (58). Thus, genetic association with a nonrelevant phenotype has proven to be a good indicator of the increased likelihood of corresponding adverse events.

RAPID DRUG TARGET IDENTIFICATION DURING A GLOBAL PANDEMIC

The COVID-19 pandemic has had a substantial impact on people's lives and health and required unprecedented response from health-care providers, scientists, and pharmaceutical companies. Less than a year after the sequencing of the SARS-CoV-2 genome, vaccines were available, and hundreds of clinical trials for different medications against severe COVID-19 have been running since the beginning of the pandemic. All the methods and approaches covered in this review have also contributed to the identification of potential drug targets for COVID-19. Early in the pandemic, scientists across the world joined forces in the COVID-19 Host Genetics Initiative to conduct large genome-wide studies that could explain why some people do not get infected by SARS-CoV-2 and why some experience more severe symptoms (59). Biobanks with existing genotypic data and the possibility of linking to EHRs were central to this effort (60). In parallel, several hospital-based studies were launched in countries that were heavily affected early on in the pandemic. The first GWASs revealed several genes that were strongly associated with COVID-19, including *SLC6A20*, *LZTFL1*, *FYCO1*, *CXCR6*, *XCR1*, and *CCR9* at 3p21.31; *FOXP4* at 6p21.1; *ABO* at 9q34.2; *OAS1*, *OAS2*, and *OAS3* at 12q24.13; *KANSL1* at 17q21.31; *DPP9*, *TYK2*, and *PPP1R15A* at 19p13.3; and *IFNAR2* and *IL10RB* at 21q22.1 (61–64). Independent multi-ancestry fine mapping implicated *OAS1* as an effector gene at 12q24.13 that influenced COVID-19 severity (65), and colocalization analysis integrating the results of a CRISPR screen (66) and cis-expression quantitative trait loci (eQTLs) highlighted *SLC6A20* and *CXCR6* as potential causative genes in the 3p21.31 locus associated with a higher risk for COVID-19. Moreover, lung-specific cis-eQTLs from GTEx v8 (67) and the Lung eQTL Consortium (68) provided further functional evidence that the COVID-19-associated variants in *FOXP4*, *ABO*, *OAS1*, and *IFNAR2*/*IL10RB* modify gene expression in the lungs (63). Further, a large-scale MR study of more than 3,000 blood proteins replicated findings of blood markers associated with COVID-19 and implicated additional markers, including higher levels of a number of adhesion molecules such as SELE, SELL, PECAM-1, and ICAM-1, for the prediction of risk for severe disease (69). A separate study of therapeutic targets relevant to COVID-19 used MR analyses with genetic instruments based on published COVID-19 GWASs and transcriptomic and proteomic data for 1,263 actionable proteins that are already targeted by drugs that have been approved or are in development (70). Their findings prioritized further trials of drugs targeting ACE2 and IFNAR2 for early treatment of COVID-19. A recent important large-scale GWAS with 756,646 individuals across four cohorts identified a rare variant in *ACE2* (frequency 0.2–2%) that reduces the risk of SARS-CoV-2 infection, thus further confirming how ACE2 expression levels influence COVID-19 risk. They demonstrated how coupling genetics with EHRs extends the knowledge of COVID-19 host genetics and how larger sample sizes provide additional power for the detection of rare variants (71). An earlier MR study of genetic instruments mimicking the inhibition of the IL-6 receptor also indicated protective effects against hospitalization due to COVID-19, indicating the potential efficacy of repurposing sarilumab or tocilizumab monoclonal antibodies, which are approved for the treatment of arthritis and other inflammatory conditions via inhibition of the IL-6 receptor. These initial results have now been supported by a meta-analysis of 10 randomized clinical trials evaluating the efficacy and safety of tocilizumab versus standard care in patients with COVID-19 (72).

