OXFORD

Systems biology

polyPK: an R package for pharmacokinetic analysis of multi-component drugs using a metabolomics approach

Mengci Li[1,2], Shouli Wang[1], Guoxiang Xie[3], Xiaohui Ma[4], Tianlu Chen[1,]*
and Wei Jia[1,3,]*

[1]Center for Translational Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China, [2]School of Biomedical Engineering and Med-X Research Institute, Shanghai Jiao Tong University, Shanghai 200030, China, [3]University of Hawaii Cancer Center, Honolulu 96813, USA and [4]State Key Laboratory of Core Technology in Innovative Chinese Medicine, Pharmacology and Toxicology Research Centre, Tasly Academy, Tasly Holding Group Co., Ltd, Tianjin 300410, China

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Abstract

Summary: Pharmacokinetics (PK) is a long-standing bottleneck for botanical drug and traditional medicine research. By using an integrated phytochemical and metabolomics approach coupled with multivariate statistical analysis, we propose a new strategy, Poly-PK, to simultaneously monitor the performance of drug constituents and endogenous metabolites, taking into account both the diversity of the drug's chemical composition and its complex effects on the mammalian metabolic pathways. Poly-PK is independent of specific measurement platforms and has been successfully applied in the PK studies of Puerh tea, a traditional Chinese medicine Huangqi decoction and many other multi-component drugs. Here, we introduce an R package, polyPK, the first and only automation of the data analysis pipeline of Poly-PK strategy. polyPK provides 10 functions for data pre-processing, differential compound identification and grouping, traditional PK parameters calculation, multivariate statistical analysis, correlations, cluster analyses and resulting visualization. It may serve a wide range of users, including pharmacologists, biologists and doctors, in understanding the metabolic fate of multi-component drugs.

Availability and implementation: polyPK package is freely available from the R archive CRAN (https://CRAN.R-project.org/package=polyPK).

Contact: wjia@cc.hawaii.edu or chentianlu@sjtu.edu.cn

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

1 Introduction

The pharmacokinetics (PK) of multi-component drugs such as herbal medicines and other natural products is more complex than compounds containing a single active component. From 2010 on, our group has proposed, step by step, a metabolomics-based strategy, Poly-PK, which is very comprehensive in exploring complex effects of multi-component drugs in the mammalian metabolic system (Jia *et al.*, 2015; Lan *et al.*, 2010, 2013; Xie *et al.*, 2012, 2017). Using this strategy, researchers will be able to monitor the performance of both drug constituents and endogenous metabolites. Its core step is to identify numerous differential metabolites and to divide them into three groups: (i) altered endogenous metabolites caused by exogenous drugs, (ii) absorbed drug-derived compounds and (iii) secondary metabolites generated by the chemical transformation of

drug compounds by hepatic enzymes and gut microbes (Jia *et al.*, 2015). Currently, this strategy has been successfully applied in examining the complex PK and pharmacodynamics profiles of a Chinese tea (Xie *et al.*, 2012), a Chinese herbal medicine Huangqi decoction (Xie *et al.*, 2017) as well as, various multi-component drugs and shows prominent advantages over conventional approaches (Lan *et al.*, 2013). Considering the hard and complex data analysis burden, especially for those not familiar to metabolomics and PK study, we developed polyPK, an R package, to automate the data analysis pipeline of Poly-PK strategy (Fig. 1). To our knowledge, this is the first and only integrated tool that uses metabolomics approach for pharmacokinetic analysis and visualization on multi-component drugs.

2 Implementation and main functions

2.1 Data importation and pre-processing

Three data files, derived from any measurement platforms (e.g. NMR, GC/LC-MS), are required for a complete polyPK study (Fig. 1A): (i) list of drug constituents, (ii) pre- and (iii) post-dose (multiple time-points) metabolomics data. A file of basic information is optional. The formats and demo contents of the four data files are illustrated in Supplementary Figures S1–S4. All the functions are independent and if some of the data files are not imported, only the related functions and results will be affected.

The pre- and post-dose metabolomics datasets are processed by the function 'DataPre'. Outliers and sparse variables are removed and then missing values are imputed (Fig. 1B). Various control parameters are ready for users with diverse aims and detailed descriptions are provided in SI. The default setting is optimal for most cases. For left-censored missing values (e.g. quantitative metabolomics data), 'QRILC' is the best imputation method.

2.2 Differential compounds screening and grouping

The core part of polyPK is noted in a white dashed box in Figure 1C. Differential metabolites of all time points are identified by parametric or nonparametric hypothesis tests ($p < 0.05$) between the pre-dose and every post-dose metabolomics data respectively (function 'GetDiffData') and are listed in weight rank order as calculated by the SAM (Significance analysis of microarrays) method. It is suggested to adjust the raw p values by the 'FDR' (or stricter 'Holm') method to eliminate possible false positive results induced by multiple tests. Subsequently, functions 'GetEndo', 'GetAbso' and 'GetSecdAbso' divide the differential metabolites into three groups by similarity analysis: (i) altered endogenous metabolites, (ii) absorbed drug compounds and (iii) secondary metabolites of the absorbed drug compounds. These findings are the crucial basis for the following analysis on the interrelationships between drug compounds and metabolites derived from drug as well as the metabolic impact of the drug to the mammalian metabolism.

