



Short communication

Medcheck: a novel software for automated de-formulation of traditional Chinese medicine (TCM) prescriptions by liquid chromatography-mass spectrometry



Xiao-lan Li ^{a, b, 1}, Jian-qing Zhang ^{a, 1}, Yun Li ^{a, 1}, Xuan-jing Shen ^{a, b}, Huan-ya Yang ^a, Lin Yang ^a, Meng Xu ^a, Qi-rui Bi ^a, Chang-liang Yao ^a, De-an Guo ^{a, b, *}

^a Shanghai Research Center for Modernization of Traditional Chinese Medicine, National Engineering Research Center of TCM Standardization Technology, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China

^b University of Chinese Academy of Sciences, Beijing, 100049, China

ARTICLE INFO

Article history:

Received 21 September 2023

Received in revised form

6 January 2024

Accepted 25 February 2024

Available online 12 March 2024

Prescriptions are the main clinical application of traditional Chinese medicines (TCMs). Common forms include Chinese patent medicines, Kampo formulas, and hospital decoctions. A new prescription called “famous classical formulas” is recently developed and expected to boom in the market. Identifying constituent medicinal plants in prescriptions is critical for new drug development and quality control [1], which could avoid safety issues from adulteration or substandard ingredients, as seen in the notorious Longdan Xiegan Pill event. However, formula clarification of consumable plant products, especially TCMs, has been challenging, relying on manual screening of metabolites and subjective evaluation [2]. Computers rarely produce random errors, and data processing often takes just minutes, which is more conducive to handling large sample batches. Most attractively, less prior knowledge from the researcher is required [3]. Reported here for the first time is the accomplishment of two aims: 1) Developing a universal workflow applicable to all herbal prescriptions; 2) Creating a user-friendly, intuitive software for automated identification of medicinal plants in TCM prescriptions, termed Medicine Check (Medcheck), freely available at [https://github.com/](https://github.com/TCMP2023/Medcheck_v.1.git)

[TCMP2023/Medcheck_v.1.git](https://github.com/TCMP2023/Medcheck_v.1.git). The application allows users to simply import the liquid chromatography-dynamic multiple reaction monitoring (LC-dMRM) data of an unknown prescription into the platform, which then reveals medicinal plant identification with a matching score.

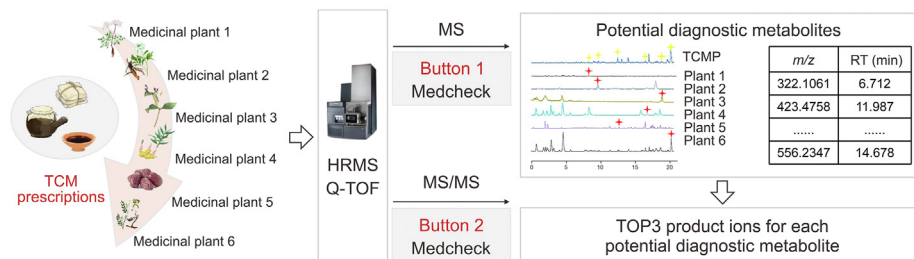
The universal workflow involves four key steps (Fig. 1). Step 1: screening potential diagnostic metabolites based on high resolution mass spectrometry (HRMS) data. LC-HRMS signals characteristically present in single plant and prescription, but absent in other plants, were considered as potential diagnostic metabolites. Step 2: transferring potential diagnostic metabolites from HRMS to triple-quadrupole mass spectrometry (TQMS). The metabolites were transferred to TQMS for pursuing a higher selectivity and economic efficiency. Predicted retention time was calculated in Medcheck based on mixed solution of 19 reference standards [4]. The retention time calibration model was assessed by another 9 standard compounds with R^2 of 0.9999 (Table S1 and Fig. S1). Taking both predicted retention time and ion response intensity into consideration, accurate retention time was determined. This outperformed traditional manual approaches that typically prefer peaks with larger peak area or closer retention time to HRMS, as evidenced by 257 > 137 (liquiritigenin) ion pair in Fig. S2A. Step 3: verifying diagnostic metabolites and method validation. The metabolites were further verified in TCM prescriptions, medicinal plants, and negative controls using TQMS. To enable rapid verification, an automated approach was utilized to determine the presence or absence of chromatographic peaks rather than traditional signal-to-noise (S/N) threshold. We analyzed 526 ion pair peaks in Danggui Jianzhong Decoction (DGJZD) and negatives. Peak presence was determined to be height >60, width >0.2 min, and area >300 based on maximal accuracy (Table S2). Multi-batch sample verification required metabolites to be detectable in at least 3 batches with peak area relative standard deviations (RSDs) <30%. To simplify analysis and eliminate interference, the top 30 ions were chosen as diagnostic metabolites. Ultimately, methodological validation was performed, including assessments of

