Computational drug repositioning for rare diseases in the era of precision medicine

Delavan, Brian ; Roberts, Ruth; Huang, Ruili; Bao, Wenjun; Tong, Weida; Liu, Zhichao

DOI:
10.1016/j.drudis.2017.10.009

Document Version
Peer reviewed version

Citation for published version (Harvard):
Delavan, B, Roberts, R, Huang, R, Bao, W, Tong, W & Liu, Z 2018, 'Computational drug repositioning for rare diseases in the era of precision medicine', *Drug Discovery Today*, vol. 23, no. 2, pp. 382-394. https://doi.org/10.1016/j.drudis.2017.10.009

Computational Drug Repositioning for Rare Diseases in the Era of Precision Medicine

Brian Delavan[1,2], Ruth Roberts[3,4], Jonathan Goldsmith[5], Hong Fang[1], Shraddha Thakkar[1], Ruili Huang[6], Wenjun Bao[7], Weida Tong[1*], Zhichao Liu[1*]

[1] National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas 72079, United States

[2] University of Arkansas at Little Rock, Little Rock, AR, 72204, United States

[3] ApconiX, BioHub at Alderley Park, Alderley Edge, SK10 4TG, UK

[4] University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

[5] Office of New Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland 20993, United States

[6] National Center for Advancing Translational Sciences, National Institutes of Health Rockville, Maryland 20850, United States

[7] SAS Institute Inc., Cary, North Carolina, USA.

*To whom correspondence should be addressed at: Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA. Telephone: (870) 543-7909. Fax: (870) 543-7854. weida.tong@fda.hhs.gov or zhichao.liu@fda.hhs.gov

Disclaimer: *The views presented in this article do not necessarily reflect current or future opinion or policy of the US Food and Drug Administration and National Institutes of Health. Any mention of commercial products is for clarification and not intended as endorsement.*

Abstract

There are tremendous unmet needs and unprecedented opportunities in drug development for rare diseases. Advances in emerging techniques such as next generation sequencing has changed the landscape of research in rare diseases, exemplified by our increasing knowledge of the genetic origins of disease. In *silico* drug repositioning is a promising approach and has been successfully applied to the development of treatments for diseases. The underlying genetic nature of rare diseases influences the treatment responses of different genetic mutation carriers, which is an important component of precision medicine. However, how to utilize this knowledge and effectively conduct and implement in *silico* drug repositioning approaches for rare disease therapies is still an open question. In this review, we will focus on the means of utilizing accumulated genomic data for accelerating and facilitating drug repositioning for the treatment of rare diseases. First, we summarize the current genome landscape of rare diseases. Second, we propose several promising bioinformatics approaches and pipelines for computational drug repositioning for rare diseases. Finally, we discuss recent regulatory incentives and other enablers in rare disease drug development and outline the remaining challenge.

Keywords: drug repositioning; rare diseases; precision medicine; genome; next generation sequencing

Introduction

Most rare diseases have a genetic etiology, affect a small proportion of the population (usually less than 200,000 in U.S. or 1/2000 in Europe), but are severe and life-threatening [1-3]. Although rare diseases are themselves infrequent by definition, collectively they are a common occurrence. There are more than 7,000 rare diseases based on the European Organization for Rare Diseases (EURORDIS) statics (http://www.eurordis.org/about-rare-diseases). However, there are only approximately 500 treatment options available after the Orphan Drug Act of 1983 was passed [4]. The average time to diagnosis of a rare disease is more than seven years. Over one-third of children with a rare disease will not live more than five years, and about 35% of these children will die within the first year of life [5].

The fundamental challenge of orphan drug development is a lack of knowledge about pathophysiology, etiology, and the natural history of rare diseases. Few patients are available and together with their geographical dispersal, clinical trials are often impractical [6]. Also, researchers also have great difficulty in gauging the genetic origin of rare diseases [1]. The causative genetic mutations are either hereditary (even when the disease has a late onset in the patient's life), or they are caused by a new mutation (*de novo*) [7]. Like common disease, heterogeneity also exists in rare diseases, which makes it extremely challenging to distinguish patients with different morphological features or genetic variants and then look for the right treatment options. One example is cystic fibrosis (CF), which is accounted for by the genetic mutation of the transmembrane conductance regulator (CFTR) gene. There are approximately 2000 identified mutations within the CFTR gene from CF patients. Among the 2000 identified CFTR mutations, the F508del mutation and G551D are major mutations that are carried by more than 90% of CF patients. However, the associated phenotypic outcomes of the two mutations are quite distinct. The F508del mutation is mainly associated with CFTR folding impairment, and stability at the endoplasmic reticulum and plasma membrane, and chloride channel gating. The G551D is mainly related to channel gating alternation [8, 9]. The only FDA approved drug, ivacaftor, is only effective to patients with the G551D mutation. Meanwhile, there are still a substantial number of CF patients carrying the F508del mutation without a treatment option.

The advent of next-generation sequencing (NGS) has changed the landscape of rare disease research, presenting the opportunity for the causative genes of rare diseases to be identified at an unprecedented pace and resolution [1]. Next generation sequencing (NGS) is also considered as a key technology for advancing precision medicine [10]. Many genetic variants of rare diseases have been detected and the data are publicly accessible. However, there is still a large number of undetected genes associated with rare diseases [7, 11, 12]. On-going efforts are being made and will lead to substantial improvement in our understanding of the genetic origin of rare diseases. For example, the International Rare Diseases Research Consortium (IRDiRC) set a goal of developing the capacity to diagnose all of the rare diseases, and to establish 200 new or repurposed therapies for rare diseases by the year 2020 [13].

How to translate the accumulated genetic knowledge to facilitate rare disease treatment development is still an open question [14]. First, to identify and validate therapeutic targets of rare diseases is a great challenge. Even if a causative genetic mutation in a patient with rare diseases is detected, there is no guarantee that a therapeutic option might arise from this knowledge. This is because the mutated protein may be unsuitable as a therapeutic target for a variety of reasons such as inaccessibility or lack of suitability as a small molecule target. [15]. In this context, the current drug design paradigm has proved generally successful in inhibiting therapeutic targets in rare diseases with gain-of-function mutations [16]. Rare diseases with gain-of-function mutations, like most common diseases, are defined as the activation of specific pathways or the ectopic activity in relation to the proteins, which aligns well with the current concept of target identification. However, there are many rare diseases that are due to loss-of-function where the impairment of a particular protein drives the etiology [16]. Therefore, a novel approach for translating knowledge of loss-of-function genetic variants nto clinical utility is urgently needed.