CONCLUSION

Taken together, large population biobanks and clinical biorepositories where genotypic data and EHRs are linked, pseudonymized, and available for research provide invaluable resources for drug discovery, validation, and repurposing. Established simple and advanced statistical methods that allow the simultaneous analysis of genotypic and clinical phenotypic data by genome- and phenome-wide analyses, colocalization analyses with QTL data from transcriptomics and proteomics data sets from different tissues, and MR are essential tools for drug development in the postgenomic era. This fact was particularly highlighted by the rapid studies that identified potential drug repurposing opportunities during the global COVID-19 pandemic.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the Estonian Research Council (grant PRG184) and by the European Union through the European Regional Development Fund (project no. 2014-2020.4.01.15-0012 GENTRANSMED).

LITERATURE CITED

1. Dowden H, Munro J. 2019. Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Discov.* 18(7):495–96
2. Wouters OJ, McKee M, Luyten J. 2020. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* 323(9):844–53
3. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, et al. 2015. The support of human genetic evidence for approved drug indications. *Nat. Genet.* 47(8):856–60
4. King EA, Davis JW, Degner JF. 2019. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLOS Genet.* 15(12):e1008489
5. Jensen PB, Jensen LJ, Brunak S. 2012. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* 13:395–405
6. Zhou W, Kanai M, Wu K-HH, Humaira R, Tsuo K, et al. 2021. Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. medRxiv 2021.11.19.21266436. **https://doi.org/10.1101/2021.11.19.21266436**
7. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, et al. 2012. Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* 30(4):317–20
8. Namba S, Konuma T, Wu K-H, Zhou W, Biobank G, et al. 2021. A practical guideline of genomics-driven drug discovery in the era of global biobank meta-analysis. medRxiv 2021.12.03.21267280. **https://doi.org/10.1101/2021.12.03.21267280**
9. Yarmolinsky J, Díez-Obrero V, Richardson TG, Pigeyre M, Sjaarda J, et al. 2022. Genetically proxied therapeutic inhibition of antihypertensive drug targets and risk of common cancers: a mendelian randomization analysis. *PLOS Med.* 19(2):e1003897

10. Swerdlow DI, Holmes MV, Kuchenbaecker KB, Engmann JEL, Shah T, et al. 2012. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 379(9822):1214–24
11. Storm CS, Kia DA, Almramhi MM, Bandres-Ciga S, Finan C, et al. 2021. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat. Commun.* 12:7342
12. Suhre K, McCarthy MI, Schwenk JM. 2021. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* 22(1):19–37
13. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, et al. 2020. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* 2(10):1135–48
14. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, et al. 2021. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* 53(12):1712–21
15. Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, et al. 2022. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLOS Biol.* 20(2):e3001547
16. Nielsen JB, Rom O, Surakka I, Graham SE, Zhou W, et al. 2020. Loss-of-function genomic variants highlight potential therapeutic targets for cardiovascular disease. *Nat. Commun.* 11:6417
17. Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, et al. 2021. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* 53(7):942–48
18. Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, et al. 2020. Evaluating drug targets through human loss-of-function genetic variation. *Nature* 581(7809):459–64
19. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536(7616):285–91
20. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581(7809):434–43
21. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, et al. 2017. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* 9(383):eaag1166
22. Abul-Husn NS, Kenny EE. 2019. Personalized medicine and the power of electronic health records. *Cell* 177(1):58–69
23. Pendergrass SA, Crawford DC. 2019. Using electronic health records to generate phenotypes for research. *Curr. Protoc. Hum. Genet.* 100(1):e80
24. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, et al. 2015. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud. Health Technol. Inform.* 216:574–78
25. Robinson JR, Denny JC, Roden DM, Van Driest SL. 2018. Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. *Clin. Transl. Sci.* 11(2):112–22
26. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, et al. 2010. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26(9):1205–10
27. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, et al. 2013. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31(12):1102–10
28. Rastegar-Mojarad M, Ye Z, Kolesar JM, Hebbring SJ, Lin SM. 2015. Opportunities for drug repositioning from phenome-wide association studies. *Nat. Biotechnol.* 33(4):342–45
29. Diogo D, Tian C, Franklin CS, Alanne-Kinnunen M, March M, et al. 2018. Phenome-wide association studies across large population cohorts support drug target validation. *Nat. Commun.* 9:4285
30. Jerome RN, Joly MM, Kennedy N, Shirey-Rice JK, Roden DM, et al. 2020. Leveraging human genetics to identify safety signals prior to drug marketing approval and clinical use. *Drug Saf.* 43(6):567–82
31. Millwood IY, Bennett DA, Walters RG, Clarke R, Waterworth D, et al. 2016. A phenome-wide association study of a lipoprotein-associated phospholipase A2 loss-of-function variant in 90 000 Chinese adults. *Int. J. Epidemiol.* 45(5):1588–99
32. Roitmann E, Eriksson R, Brunak S. 2014. Patient stratification and identification of adverse event correlations in the space of 1190 drug related adverse events. *Front. Physiol.* 5:332

33. Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. 2012. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. *Br. J. Clin. Pharmacol.* 73(5):674–84

34. Wasylewicz A, van de Burgt B, Weterings A, Jessurun N, Korsten E, et al. 2022. Identifying adverse drug reactions from free-text electronic hospital health record notes. *Br. J. Clin. Pharmacol.* 88(3):1235–45

35. Krebs K, Bovijn J, Zheng N, Lepamets M, Censin JC, et al. 2020. Genome-wide study identifies association between HLA-B*55:01 and self-reported penicillin allergy. *Am. J. Hum. Genet.* 107(4):612–21

36. Hebbring SJ, Rastegar-Mojarad M, Ye Z, Mayer J, Jacobson C, Lin S. 2015. Application of clinical text data for phenome-wide association studies (PheWASs). *Bioinformatics* 31(12):1981–87

37. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, et al. 2010. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26(9):1205–10

38. Wei W-Q, Bastarache LA, Carroll RJ, Marlo JE, Osterman TJ, et al. 2017. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLOS ONE* 12(7):e0175508

39. Wu P, Gifford A, Meng X, Li X, Campbell H, et al. 2019. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med. Inform.* 7(4):e14325

40. Abifadel M, Varret M, Rabès JP, Allard D, Ouguerram K, et al. 2003. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia. *Nat. Genet.* 34(2):154–56

41. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. 2005. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nat. Genet.* 37(2):161–65

42. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. 2006. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* 354(12):1264–72

43. Sanna S, Li B, Mulas A, Sidore C, Kang HM, et al. 2011. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLOS Genet.* 7(7):e1002198

44. Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, et al. 2015. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N. Engl. J. Med.* 372(16):1489–99

45. Brunkow ME, Gardner JC, Van Ness J, Paeper BW, Kovacevich BR, et al. 2001. Bone dysplasia sclerosteosis results from loss of the *SOST* gene product, a novel cystine knot-containing protein. *Am. J. Hum. Genet.* 68(3):577–89

46. McClung MR, Grauer A, Boonen S, Bolognese MA, Brown JP, et al. 2014. Romosozumab in postmenopausal women with low bone mineral density. *N. Engl. J. Med.* 370(5):248–59

47. Amgen. 2019. *FDA approves EVENITY^{TM} (romosozumab-aqqg) for the treatment of osteoporosis in postmenopausal women at high risk for fracture*. News Release, Apr. 9. **https://www.prnewswire.com/news-releases/fda-approves-evenity-romosozumab-aqqg-for-the-treatment-of-osteoporosis-in-postmenopausal-women-at-high-risk-for-fracture-300828376.html**

48. Amgen. 2019. *European Commission approves EVENITY® (romosozumab) for the treatment of severe osteoporosis in postmenopausal women at high risk of fracture*. Press Release, Dec. 11. **https://www.amgen.com/newsroom/press-releases/2019/12/european-commission-approves-evenity-romosozumab-for-the-treatment-of-severe-osteoporosis-in-postmenopausal-women-at-high-risk-of-fracture**