2.3 Analysis and visualization of differential compounds

The resulting differential compounds or groups of compounds are analyzed in four independent ways (Fig. 1C). Seven conventional PK parameters (Cmax, Tmax, AUC, CL, Tlast, Tfirst and Cmin) and the concentration-time profiles of every user-specified compound are calculated and plotted by the function 'PKs' [Fig. 1D(a, b)]. This is an extension of the traditional PK analysis as both drug compounds and endogenous metabolites can be analyzed.

The scores and trajectory plots [Fig. 1D(c)] of principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) which are commonly used in metabolomics studies are generated by the function 'ScatPlot'. These figures are effective ways to show the classification and alteration trend of metabolic profiles across time-points of interest. Permutation plot [Fig. 1D(d)] and error rates of PLS-DA are also generated to check the model performance.

The relationships between any two groups of differential metabolites are evaluated by correlation coefficients (Pearson, Spearman or Kendall) and diverse forms of correlation diagrams [Fig. 1D(f–h)], which may provide additional insights to the interplay of multi-component drugs and the metabolic system of mammalian. Additionally, heatmap with Hierarchical Cluster Analysis (HCA) [Fig. 1D(e)] is also a good way to visualize the similarity of groups or compounds by detailed data (e.g. all or any group of differential compounds).

Gender, daily routine and diet may impact the performance and efficacy of drugs. So, data analysis is conducted in all, male and female samples, separately. Times of meal and sleep are marked in the figures if corresponding information is given in the basic information file.

Fig. 1. Workflow of polyPK. (**A**) Data importation. Four data files, three required (solid boxes) and one optional (black dashed box), are imported for analysis. (**B**) Data pretreatment. (**C**) Poly-PK analyses. Differential metabolites are identified and are classified into three groups (white dashed box). The resulting compounds or groups of compounds are analyzed in four independent ways. (**D**) Results and visualization. Some figures (.pdf) and tables (.xlsx) are generated during step C. (**a**) PK parameters of user-specified compounds (e.g. absorbed drug compounds). (**b**) Concentration-time profiles of user-specified compounds. (**c**) Trajectory of user-specified groups of compounds (e.g. altered endogenous metabolites). (**d**) R2-Q2 scatter plot of the permutation test. (**e**) Heatmap and cluster of user-specified groups of compounds. (**f**) Pairwise correlation coefficients of user-specified compounds (e.g. altered endogenous metabolites and absorbed drug compounds). Bubble (**g**) and chord (**h**) diagrams of correlation coefficients. The functions of each step are in white italic. All the intermediate and final results are saved separately in corresponding files and folders

3 Conclusion

An R package, polyPK, was developed to automate the complete data analysis and resulting visualization step of Poly-PK pipeline. It will serve as an easy and effective tool for PK study of multi-component drugs by using a systematic omics approach. More descriptions are provided in the Supplementary Material.
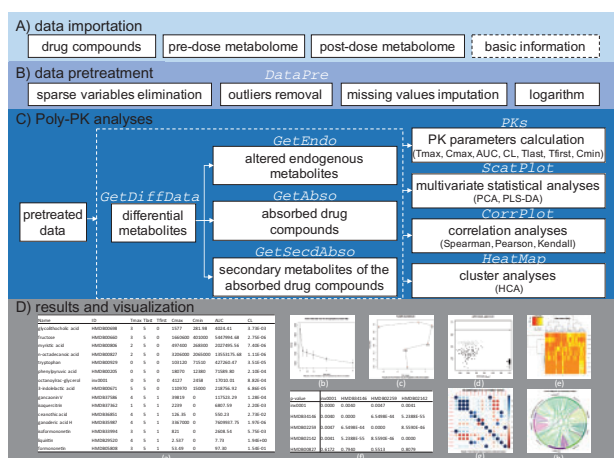
Funding

References

Jia,W. *et al*. (2015) The polypharmacokinetics of herbal medicines. *Science*, **350**, S76–S79.

Lan,K. *et al*. (2010) An integrated metabolomics and pharmacokinetics strategy for multi-component drugs evaluation. *Curr. Drug Metab.*, **11**, 105–114.

Lan,K. *et al*. (2013) Towards polypharmacokinetics: pharmacokinetics of multicomponent drugs and herbal medicines using a metabolomics approach. *Evid. Based Complement. Alternat. Med.*, **2013**, 819147.

Xie,G. *et al*. (2017) Poly-pharmacokinetic study of a multicomponent herbal medicine in healthy Chinese volunteers. *Clin. Pharmacol. Ther.*, doi: 10.1002/cpt.784

Xie,G. *et al*. (2012) Metabolic fate of tea polyphenols in humans. *J. Proteome Res.*, **11**, 3449–3457.