Drug repositioning that aims to find new uses for existing drugs is considered as an effective and alternative paradigm of drug development [17]. Computational drug repositioning provides a systematic and rational solution for identifying treatment options as compared to conventional drug repositioning approaches arising from serendipity or close clinical observation[18-21]. Linking the genetic findings of rare disease and drug repositioning into the

same framework to accelerate drug development for rare diseases is imperative and is also a necessary practice for precision medicine. In this review, first, we summarize current progress in research on the genetic origins of rare diseases. Second, we propose several novel strategies to integrate these accumulated genetic findings into computational drug repositioning frameworks for the development of treatments for rare diseases (**Figure 1**). Finally, we will discuss the remaining challenges and future perspectives in this field.

The genetic landscape of rare diseases

In the past decade, much progress has been made in the detection the genetic origin of rare diseases even though patient recruitment is a challenge both for obtaining samples and for carrying out clinical studies for the development of treatment options. This has resulted from the advancement of new techniques, the assistance of social media, and the policy shifts of regulatory agencies [1, 22, 23]. Particularly, NGS techniques have greatly enabled the detection of the possible genetic basis of rare diseases [24]. **Table 1** summarizes the public available resources and efforts of rare disease genetics.

Up to date, the molecular level etiology information of around one third of rare disease has been uncovered, although many causative genes of rare diseases remain to be identified [25, 26]. Based on the Orphanet data [27], there are a total of 6,289 rare diseases with a causative gene relationship, which corresponds to 3343 rare diseases and 3,398 genes. Among 3343 rare diseases, 2,442 (2442/3343 = 73.0%) have a single causative gene (**Figure 2(A)**). Among 6,289 rare disease and causing gene relationships, 5,032 (4,171 unclassified + 715 loss of function + 146 gain of function) belong to germline mutation in the causative genes, which account for more than 80% of mutation types (**Figure 2(B)**). It could be also seen that 4,171 unclassified mutation (4,171/5,032 = 82.9% of total germline mutations) remain to be annotated at the functional level.

Genetic structural variants have been implicated in mutation functions and phenotypic outcomes. However, genetic structural variants are still considered as one of most difficult to interpret with regards to their functional consequence [22]. Structural variants comprise different unbalanced forms of variants such as deletion, insertion, reduplication, and also

balanced forms such as translocation and inversion. ClinVar is a database for the clinical significance of mutations [11]. Based on ClinVar, there are a total 52,944 genetic mutations from 3,502 unique rare disease-associated genes, which distributes into different chromosome locations. The types of 52,944 rare disease structure variants includes single nucleotide variant (SNV), deletion, duplication, insertion, insertion, indel, undetermined variant, NT expansion, protein only, copy number loss, copy number gain, inversion, short repeat, structural variant. Among 13 mutation types, SNV, deletion, and duplication are the three most frequent mutation types (**Figure 2(C)**).

Paths toward to rare disease therapy

The emerging techniques have accelerated the pace of the identification of rare diseases genetic variants [1]. However, the majority of the detected variants remain to be translated into treatment options. . Here, we summarize and propose several computational drug repositioning approaches for facilitating this process (**Figure 1**).

Phenome-wide association

The candidate gene and genome-wide association studies (GWAS) studies have identified a large number of SNP-trait/disease relationships [28], which could be used to prioritize genetic findings and further identify therapeutic targets [29-31]. Sanseau *et. al.* [32] assessed the utility of GWAS for identifying alternative uses of existing drugs. It was found that a list of 155 genes identified from GWAS studies had been targeted by at least one existing drug or candidate in clinical trials. For 92 of 155 genes, the suggested drug indication was different from the original disease trait identified by GWAS, which implies that these new drug-indication pairs should be further verified for identifying new disease treatment options [33]. Similarly, Nelson *et.al.* [31] filtered SNP-trait/disease relationships from GWASdb with OMIM and obtained rare diseases related SNP-trait/disease relationships. Then, these were linked to drug–target-disease relationships to determine whether the known genetic associations could play a role in drug development. It was found that drug mechanisms with plausible genetic associations were twice as successful as those where this associated was missing.

However, results from GWASs contain a high false positive rate due to the limitations posed by both technique or sample size [34].

Integration of electronic health records (EHR) of various disease types from different ethnic groups to the dense genomic information presents a new vision of precision medicine [35]. Denny *et. al.* [36] reported a novel paradigm named phenome-wide association study (PheWAS), which incorporated SNP-trait relationship identified from GAWS studies with the electronic medical records of genetic scanning from a large cohort of people with European ancestry. The PheWAS not only provided an extra verification of the results from GWA studies, but also revealed some potentially interesting associations. The PheWAS tremendously expanded the scale of SNP-trait relationship and provided more opportunities for looking for new uses of existing drugs. Rastegar-Mojarad *et. al.* [37] combined PheWAS and DrugBank [38] to identify repositioning candidates for rare and common diseases. A total of 52,966 drug-disease pairs were enriched by the approach. Approximate 30% of 52,966 drug pairs were verified for known drug-disease relationship, on-going clinical trial or literature reports. About 70% of drug pairs could be candidates for drug repositioning.

Pathway/network based approaches

Genes with genetic variants may not be suitable “druggable” targets. However, pathway or network approaches can be helpful in finding genes involved in general signaling networks or biological pathways, and could provide a list of proteins for therapeutic target identification [39]. For example, the Ras/MAPK syndromes (Noonan, LEPAROD, Costello and cardio-facio-cataneous syndromes) are a class of rare developmental disorders caused by germline mutations of genes including PTPN11, PTPN11, SOS1, RASA1, NF1, KRAS, HRAS, NRA S, BRAF, RAF1, MAP2K1, MAP2K2, SPRED1, RIT1, SHOC2 and CBL. Ras/MAPK signaling pathways deregulated by cancerous somatic mutations exist in approximately one-third of all cancer types [40, 41]. Naturally, it is assumed oncology drugs that could inhibit the Ras/MAPK signaling pathways components could be used to treat RASopathy related rare development disorders. A mouse model was developed for verification of the oncology drug rapamycin for treating LEOPARD syndrome (LS) [42]. Specifically, mice carrying the *ptpn11* mutation developed LS

symptoms, and experiments verified that the mTOR inhibitor rapamycin could reverse some of these, such as hypertrophic cardiomyopathy (HCM).

