Bioinformatics, 35(19), 2019, 3821–3823
doi: 10.1093/bioinformatics/btz166
Advance Access Publication Date: 11 March 2019
Applications Note

Genome analysis

GREP: genome for REPositioning drugs

Saori Sakaue[1,2,3] and Yukinori Okada[1,2,4,*]

[1]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, [2]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan, [3]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-8655, Japan and [4]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on June 14, 2018; revised on February 14, 2019; editorial decision on March 6, 2019; accepted on March 9, 2019

Abstract

Summary: Making use of accumulated genetic knowledge for clinical practice is our next goal in human genetics. Here we introduce GREP (**G**enome for **REP**ositioning drugs), a standalone python software to quantify an enrichment of the user-defined set of genes in the target of clinical indication categories and to capture potentially repositionable drugs targeting the gene set. We show that genes identified by the large-scale genome-wide association studies were robustly enriched in the approved drugs to treat the trait of interest. This enrichment analysis was also highly applicable to other sets of biological genes such as those identified by gene expression studies and genes somatically mutated in cancers. This software should accelerate investigators to reposition drugs to other indications with the guidance of known genomics.

Availability and implementation: GREP is available at https://github.com/saorisakaue/GREP as a python source code.

Contact: yokada@sg.med.osaka-u.ac.jp

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

After the fruitful decades of genome-wide association studies (GWASs) identifying thousands of loci robustly associated with human complex traits (Welter *et al.*, 2014), we are now heading to the next step to apply *in silico* genetic knowledge for clinical practice. Given the fact that more than half of clinical trials fail because of lack of efficacy or adverse events (Nelson *et al.*, 2015), the method to discover new therapeutics with the guidance of genomics should be warranted (Malik *et al.*, 2018; Okada *et al.*, 2014). Repositioning known drugs to another indication is an effective way to bring them to bedside, because their efficacy and potential adverse effects have already been investigated for the current indication.

Here we introduce a software GREP (**G**enome for **REP**ositioning drugs), which (i) quantifies an enrichment of the user-defined set of genes in the target of clinical indication categories and (ii) captures potentially repositionable drugs targeting the gene set. By applying GREP to a large-scale GWAS of stroke, a gene expression analysis on human tissues, and somatic mutations in cancers, we show that genes implicated in the GWAS were robustly enriched in the indicated medications for the trait of interest. GREP also unraveled new insights into the biology of the tissue-specific gene expression and the genetic landscape of cancers.

2 Materials and methods

The overview of our method is shown in Figure 1a. We have collected and curated target chemical information on medications of current use or developed in the past, from two major drug databases, Drug Bank (Wishart *et al.*, 2018) and Therapeutic Target Database (Li *et al.*, 2018). The detailed process for the curation is described in Supplementary Material, and the source code for reformatting the downloaded drug database is distributed with the software so that users can update drug database information on their own. Target chemical information was converted to gene symbols, which made information on 22 300 drugs and 2029 genes

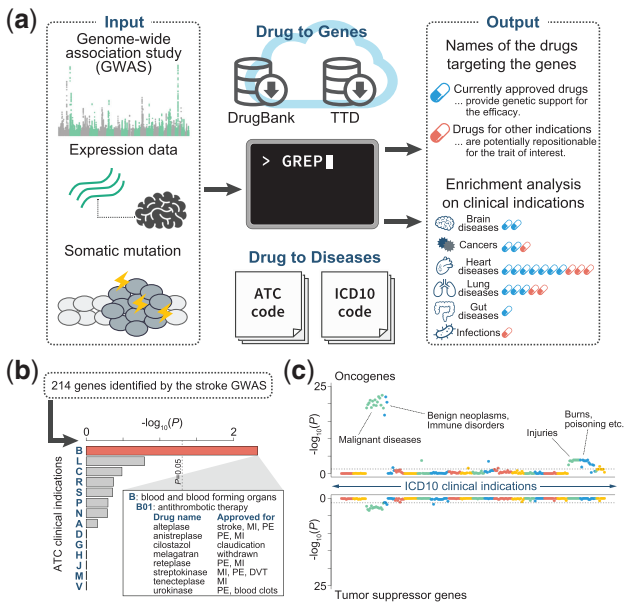

Fig. 1. The overview of GREP and its application to GWAS and cancer somatic mutations. **(a)** Schematic presentation of GREP software. TTD, Therapeutic Target Database; ATC, Anatomical Therapeutic Chemical Classification; ICD10, International Classification of Diseases 10. **(b)** Enrichment analysis of stroke-risk suggestive genes by GREP. The inset shows the approved drugs targeting the gene set and their current indications. MI, myocardial infarction; PE, pulmonary embolism; DVT, deep vein thrombosis. **(c)** Comparative enrichment analysis of oncogenes and tumor suppressor genes by GREP. Each circle represents an enrichment *P* value of the drug indication category by ICD10

of their target in total. We next categorized these drugs by their clinical indication for treatment based on the two widely used classification systems Anatomical Therapeutic Chemical Classification System (ATC), World Health Organization (WHO) and International Classification of Diseases 10 (ICD10) diagnostic code.

