analytical chemistry

pubs.acs.org/ac Article

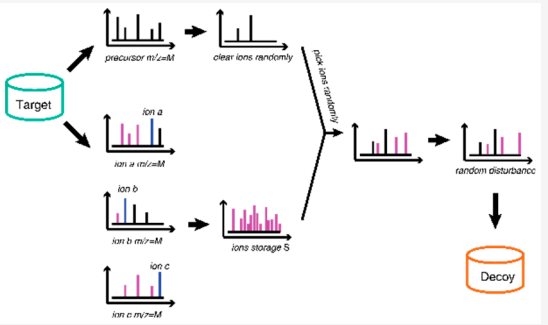
XY-Meta: A High-Efficiency Search Engine for Large-Scale Metabolome Annotation with Accurate FDR Estimation

Dehua Li,* Binghang Liu, Hancheng Zheng, Xu Xiao, Zhenyu Li, Enhui Luan, Wei Li, Yaling Yang, Yalan Wang, Qiaoyun Long, Jiaping Song, and Gong Zhang*

Cite This: *Anal. Chem.* 2020, 92, 5701−5707 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: FDR control has been a huge challenge for large-scale metabolome annotation. Although recent research indicated that the target−decoy strategy could be implemented to estimate FDR, it is hard to perform FDR control due to the difficulty of getting a reliable decoy database because of the complex fragmentation mechanism of metabolites and ubiquitous isomers. To tackle this problem, we developed a decoy generation method, which generates forged spectra from the reference target database by preserving the original reference signals to simulate the presence of isomers of metabolites. Benchmarks on GNPS data sets in Passatutto showed that the decoy database generated by our method is closer to the actual FDR than other methods, especially in the low FDR range (0−0.05). Large-scale metabolite annotation on 35 data sets showed that strict FDR reduced the number of annotated metabolites but increased the spectral efficiency, indicating the necessity of quality control. We recommended that the FDR threshold should be set to 0.01 in large-scale metabolite annotation. We implemented decoy generation, database search, and FDR control into a search engine called XY-Meta. It facilitates large-scale metabolome annotation applications.

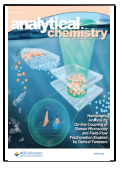


Metabolomics is mainly classified into two types: targeted metabolomics[1] and untargeted metabolomics.[2] The targeted metabolomics are to quantify and analyze the known metabolites of interest, while the untargeted metabolomics focus on identifying as much as possible metabolites of whole metabolomes. However, the lack of many metabolite standards leads to a large amount of false positives in untargeted metabolomics, which requires a way to estimate the FDR of metabolome annotation. However, lacking an accurate FDR estimation method is a challenge hindering the development of metabolomics.[3,4] Also, literature has reported that about 80% of the metabolites could not be annotated in many biological samples, which could have been served as a valuable pool for biomarker discovery.[3]

In the past decade, a number of analytical tools have been developed for metabolome annotation and quantitative analysis, such as XCMS,[5] MZmine,[6] MIDAS,[7] MAIT,[8] and CSI: FingerID.[9] Most of them used the database search method for metabolome annotation. Thus, obtaining standard mass spectra of metabolites is preferred because the structure and the fragmentation scheme of metabolites are too complex to be computationally predicted. Due to the lack of the standard mass spectrum, spectrum prediction tools of metabolites, like CFM-ID,[10] can be used to build a reference database since these tools integrated the complicated mass spectrometric decomposition processes by model training from an enormous amount of mass spectra of metabolites. However, the spectral matching often reports many misannotations mainly due to two reasons: random noise contamination and lack of reference spectra. This necessitates FDR control, which is one of the major challenges in untargeted metabolomics.[4,11]

Generally, large-scale metabolomics annotation uses the empirical Bayesian method to evaluate the FDR of the annotation results, which is the simplest and relatively robust strategy.[12] The empirical Bayes approximation of the Bayesian method are often fixed from historical samples, while in real situations the experimental samples are different from the historical ones in various aspects, for example, different sample types, preprocessing techniques, and biological backgrounds. Thus, the empirical Bayesian method is not an appropriate option to estimate FDR when samples have various and complex characteristics. Palmer et al. suggested to generate a decoy set by adding implausible adducts.[11] The decoys obtained by this method have fragment ions characteristics of metabolites that do not exist in the target database. Therefore, this method of generating a decoy database is a huge breakthrough in the application of a target−decoy strategy in molecular formula annotation. Subsequently,

Received: July 24, 2019
Accepted: March 26, 2020
Published: March 26, 2020



https://dx.doi.org/10.1021/acs.analchem.9b03355
Anal. Chem. 2020, 92, 5701−5707

Scheubert et al. created a tool, named Passatutto, to evaluate the performance of the generated decoy database. After comparing the performance of the decoy database generated by different methods, they found the ion fragmentation tree method seems to be the best way to build a decoy database.[12] Also, they proposed that the ion fragmentation tree method can be used as a standard to perform the target−decoy strategy in MS/MS-based metabolomics. Morever, Scheubert et al. tried four different target−decoy approaches for the MS/MS spectral match. Fragmentation tree was the most efficient one. Meanwhile, they suggested that the FDR threshold can be adjusted in different projects.[12] Wang et al. proposed a method, using the fake compounds violating the octet rule of chemistry, to modify the chemical formula of the target metabolites and then used MetFrag to generate an MS/MS pattern of the decoy database.[13] Finally, they demonstrated the feasibility of this decoy database generation method and developed a tool for metabolome annotation together with quality control, called JUMPm.