Linking the common disease with the rare disease based on a shared gene is an idea originally proposed by Goh [43] which developed as a concept to identify the disease-disease relationship based on their shared pathways [44]. However, there is little knowledge about the underlying molecular mechanism of the influence of genetic variants on the pathways. This knowledge is crucial to understanding the pathogenesis of diseases. Kiel *et. al.* developed a structure-energy-based prediction and network modeling framework to uncover the different degrees of perturbation of the Ras/MAPK pathway by germline mutations and somatic mutations. By measuring quantitative activity changes in the pathway based on mutated 3D protein structure, the difference between germline RASopathy mutation and cancer mutations could be explained by switching the genes on and off and assessing the degree of protein-protein interactions.. Furthermore, the binding constants and affinities could be quite different for the same protein with different disease related mutations. In addition, the energy change noted in a pathway was higher with a somatic mutation compared to a germline mutation. Overall, these pathway/network-based methodologies and conclusions are of great value in uncovering the impact of genetic variants on pathways, further facilitating target identification and subsequent treatment development for rare diseases.

Genomic data integration

Deciphering the effect of genetic variants on cellular processes such as gene expression at the cellular or organism level is crucial in dissecting genetic contributions to phenotypic endpoints [45, 46]. This also paves the way for linking genetic variants to treatment development since vast amounts of drug transcriptome data in different cell types and organisms are publicly available [47-49].

The correlation between genetic variants and gene expression has been discussed and applied in the cancer genome field (**Table 2**). Although the proposed approaches are tailored to driver gene enrichment and patient survival, it also could be applied for treatment development. For example, Masica *et. al.* [50] proposed a statistical strategy with network analysis for correlating

somatic mutation and gene expression and applied it to 149 human glioblastoma (GBM) samples. They found that somatic mutations of 41 genes were highly related to GBM progression and patient survival. Bertrand *et. al.* [51] developed a network approach by integrating SNP, CNV and gene expression for driver gene enrichment. The proposed methodology was also applied to GBM and a novel driver gene TRIM24 was found and experimentally verified. In addition, the methodology was used for more than 1000 tumor samples from 5 different cancer types for identifying modes of synergistic action, which could be potentially used for combination drug design for cancer treatment. Peng *et. al.* [52] developed a hybrid integrative approach named CMDD by combining partial least squares regression and network methods covering multiple- omics profiles such as CNV, DNA methylation, miRNA and gene expression. CMDD was also applied to GBM and six other cancer types and the genes involved in the enriched modules were correlated with overall patient survival. Ding *et. al.* [53] presented a novel hierarchical Bayes graphic modeling approach for symmetrically qualifying the effect of somatic mutation on gene expression across 12 ~~pan~~ cancers. Some very interesting conclusion were drawn: (1) the patients carried the same somatic mutations, which influenced different downstream gene expression; (2) some somatic mutations are conserved across cancer types. Gerstung *et. al.* [54] developed a computational approach for detecting the phenotypic heterogeneity caused by distinct genotype and applied it to 124 patients with myelodysplastic syndromes (a rare cancer) and with TCGA acute myeloid leukemia (AML). It was found that the one or more genetic variants were correlated with around 20% of all genes, which dictated 20~65% of gene expression variability. These proposed methodologies have been successfully used to uncover genetic mutation and gene expression relationships for common or rare cancers. It is worth investigating the utility of these approaches in the rare diseases field to decipher germline mutations and their influencing on gene expression profiles.

Furthermore, dysfunctional non-coding RNA such as miRNAs and lncRNA in different biological process often leads to disease [55]. The genetic mutation may change the binding affinity to miRNA impairing gene expression, and contributing to the phenotypic expression of the diseases [56]. Liu *et. al.* [57] introduced a feed-forward loop concept into the drug repositioning

field and applied it for the development of treatment for cystic fibrosis (CF) by integrating information including germline mutation, miRNA, transcription factors (TF) and gene expression. Then, 15 CF specific miRNA-TF feed-forward loops were enriched by using a cumulative hypergeometric test. Finally, by investigating the perturbation of obtained CF specific FFLs with small molecules, a list of 48 CF repurposed candidates were proposed. Among the 48 repurposed candidates for CF, 26 candidates were verified by literature survey and existing clinical trials.

Once the correlation between genetic variants and gene expression, drug transcriptome data could be applied to look for repositioning opportunities (**Table 3**). The Connectivity Map (CMap) [49] as the key source has been successfully applied to drug repositioning fields [58, 59]. For example, Dudley *et. al.* [60] proposed a novel approach that aims to look for inverse drug diseases relationship by comparing the disease signature generated from the Gene Expression Omnibus (GEO) databases [61] and drug signatures obtained from CMap. They found several repurposing candidates for treating inflammatory bowel disease (IBD) and these were verified by in *vitro* assays.

Besides CMap, several large toxicogenomics efforts such as TG-GATEs [48] and DrugMatrix [47] have accumulated hundreds of drugs transcriptome data profiles at multiple time/dose/assay type points. Iskar *et. al.* [62] identified a large set of drug induced transcriptional modules with CMap and DrugMatrix data that are from human cancer cell lines and from rat liver in vivo. They found that 70% of drug induced transcriptional modules were conserved in both assay types, which suggests that toxicogenomics data could be also used for drug repositioning, although further comprehensive assessment is needed. Furthermore, miRNAs have been considered as novel and promising therapeutic targets against various diseases [63, 64], several miRNA and small molecule relationship databases such as SM2miR [65] and Pharmaco-miR [66] were constructed by curation from literature or *in silico* prediction.

Discussion

Computational drug repositioning provides a rapid turnaround list of repositioning candidate drugs. The challenge is to experimentally verify the efficacy and safety of these and to move the

drugs forward into clinical trials. Currently, most *in silico* drug repositioning approaches are verified by either animal-based *in vitro* or *in vivo* models [59, 60, 67-69]. Moving these *in silico* findings towards clinical application is challenging due to difficulties in patient recruitment, which are especially hard with patients with rare diseases. About 30% of clinical Phase 3 studies fail due to patient enrollments [70]. Therefore, a lot of proposed repositioning candidates remain at the report or literature level. Patient registries, which have been created by patient advocacy groups, none-profit organization, government agencies and companies, facilitate progress in the enrollment and retention of patients with rare diseases. For example, the National Institutes of Health (NIH) established the Rare Diseases Clinical Research Network I (RDCRN I, http://www.rarediseasesnetwork.org/) to address the unique challenges of research on rare diseases. RDCRN studies more than 90 rare diseases at about 100 academic institutions. Patient advocacy groups actively participate in the research.