Using a text file of genes as a user input together with these pre-formatted databases, the GREP software automatically performs a series of Fisher's exact tests to examine whether the gene set is enriched in genes targeted by drugs in a clinical indication category [by ATC or ICD10] to treat a certain disease or condition. Further, GREP outputs the names of the drugs targeting the gene set, which are thus considered to have an association with the basis on which the gene set was selected. Our method is novel in that (i) users can input a gene set of any origin, and (ii) the drugs in the database are organized by their clinical indications, so as to grasp the whole picture of the pharmacological enrichment. We note that our implementation to identify pharmacologically associated drug indication classes or anatomical classes from the flexible gene set is the novel feature when compared with the previous gene-drug-disease databases (as in Koscielny *et al.*, 2017). GREP is implemented as a python module and runs all the analysis in a few seconds on Linux or Mac OS environment. The benchmarking shows that the standard GREP analysis takes 1.71 s on MacBook Pro (2.4 GHz, Intel Core i7) using a single CPU. To underscore the applicability of our tool, we tested a wide range of gene sets including large-scale GWASs, tissue-specific gene expression data released by GTEx Consortium (GTEx Consortium *et al.*, 2017), and a somatic mutation catalog of various cancers by COSMIC Consortium (Forbes *et al.*, 2017).

3 Results

First, as an illustrative example, we applied GREP to the risk genes identified by a recent multi-ancestry genome-wide association meta-analysis of stroke by MEGASTROKE consortium (Malik *et al.*, 2018). Using a set of 214 stroke-risk suggestive genes as an input, GREP enrichment analysis showed that these risk genes were enriched in the targets of approved medications for 'blood and blood forming organs' in ATC group B (odds ratio = 5.3, Fisher's exact *P* value = 4.7×10^{-3}; Fig. 1b) This enrichment was driven by the indicated drugs for the 'antithrombotic therapy', replicating the result we have reported in the original article. In addition to the enrichment, GREP also revealed that the drugs targeting the GWAS-identified genes were currently approved for stroke and other thrombotic diseases (e.g. alteplase, anistreplase cilostazol, melagatran, reteplase, streptokinase, tenecteplase and urokinase). This list of drugs would support the approved drugs' efficacy on the indicated trait itself from genetics on one hand, and further provide a potential for repositioning the drugs if they are yet to be indicated for the trait under investigation on the other hand. We also note that the enrichment result suggests which group of drugs is likely to be supported by genetics, supposedly useful to prioritize the candidate drugs.

Second, to expand the applicability of GREP from genome to transcriptome, we utilized tissue-specifically expressed genes (SEGs) obtained from GTEx expression data as an input gene set for GREP. We defined top 5% of all genes as SEGs in each tissue as described elsewhere (Finucane *et al.*, 2018). Supplementary Figure S1 illustrates in what drug categories defined by ATC the input SEGs in each tissue are enriched. Here again we observed strong associations of the tissue-SEGs and disease categories of relevance to the tissue, such as SEGs in brain cortex and drugs used for nervous system diseases, SEGs in aorta and those for cardiovascular diseases, and SEGs in stomach and colon and those indicated for alimentary tract diseases. Therefore, the drugs targeting the SEGs in each tissue are likely to be approved for treating the diseases linked to that tissue.

Third, to further broaden the application to somatic mutations, we investigated whether somatically mutated genes associated with cancers have any enrichment in known drug indications. By separately assessing oncogenes and tumor suppressor genes, while both of the gene sets showed the strongest enrichment in the medications to treat malignant neoplasms, we highlighted the striking contrast of the strength in enrichment, oncogenes being much stronger (Fig. 1c). This observation would recapitulate the fact that current anti-cancer drugs are rationally designed to alter or suppress aberrant oncogene activity (Yildirim *et al.*, 2007), and might further help researchers respond to an emergent need for new strategies to target tumor suppressor genes (Ashworth *et al.*, 2011).

As a potential limitation, we note that the current version of the software does not account for the directional effects of both genes and drugs. Future implementation of such information with the advent of comprehensive catalog of the functional annotation of variants and drug databases would make our method further fine-tuned and practical for repurposing.

4 Conclusions

We implemented an easy-to-use standalone software, GREP and successfully depicted the relationships between a wide range of genomic knowledge and clinical indication categories of medications of current use. Further, this software should guide investigators to determine the best path forward for the efficient drug discovery.

Acknowledgements

We thank Dr Robert M. Plenge, Prof Kazuhiko Yamamoto and Dr Yoichiro Kamatani for their kind supports on the study.

Funding

This study is supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI [15H05911], AMED [16gm6010001h0001 and 17ek0410041h0001].

Conflict of Interest: none declared.

References

Ashworth,A. *et al*. (2011) Genetic interactions in cancer progression and treatment. *Cell*, **145**, 30–38.

Finucane. *et al*. (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet*., **50**, 621–629.

Forbes,S.A. *et al*. (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*., **45**, D777–D783.

GTEx Consortium. *et al*. (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.

Koscielny,G. *et al*. (2017) Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res*., **45**, D985–D994.

Li,Y.H. *et al*. (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res*., **46**, D1121–D1127.

Malik,R. *et al*. (2018) Multiancestry genome-wide association study of 520, 000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet*., **50**, 524–537.

Nelson,M.R. *et al*. (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet*., **47**, 856–860.

Okada,Y. *et al*. (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376–381.

Welter,D. *et al*. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*., **42**, D1001–D1006.

Wishart,D.S. *et al*. (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*., **46**, D1074–D1082.

Yildirim,M.A. *et al*. (2007) Drug-target network. *Nat. Biotechnol*., **25**, 1119–1126.