* Corresponding author. Shanghai Research Center for Modernization of Traditional Chinese Medicine, National Engineering Research Center of TCM Standardization Technology, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China

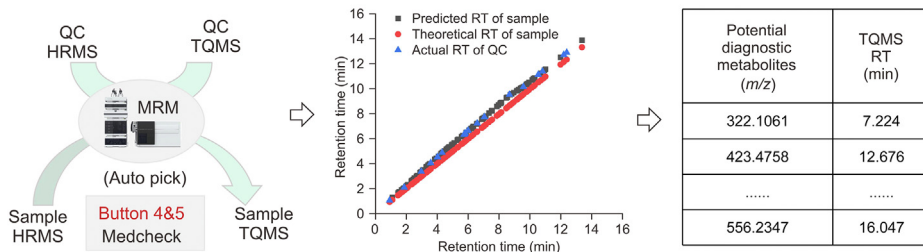
E-mail address: daguo@simm.ac.cn (D.-a. Guo).

¹ These authors contributed equally to this work.

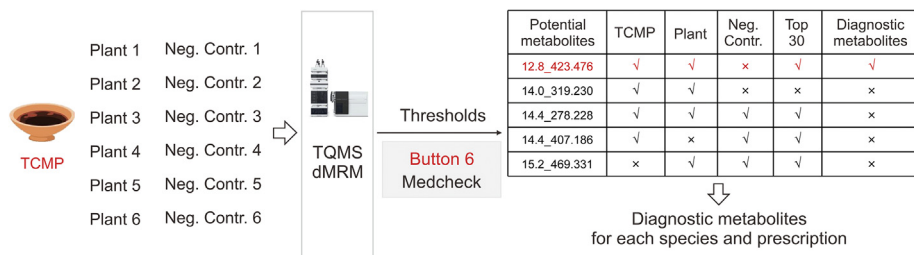
Step 1: Screening potential diagnostic metabolites in HRMS