The NGS technologies have driven a dramatic shift in our understanding of rare diseases at a genome-wide scale [71]. Bioinformatics play a central role and has become an important component in NGS data analysis, generating many algorithms and workflows. However, building a standard bioinformatics solution for NGS analysis and application to clinical practice remains to be carried out. Accurate and reliable NGS analysis ensures patients with rare diseases receive the correct diagnosis. Accurate and reliable NGS further facilitates the practice of precision medicine. However, inaccurate NGS testing can lead to poor or misleading results. Therefore, the drug makers, scientific researches, and reviewers need to collaboratively to standardize NGS techniques and performance evaluation approaches. The FDA (https://precision.fda.gov/) and the NIH Precision Medicine Initiative Cohort Program (https://www.nih.gov/precision-medicine-initiative-cohort-program) have been created to provide insightful vision on precision medicine taking advantage of emerging techniques.

Government-sponsored initiatives and accompanying policy shifts have also had a great impact on the development of treatments for rare diseases. For example, FDA awarded 18 new research grants for the development of rare disease products or biomarkers or to efray the cost of clinical trials

(http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm463539.htm). So far

40 orphan disease products were partially funded by grants from the “orphan products grant program”. Furthermore, the FDA has developed four distinct and promising routes, which enacts faster drug review and approval, shortening rare disease therapy development. In addition, there are other government-sponsored initiatives such as the Medical Research Council in the United Kingdom and the NIH National Center for Advancing Translational Sciences (NCATS). The NIH has established partnerships among public funders, the pharmaceutical industry and academic investigators, which will also be beneficial for the development of therapies for rare diseases [72].

Closing Remarks

In summary, under the precision medicine umbrella, the landscape of rare diseases has been redrawn by applying NGS techniques. The accumulated genomic data provides great opportunities for the development of treatments for rare diseases by providing insight into the possibility of drug repositioning. Enabling the translation of these novel findings to clinical practice of rare disease treatment development is the real practice of precision medicine. Several promising bioinformatics approaches as summarized, have shown great potential in tailoring genomic findings to developing therapies for rare diseases. Combined with other established drug repositioning approaches and efforts form scientific communities, government agencies, and pharmaceutical companies, the timing is excellent for furthering the development of innovative approaches and clinical practice towards precision medicine for rare diseases.

Competing interests

Dr. Ruth Roberts is co-founder and co-director of Apconix, an integrated toxicology and ion channel company that provides expert advice on nonclinical aspects of drug discovery and drug development to academia, industry and non-for-profit organisations.

Acknowledgement

We thank Dr. Carole Christian from The Office of Portfolio Analysis, Division of Program Coordination, Planning, and Strategic Initiatives at National Institutes of Health for her help in editing.

References

1. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE: **Rare-disease genetics in the era of next-generation sequencing: discovery to translation**. *Nat Rev Genet* 2013, **14**(10):681-691.
2. Basch E: **The Missing Voice of Patients in Drug-Safety Reporting**. *New England Journal of Medicine* 2010, **362**(10):865-869.
3. Lesko LJ, Atkinson AJ: **USE OF BIOMARKERS AND SURROGATE ENDPOINTS IN DRUG DEVELOPMENT AND REGULATORY DECISION MAKING: Criteria, Validation, Strategies1**. *Annual Review of Pharmacology and Toxicology* 2001, **41**(1):347-366.
4. Mullard A: **2012 FDA drug approvals**. *Nat Rev Drug Discov* 2013, **12**(2):87-90.
5. Schadow G: **Assessing the impact of HL7/FDA Structured Product Label (SPL) content for medication knowledge management**. In: *AMIA Annual Symposium Proceedings: 2007*. American Medical Informatics Association: 646.
6. Haffner ME: **Adopting Orphan Drugs — Two Dozen Years of Treating Rare Diseases**. *New England Journal of Medicine* 2006, **354**(5):445-447.
7. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders**. *Nucleic Acids Research* 2005, **33**(suppl 1):D514-D517.
8. Rowe SM, Verkman AS: **Cystic Fibrosis Transmembrane Regulator Correctors and Potentiators**. *Cold Spring Harbor Perspectives in Medicine* 2013, **3**(7).
9. Rowe SM, Miller S, Sorscher EJ: **Cystic Fibrosis**. *New England Journal of Medicine* 2005, **352**(19):1992-2001.
10. Sboner A, Elemento O: **A primer on precision medicine informatics**. *Briefings in Bioinformatics* 2015.
11. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR: **ClinVar: public archive of relationships among sequence variation and human phenotype**. *Nucleic Acids Research* 2014, **42**(D1):D980-D985.
12. Cheung-Ong K, Giaever G, Nislow C: **DNA-Damaging Agents in Cancer Chemotherapy: Serendipity and Chemical Biology**. *Chemistry & Biology* 2013, **20**(5):648-659.
13. Potter BK, Khangura SD, Tingley K, Chakraborty P, Little J: **Translating rare-disease therapies into improved care for patients and families: what are the right outcomes, designs, and engagement approaches in health-systems research?** *Genet Med* 2016, **18**(2):117-123.
14. Jamuar SS, Tan E-C: **Clinical application of next-generation sequencing for Mendelian diseases**. *Human Genomics* 2015, **9**(1):10.
15. Briggs MD, Bell PA, Wright MJ, Pirog KA: **New therapeutic targets in rare genetic skeletal diseases**. *Expert Opinion on Orphan Drugs* 2015, **3**(10):1137-1154.
16. Segalat L: **Loss-of-function genetic diseases and the concept of pharmaceutical targets**. *Nature Reviews Genetics* 2007, **8**(5).
17. Ashburn TT, Thor KB: **Drug repositioning: Identifying and developing new uses for existing drugs**. *Nature Reviews Drug Discovery* 2004, **3**(8):673-683.
18. Liu Z, Fang H, Reagan K, Xu X, Mendrick DL, Slikker Jr W, Tong W: **In silico drug repositioning – what we need to know**. *Drug Discovery Today* 2013, **18**(3–4):110-115.
19. Li Y, Jones S: **Drug repositioning for personalized medicine**. *Genome Medicine* 2012, **4**(3):27.