49. Harper AR, Nayee S, Topol EJ. 2015. Protective alleles and modifier variants in human health and disease. *Nat. Rev. Genet.* 16(12):689–701

50. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, et al. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* 40(2):189–97

51. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, et al. 2006. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314(5804):1461–63

52. Sandborn WJ, Feagan BG, Fedorak RN, Scherl E, Fleisher MR, et al. 2008. A randomized trial of Ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with moderate-to-severe Crohn's disease. *Gastroenterology* 135(4):1130–41

53. Wu P, Feng QP, Kerchberger VE, Nelson SD, Chen Q, et al. 2022. Integrating gene expression and clinical data to identify drug repurposing candidates for hyperlipidemia and hypertension. *Nat. Commun.* 13:46

54. Bovijn J, Krebs K, Chen C-Y, Boxall R, Censin JC, et al. 2020. Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics. *Sci. Transl. Med.* 12(549):eaay6570
55. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, et al. 2010. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* 375(9716):735–42
56. Liu G, Shi M, Mosley JD, Weng C, Zhang Y, et al. 2021. A Mendelian randomization approach using 3-HMG-coenzyme-A reductase gene variation to evaluate the association of statin-induced low-density lipoprotein cholesterol lowering with noncardiovascular disease phenotypes. *JAMA Netw. Open* 4(6):e2112820
57. Chong M, Sjaarda J, Pigeyre M, Mohammadi-Shemirani P, Lali R, et al. 2019. Novel drug targets for ischemic stroke identified through Mendelian randomization analysis of the blood proteome. *Circulation* 140(10):819–30
58. Nguyen PA, Born DA, Deaton AM, Nioi P, Ward LD. 2019. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat. Commun.* 10(1):1579
59. COVID-19 Host Genet. Initiative. 2020. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. 2020. *Eur. J. Hum. Genet.* 28(6):715–18
60. Satterfield BA, Dikilitas O, Kullo IJ. 2021. Leveraging the electronic health record to address the COVID-19 pandemic. *Mayo Clin. Proc.* 96(6):1592–608
61. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, et al. 2021. Genetic mechanisms of critical illness in COVID-19. *Nature* 591(7848):92–98
62. Severe Covid-19 GWAS Group. 2020. Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* 383(16):1522–34
63. Niemi MEK, Karjalainen J, Liao RG, Neale BM, Daly M, et al. 2021. Mapping the human genetic architecture of COVID-19. *Nature* 600(7889):472–77
64. Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Sonmez T, et al. 2021. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* 53(6):801–8
65. Huffman JE, Butler-Laporte G, Khan A, Pairo-Castineira E, Drivas TG, et al. 2022. Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19. *Nat. Genet.* 54(2):125–27
66. Daniloski Z, Jordan TX, Wessels HH, Hoagland DA, Kasela S, et al. 2021. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* 184(1):92–105.e16
67. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, et al. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369(6509):1318–30
68. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, et al. 2012. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLOS Genet*. 8(11):e1003029
69. Palmos AB, Millischer V, Menon DK, Nicholson TR, Taams LS, et al. 2022. Proteome-wide Mendelian randomization identifies causal links between blood proteins and severe COVID-19. *PLOS Genet*. 18(3):e1010042
70. Gaziano L, Giambartolomei C, Pereira AC, Gaulton A, Posner DC, et al. 2021. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat. Med.* 27(4):668–76
71. Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GHL, et al. 2022. Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* 54:382–92
72. Vela D, Vela-Gaxha Z, Rexhepi M, Olloni R, Hyseni V, Nallbani R. 2022. Efficacy and safety of tocilizumab versus standard care/placebo in patients with COVID-19: a systematic review and meta-analysis of randomized clinical trials. *Br. J. Clin. Pharmacol.* 88(5):1955–63