Step 2: Transferring potential diagnostic metabolites to TQMS



Step 3: Verifying diagnostic metabolites and method validation



Step 4: Database construction and plant authentication with scores

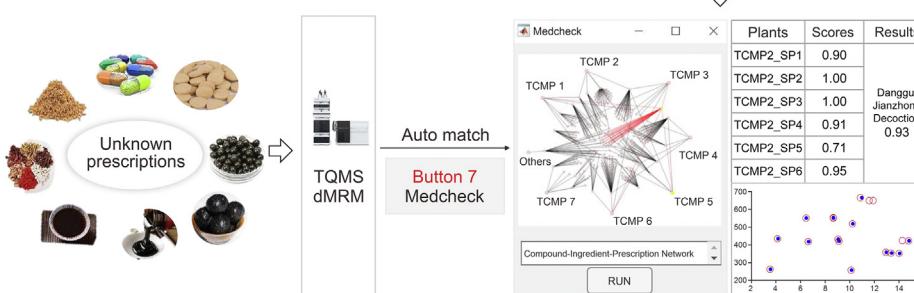


Fig. 1. Overview of the data analysis workflow for the Medcheck software used to de-formulate TCM prescriptions. Step 1: TCM prescriptions and medicinal plants were analyzed by HRMS to screen common signals as potential diagnostic metabolites. MS¹ was obtained automatically via Button 1, MS² via Button 2. Step 2: Potential metabolites were transferred to TQMS with automatic retention time calibration based on 19 reference standards. Predicted retention times were obtained by Button 4, actual retention time by Button 5. Step 3: Diagnostic metabolites were obtained after verification with negatives on TQMS by Button 6. Step 4: Diagnostic metabolites populated the Medcheck database. LC-dMRM data were directly imported into Medcheck for de-formulation by matching scores against the database using Button 7. HRMS: high resolution mass spectrometry; TQMS: triple-quadrupole mass spectrometry; QC: quality control sample; Auto. Pick: automatic pick peaks; TCMP: traditional Chinese medicine prescriptions; Neg. Contr.: negative Control sample; dMRM: dynamic multiple reaction monitoring.

linearity, repeatability, and stability (Table S3 and Fig. S3) Table S3. Step 4: database construction and plant authentication with scores. Users can automatically construct an in-house database using Medcheck following Step 1–Step 3. LC-dMRM data can then be directly imported into Medcheck to automatically match and score samples. Determining an optimal scoring threshold is challenging [5]. We analyzed 1,100 matching results from cross-combining 25 plant species across 7 prescriptions and 41 negative controls. An accuracy of 99% was achieved when thresholds were between 0.46 and 0.60 (Fig. S2B). Therefore, the matching score threshold was set at 0.5.

The Medcheck software was developed in MATLAB App Designer with two tabs –“Process” and “Identification” Fig. S3 (Fig.

S4). “Process” is used to process raw data to obtain diagnostic metabolites and build a database containing prescriptions, medicinal plants, retention times, and MS¹/MS² ions. “Identification” applies database to identify unknown samples, with visualized results including identified prescriptions and ingredients, matched ion number, and matching scores. The user guide is available on the web page, https://github.com/TCMP2023/Medcheck_v.1.git.

The universal workflow and Medcheck software were validated by 150 batches of samples including 7 homemade prescriptions (35 batches), 27 plants, 41 negatives, 30 blinds, and 17 commercial products (Table S4). Seven prescriptions each with at least three batches, along with 27 plants and 41 negatives, were analyzed by LC-TQMS. The results were processed through “Medcheck” to

screen diagnostic metabolites, compiling them into an authentication database. Metabolites obtained for DGJZD, Banxia Xiexin Decoction (BXXXD), Fuzi Decoction (FZD), Qingweisan Decoction (QWSD), Wenpi Decoction (WPD), Jinshui Liuju Decoction (JSLJD), and Taohong Siwu Decoction (THSWD) **Annex 1** were 68, 89, 66, 44, 84, 49, 48 in positive and 80, 81, 46, 88, 55, 85, 90 in negative mode, respectively **Fig. S4A**(**Fig. S5A**). Ultimately, a local database was constructed containing all selected metabolites (448 positive, 525 negative) from 25 out of 27 medicinal plants (**Table S5**). Furthermore, each diagnostic metabolite was assigned a unique label composed of a sequential number and abbreviated source name for rapid access and evaluation. For example, the transition 381.2057 > 135.0431 at 15.123 min in DGJZD under positive mode was labeled as DGJZD_DG_61_2_POS, indicating that *m/z* 135.0432 was the second intensive fragment for the 61st diagnostic metabolite (*m/z* 381.2057) from *Angelicae Sinensis Radix*. After systematic labeling, a TCM prescriptions-single plants-diagnostic metabolites network **Fig. S4B**(**Fig. S5B**) was created to identify ingredients by matching metabolites to database.

The validated Medcheck software was tested on 30 blind samples. The identified results and corresponding scores were summarized into heat maps (**Fig. S4C**(**Fig. S5C**, details in **Table S5**(**Table S6**). Results showed all 25 plants-related samples were accurately identified. For example, sample 30 was identified as Wenpi Decoction (0.97), with all constituents detected: Rhei Radix Et Rhizoma (1.00), Radix Aconiti Lateralis Preparata (1.00), Licorice Root (0.97), Zingiberis Rhizoma (0.86), and Ginseng (1.00). However, inadequate ingredients were detected in some cases. This is because suitable diagnostic metabolites could not be screened for Pinelliae Rhizoma and Poria, which showed weak mass spectral response, nonspecific ions, and low yield.