20. Ekins S, Williams AJ, Krasowski MD, Freundlich JS: **In silico repositioning of approved drugs for rare and neglected diseases**. *Drug Discovery Today* 2011, **16**(7-8):298-310.
21. Sardana D, Zhu C, Zhang M, Gudivada RC, Yang L, Jegga AG: **Drug repositioning for orphan diseases**. *Briefings in Bioinformatics* 2011, **12**(4):346-356.
22. Weischenfeldt J, Symmons O, Spitz F, Korbel JO: **Phenotypic impact of genomic structural variation: insights from and for human disease**. *Nat Rev Genet* 2013, **14**(2):125-138.
23. Schumacher KR, Stringer KA, Donohue JE, Yu S, Shaver A, Caruthers RL, Zikmund-Fisher BJ, Fifer C, Goldberg C, Russell MW: **Social media methods for studying rare diseases**. *Pediatrics* 2014, **133**(5):e1345-e1353.
24. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C: **Ten years of next-generation sequencing technology**. *Trends in Genetics*, **30**(9):418-426.
25. Aymé S: **Orphanet, an information site on rare diseases**. *Soins; la revue de référence infirmière* 2003(672):46.
26. Wei C-Y, Michael Lee M-T, Chen Y-T: **Pharmacogenomics of adverse drug reactions: implementing personalized medicine**. *Human Molecular Genetics* 2012, **21**(R1):R58-R65.
27. Frueh FW, Amur S, Mummaneni P, Epstein RS, Aubert RE, DeLuca TM, Verbrugge RR, Burckart GJ, Lesko LJ: **Pharmacogenomic Biomarker Information in Drug Labels Approved by the United States Food and Drug Administration: Prevalence of Related Drug Use**. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 2008, **28**(8):992-998.
28. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits**. *Proceedings of the National Academy of Sciences* 2009, **106**(23):9362-9367.
29. Plenge RM, Scolnick EM, Altshuler D: **Validating therapeutic targets through human genetics**. *Nat Rev Drug Discov* 2013, **12**(8):581-594.
30. Wang Z-Y, Zhang H-Y: **Rational drug repositioning by medical genetics**. *Nat Biotech* 2013, **31**(12):1080-1082.
31. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J *et al*: **The support of human genetic evidence for approved drug indications**. *Nat Genet* 2015, **47**(8):856-860.
32. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, Mooser V: **Use of genome-wide association studies for drug repositioning**. *Nat Biotech* 2012, **30**(4):317-320.
33. Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P: **Computational Drug Repositioning: From Data to Therapeutics**. *Clinical Pharmacology & Therapeutics* 2013, **93**(4):335-341.
34. Korte A, Farlow A: **The advantages and limitations of trait analysis with GWAS: a review**. *Plant Methods* 2013, **9**(1):1-9.
35. Roden DM, Denny JC: **Integrating electronic health record genotype and phenotype datasets to transform patient care**. *Clinical Pharmacology & Therapeutics* 2015.
36. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E *et al*: **Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data**. *Nat Biotech* 2013, **31**(12):1102-1111.
37. Rastegar-Mojarad M, Ye Z, Kolesar JM, Hebbring SJ, Lin SM: **Opportunities for drug repositioning from phenome-wide association studies**. *Nat Biotech* 2015, **33**(4):342-345.
38. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V *et al*: **DrugBank 4.0: shedding new light on drug metabolism**. *Nucleic Acids Research* 2014, **42**(D1):D1091-D1097.
39. Jin G, Wong ST: **Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines**. *Drug discovery today* 2014, **19**(5):637-644.

40. Dhillon AS, Hagan S, Rath O, Kolch W: **MAP kinase signalling pathways in cancer**. *Oncogene* 0000, **26**(22):3279-3290.
41. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA *et al*: **Mutational landscape and significance across 12 major cancer types**. *Nature* 2013, **502**(7471):333-339.
42. Marin TM, Keith K, Davies B, Conner DA, Guha P, Kalaitzidis D, Wu X, Lauriol J, Wang B, Bauer M *et al*: **Rapamycin reverses hypertrophic cardiomyopathy in a mouse model of LEOPARD syndrome–associated PTPN11 mutation**. *The Journal of Clinical Investigation*, **121**(3):1026-1043.
43. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L: **The human disease network**. *Proceedings of the National Academy of Sciences* 2007, **104**(21):8685-8690.
44. Li Y, Agarwal P: **A Pathway-Based View of Human Diseases and Disease Relationships**. *Plos One* 2009, **4**(2).
45. Rockman MV, Kruglyak L: **Genetics of global gene expression**. *Nat Rev Genet* 2006, **7**(11):862-872.
46. Montgomery SB, Dermitzakis ET: **The resolution of the genetics of gene expression**. *Human Molecular Genetics* 2009, **18**(R2):R211-R215.
47. Natsoulis G, Pearson CI, Gollub J, Eynon BP, Ferng J, Nair R, Idury R, Lee MD, Fielden MR, Brennan RJ *et al*: **The liver pharmacological and xenobiotic gene response repertoire**. *Molecular Systems Biology* 2008, **4**.
48. Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, Yamada H: **Open TG-GATEs: a large-scale toxicogenomics database**. *Nucleic Acids Research* 2015, **43**(D1):D921-D927.
49. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN *et al*: **The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease**. *Science* 2006, **313**(5795):1929-1935.
50. Masica DL, Karchin R: **Correlation of Somatic Mutation and Expression Identifies Genes Important in Human Glioblastoma Progression and Survival**. *Cancer Research* 2011, **71**(13):4550-4561.
51. Bertrand D, Chng KR, Sherbaf FG, Kiesel A, Chia BKH, Sia YY, Huang SK, Hoon DSB, Liu ET, Hillmer A *et al*: **Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles**. *Nucleic Acids Research* 2015, **43**(7).
52. Ping Y, Deng YL, Wang L, Zhang HY, Zhang Y, Xu CH, Zhao HY, Fan HH, Yu FL, Xiao Y *et al*: **Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data**. *Nucleic Acids Research* 2015, **43**(4):1997-2007.
53. Ding J, McConechy MK, Horlings HM, Ha G, Chun Chan F, Funnell T, Mullaly SC, Reimand J, Bashashati A, Bader GD *et al*: **Systematic analysis of somatic mutations impacting gene expression in 12 tumour types**. *Nat Commun* 2015, **6**.
54. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MGD, Jädersten M, Dolatshad H, Verma A, Cross NCP, Vyas P *et al*: **Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes**. *Nat Commun* 2015, **6**.
55. Soifer HS, Rossi JJ, Saetrom P: **MicroRNAs in Disease and Potential Therapeutic Applications**. *Mol Ther* 2007, **15**(12):2070-2079.
56. Amato F, Seia M, Giordano S, Elce A, Zarrilli F, Castaldo G, Tomaiuolo R: **Gene Mutation in MicroRNA Target Sites of CFTR Gene: A Novel Pathogenetic Mechanism in Cystic Fibrosis?** *Plos One* 2013, **8**(3).
57. Liu Z, Borlak J, Tong W: **Deciphering miRNA transcription factor feed-forward loops to identify drug repurposing candidates for cystic fibrosis**. *Genome Medicine* 2014, **6**(12):94.

58. Qu XA, Rajpal DK: **Applications of Connectivity Map in drug discovery and development**. *Drug Discovery Today* 2012, **17**(23–24):1289-1298.
59. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A *et al*: **Discovery of drug mode of action and drug repositioning from transcriptional responses**. *Proceedings of the National Academy of Sciences* 2010, **107**(33):14621-14626.
60. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, Morgan AA, Sarwal MM, Pasricha PJ, Butte AJ: **Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease**. *Science Translational Medicine* 2011, **3**(96).
61. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: **NCBI GEO: mining tens of millions of expression profiles - database and tools update**. *Nucleic Acids Research* 2007, **35**:D760-D765.
62. Iskar M, Zeller G, Blattmann P, Campillos M, Kuhn M, Kaminska KH, Runz H, Gavin AC, Pepperkok R, van Noort V *et al*: **Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding**. *Molecular Systems Biology* 2013, **9**.
63. Ling H, Fabbri M, Calin GA: **MicroRNAs and other non-coding RNAs as targets for anticancer drug development**. *Nat Rev Drug Discov* 2013, **12**(11):847-865.
64. Li Z, Rana TM: **Therapeutic targeting of microRNAs: current status and future challenges**. *Nat Rev Drug Discov* 2014, **13**(8):622-638.
65. Liu XY, Wang SY, Meng FL, Wang JZ, Zhang Y, Dai EY, Yu XX, Li X, Jiang W: **SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression**. *Bioinformatics* 2013, **29**(3):409-411.
66. Rukov JL, Wilentzik R, Jaffe I, Vinther J, Shomron N: **Pharmaco-miR: linking microRNAs and drug effects**. *Briefings in Bioinformatics* 2014, **15**(4):648-659.
67. Panic G, Vargas M, Scandale I, Keiser J: **Activity Profile of an FDA-Approved Compound Library against Schistosoma mansoni**. *Plos Neglected Tropical Diseases* 2015, **9**(7).
68. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P: **Drug Target Identification Using Side-Effect Similarity**. *Science* 2008, **321**(5886):263-266.
69. Clohessy JG, Pandolfi PP: **Mouse hospital and co-clinical trial project[mdash]from bench to bedside**. *Nat Rev Clin Oncol* 2015, **12**(8):491-498.
70. Lynam EB, Leaw J, Wiener MB: **A Patient Focused Solution for Enrolling Clinical Trials in Rare and Selective Cancer Indications: A Landscape of Haystacks and Needles**. *Drug information journal* 2012, **46**(4):472-478.
71. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER: **The Next-Generation Sequencing Revolution and Its Impact on Genomics**. *Cell* 2013, **155**(1):27-38.
72. Frail DE, Brady M, Escott KJ, Holt A, Sanganee HJ, Pangalos MN, Watkins C, Wegner CD: **Pioneering government-sponsored drug repositioning collaborations: progress and learning**. *Nat Rev Drug Discov* 2015, **14**(12):833-841.

Table 1 Public available resources and efforts in rare disease genetics

Databases/Consortiums	Web link	Remarks
Public databases		
Orphanet	http://www.orphadata.org/cgi-bin/index.php	Orphanet is a comprehensive resource on rare diseases, which provides rare disease information including rare disease associated genes, clinical signs, epidemiological data and rare disease classification.
Online Mendelian Inheritance in Man (OMIM)	http://www.omim.org/	OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes, which contains the information about all the mendelian diseases and over 15,000 genes and their variants
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/intro/	ClinVar is provides a free accessible, reported relationships among human variations and phenotypes, with supporting evidence. The human variants information is also linked to Orphanet and OMIM databases.
COSMIC	http://cancer.sanger.ac.uk/cosmic	COSMIC is designed to store and display somatic mutation information and related details and contains information relating to human cancers. The somatic mutation on rare cancers could be retrieved from COSMIC
Database of genomic variation and phenotype in humans using ensemble Resources (DECIPHER)	https://decipher.sanger.ac.uk/	DECIPHER is an interactive web-based database which incorporates a suite of tools designed to aid the interpretation of genomic variants of rare disease.
The NHGRI GWAS Catalog	www.genome.gov/gwastudies	A Catalog of Published Genome-Wide Association Studies that provides SNP-traits association
DisGeNET	http://www.disgenet.org/web/DisGeNET/menu	DisGeNET is a curated efforts and aim to integrate the disease and gene relationship from public database and literature mining.
Consortiums efforts		
Care for Rare	http://care4rare.ca/	CARE for RARE is a Canadian nation-wide research program focusing on the improvement of both the diagnosis and treatment of rare diseases. Currently, their researches embrace 637 different rare disease studies with more than 1000 rare disease patients with 81 novel rare diseases causing genes identified.
Finding of Rare Disease Genes in Canada	http://www.cpgdsconsortium.com/default.aspx	FORGE Canada (Finding of Rare Disease Genes) is a

(FORGE Canada)		national consortium of clinicians and scientists using next-generation sequencing technology to identify genes responsible for a wide spectrum of rare pediatric-onset disorders present in the Canadian population.
The Centers for Mendelian Genomics (CMG)	http://www.mendelian.org/	The CMG aim to discover genetic basis of Mendelian disorders in two main ways including applying novel sequencing technique for rare diseases researches and collaboration with other rare disease research consortiums.
The Global Alliance for Genomics and Health	https://genomicsandhealth.org/	The Global Alliance for Genomics and Health (Global Alliance) was formed to help accelerate the potential of genomic medicine to advance human health by using emerging sequencing technique.
The UK 100,000 Genomes Project	http://www.genomicsengland.co.uk/	The project will sequence 100,000 genomes from around 70,000 people. Participants are NHS patients with a rare disease, plus their families, and patients with cancer.
Deciphering Developmental Disorders (DDD)	http://www.ddduk.org/	The aim of the DDD study is to advance clinical genetic practice for children with developmental disorders by the systematic application of the latest microarray and sequencing methods while addressing the new ethical challenges raised.
The International Rare Diseases Research Consortium (IRDiRC)	http://www.irdirc.org/	IRDiRC teams up researchers and organizations investing in rare diseases research in order to achieve two main objectives by the year 2020, namely to deliver 200 new therapies for rare diseases and means to diagnose most rare diseases.
The Genetic Disorders of Mucociliary Clearance Consortium (GDMCC)	http://www.rarediseasesnetwork.org/cms/GDMCC	The Genetic Disorders Of Mucociliary Clearance Consortium is a clinical research network created to improve the diagnostic testing and treatment of rare airway diseases, including primary ciliary dyskinesia (PCD), variant forms of cystic fibrosis (CF), pseudohypoaldosteronism (PHA), and now idiopathic bronchiectasis and NTM pulmonary disease.