Medcheck was applied to de-formula of 14 commercial and 3 homemade prescriptions, containing 44 plants, with 25 kinds present in database. All results agreed with product labeling, achieving 99.7% accuracy with 0.19% false positive **Table S6** (**Table S7**). Two prescriptions were identified fewer ingredients than expected, including Rehmanniae Radix in Anshen pills and Ginseng in Wenpi Decoction, which may suggest potential issues. For Wenpi Decoction, Medcheck identified no Ginseng with a 0.43 match score. Following further verification through high-performance liquid chromatography (HPLC) and thin-layer chromatography (TLC), it was determined that American Ginseng was present instead of Ginseng **Fig. S5** (**Fig. S6**).

To conclusion, Medcheck provides an automated and streamlined workflow for automated de-formulation through an intuitive MATLAB-based platform, with steps of potential metabolites screening on HRMS, metabolites transfer to TQMS, diagnostic metabolites verification, database construction, and sample authentication with matching scores. The workflow provided significant advantages as follows. Diagnostic metabolites were automatically screened by cross-validating formulas, plants, and negatives. A robust LC-dMRM method was developed to detect metabolites from any prescriptions. Automated retention time calibration and

peak confirmation was implemented for routine analysis. Raw data can be directly imported into Medcheck, with automatic presentation of identified prescriptions, ingredients, and scores. All procedures were validated by systematic experiments. 25 plants-related samples achieved 99.7% accuracy. Medcheck was applicable for Chinese patent medicines, and Kampo formulas, and had potential to identify other commercial products like decoction slices, and dispensing granules.

CRediT author statement

Xiao-lan Li: Data curation, Formal analysis, Writing - Original draft preparation; **Jian-qing Zhang**: Investigation, Validation, Writing - Reviewing and Editing; **Yun Li**: Data curation, Formal analysis; **Xuan-jing Shen**: Data curation, Visualization; **Lin Yang**: Data curation; **Huan-ya Yang**: Data curation, Visualization; **Meng Xu**: Data curation; **Qi-rui Bi**: Methodology; **Chang-liang Yao**: Methodology; **De-an Guo**: Conceptualization, Funding acquisition.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the Guangxi Science and Technology Major Program (Grant No.: GUIKEAA23023035), Key Program of National Natural Science Foundation of China (Grant No.: 82130111), Guangxi Science and Technology Major Project (Grant No.: GUIKEAA22096029), and Science and Technology Major Project of Inner Mongolia (Grant No.: 2021ZD0017).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpha.2024.02.012>.

References

- [1] X. Lu, Y. Jin, Y. Wang, Y. Chen, X. Fan, Multimodal integrated strategy for the discovery and identification of quality markers in traditional Chinese medicine, *Journal of Pharmaceutical Analysis* 12 (5) (2022) 701–710.
- [2] S. Beyramysoltan, M.I. Chambers, A.M. Osborne, M.I. Ventura, R.A. Musah, Introducing “DoPP”: a graphical user-friendly application for the rapid species identification of psychoactive plant materials and quantification of psychoactive small molecules using DART-MS data, *Anal. Chem.* 94 (48) (2022) 16570–16578.
- [3] D. Eilertz, M. Mitterer, J.M. Buescher, automRm: an R package for fully automatic LC-QQQ-MS data preprocessing powered by machine learning, *Anal. Chem.* 94 (16) (2022) 6163–6171.
- [4] F. Zheng, X. Zhao, Z. Zeng, L. Wang, W. Lv, Q. Wang, et al., Development of a plasma pseudotargeted metabolomics method based on ultra-high-performance liquid chromatography–mass spectrometry, *Nat. Protoc.* 15 (8) (2020) 2519–2537.
- [5] Q. Song, J. Li, Y. Cao, W. Liu, H. Huo, J.-B. Wan, et al., Binary code, a flexible tool for diagnostic metabolite sequencing of medicinal plants, *Anal. Chim. Acta* 1088 (2019) 89–98.