Table 2 Data integration strategy for correlating genetic mutation and gene expression

Data profiles	Diseases	Methodology	Notes	PubMed ID
Somatic mutation and gene expression	glioblastoma (GBM)	Fisher's exact test with network analysis	The somatic mutation and gene expression are needed for each patients	21555372
SNP, CNV and gene expression	glioblastoma (GBM) and other five cancer types	Network analysis for integrative data including SNP, CNV and gene expression for driver genes enrichment	OncoIMPACT is developed and source code is available from http://sourceforge.net/projects/oncoimpact.	25572314
CNV, methylation, miRNA and gene expression	glioblastoma (GBM) and other six cancer types	Partial least squares regression and network analysis	The results were verified by survival analysis and a core gene module of 17 genes was enriched for candidate GBM driver genes.	25653168
Somatic mutation and gene expression	12 pan cancers	hierarchical Bayes statistical model http://compbio.bccrc.ca/software/xseq/	Patient genetic heterogeneity was observed and the some mutation types was conserved across cancer types	26436532
Germline mutation, miRNA, transcription factor, gene expression	Cystic fibrosis	miRNA transcription factor feed-forward loop construction by cumulative hypergeometric test	The 48 repurposing candidate were enriched for cystic fibrosis treatment, 26 of 48 candidates were verified by literature survey or existing clinical trails	25484921
Somatic mutation and gene expression	myelodysplastic syndromes and acute myeloid leukaemia (AML)	principal component analysis (PCA) with schematic linear decomposition	One or more genetic lesions correlate with expression levels of ~20% of all genes, explaining 20–65% of observed expression variability. Differential expression patterns vary between mutations and reflect the underlying biology	25574665

Table 3 Drug transcriptome and drug–miRNAs relationship resources

Databases	Web link	Remarks
The Connectivity Map (CMap)	https://www.broadinstitute.org/cmap/	Provide an comprehensive drug transcriptional responses of 1309 drugs or lead compounds in the clinical trials to six or seven different cancer cell lines
Open TG-GATEs	http://toxico.nibio.go.jp/english/index.html	TG-GATEs consists of the comprehensive toxicogenomic profiles of 170 compounds with four different assay types (human/rat in vitro/vivo) and multiple time and dose points in rat liver and kidney. The histopathological profiles for compounds are also available.
DrugMatrix	https://ntp.niehs.nih.gov/drugmatrix/index.html	DrugMatrix contains toxicogenomic profiles for 638 different compounds from both Codelink and Affymetrix platforms, which covers multiple organism including liver, kidney, heart, bone marrow, spleen and skeletal muscle.
SM2miR	http://210.46.85.180:8080/sm2mir/index.jsp	SM2miR is a manual curated database which collects and incorporates the experimentally validated small molecules and miRNA relationship from around twenty species by literature survey.
Pharmaco-miR	http://www.pharmaco-mir.org/	Pharmaco-miR identifies associations of miRNAs, genes and drugs by integrating PharmaGKB database and *in silico* prediction.

Figure Captions:

Figure 1 **The proposed computational drug repositioning approaches for rare disease therapy**

Figure 2 **The statistics of rare diseases genetic information**: (A) The relationship between rare disease and its causative genes based on Orphadata; (B) the known mutation origin and functions of rare diseases based on Orphadata; (C) the structure variants distribution of rare diseases based on ClinVar.

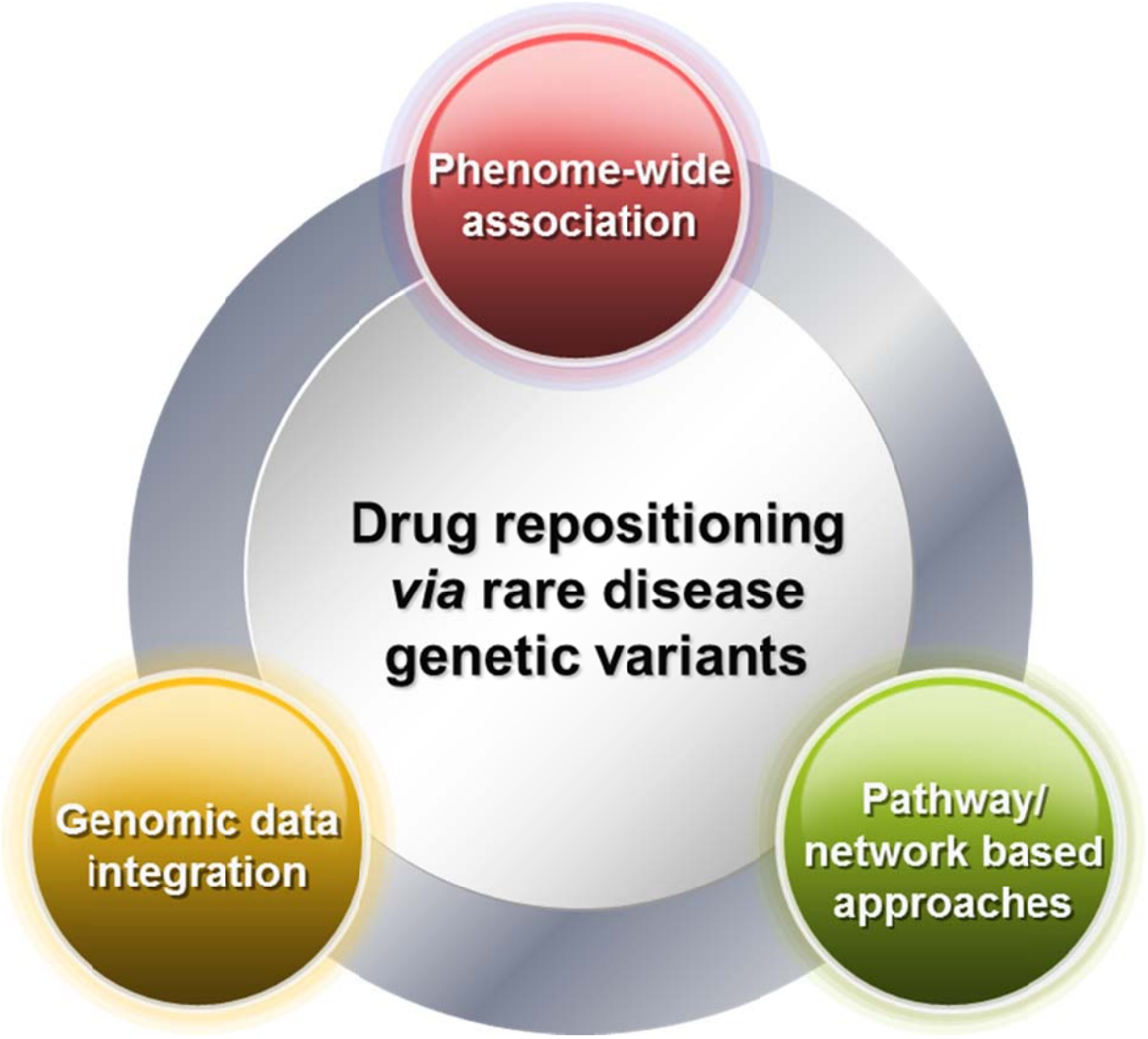
Phenome-wide association
Drug repositioning *via* rare disease genetic variants
Genomic data integration
Pathway/ network based approaches

Figure 1

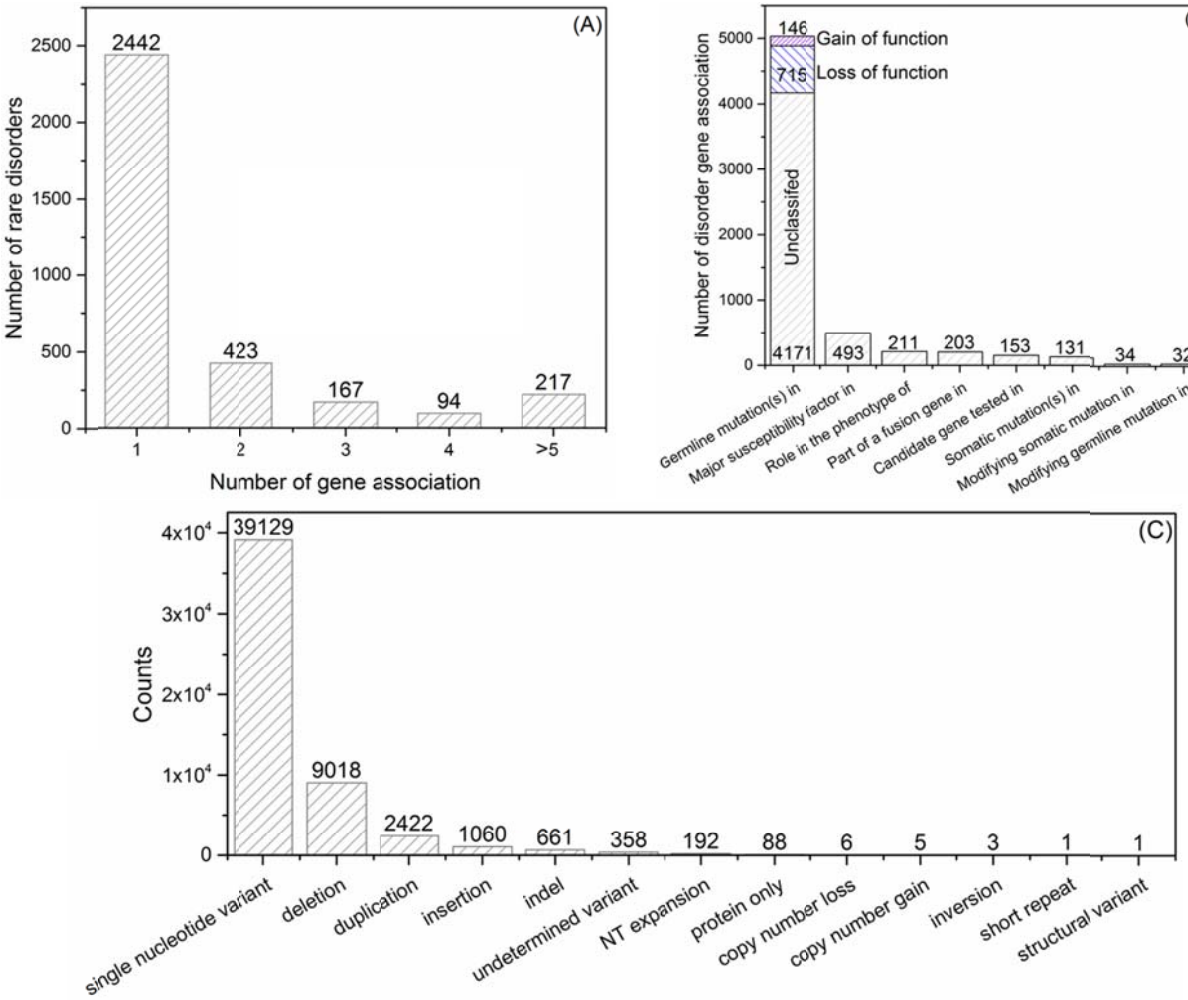
(A)
Number of rare disorders
2442
423
167
94
217
1
2
3
4
>5
Number of gene association
(B)
Number of disorder gene association
146
Gain of function
715
Loss of function
Unclassifed
4171
493
211
203
153
131
34
32
Germline mutation(s) in
Major susceptibility factor in
Role in the phenotype of
Part of a fusion gene in
Candidate gene tested in
Somatic mutation(s) in
Modifying somatic mutation in
Modifying germline mutation in
(C)
Counts
39129
9018
2422
1060
661
358
192
88
6
5
3
1
1
single nucleotide variant
deletion
duplication
insertion
indel
undetermined variant
NT expansion
protein only
copy number loss
copy number gain
inversion
short repeat
structural variant

Figure 2