Here, we demonstrate a decoy-generation method based on the target database. The generated decoy database has similar characteristics to the target database, which can effectively simulate the scenario of random matches between spectra. We tested the performance of the decoy database. The results confirmed that the decoy database based on this method has reliable performance. Furthermore, Passatuto's data set was used to evaluate the FDR performance of XY-Meta. It turns out that its performance of FDR estimation is reliable whether using the concatenated target−decoy mode or the separated target−decoy mode. Finally, we developed an efficient tool for large-scale metabolome annotation with FDR evaluation, XY-Meta (https://github.com/DLI-ShenZhen/XY-Meta), which is easy to operate and efficient.

EXPERIMENTAL SECTION

Spectrum Matching Algorithm. The spectra matching algorithm integrated in XY-Meta is based on a vector dot product algorithm.[14] The score for a pair of spectra match is determined by the quality of the matched ion signals in the spectrum, which is written as

$$S_{\text{Match}} = \sum_{i=1}^{n} w_i \frac{\vec{a}_i \cdot \vec{b}_i}{d_i}$$

where $\vec{a}$ is the fragment ion signals of the query spectrum, $\vec{b}$ is the fragment ion signals of the reference spectrum, w is the weight of ion signals intensity, d is the difference between the projections of the two vectors on the x-axis, and n is the number of matched ion signals in the reference spectrum.

Research has shown that in spectra matching, the matching score of the spectra signal charge-to-mass ratio contributes much more to the spectra recognition than the matching score of the spectra signal intensity.[15] However, when increasing the weight of the spectra signal charge-to-mass ratio matching score, the weights of different intensity fragment ions should be also considered. To calculate the weights of ion signal intensities, all fragment ion signal intensity values in the reference database consist of a background signal distribution, used in the measurement of the weights with the Grubbs test as follows

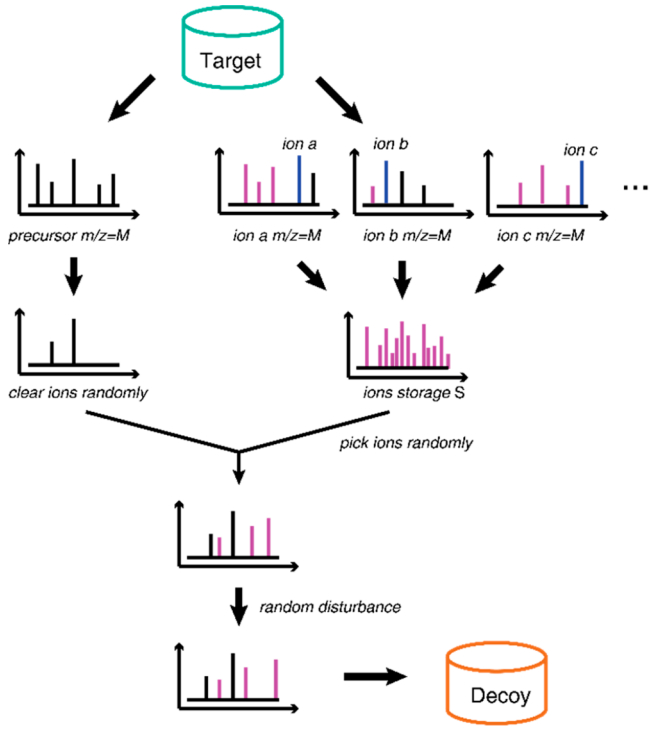


Figure 1. Workflow of decoy database generation.

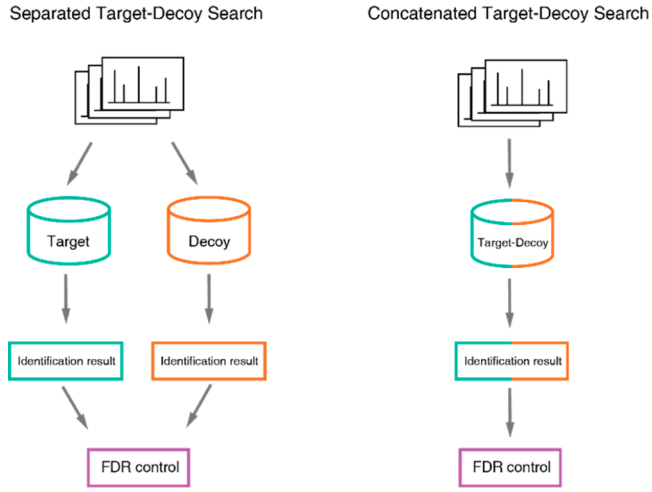


Figure 2. Schematic diagram of two different modes of target−decoy search strategy implementation: separated and concatenated.

Table 1. Parameter of Database Search

Parameter	Concatenated Search	Separated Search
Precursor ion mass tolerance	10−100 ppm	10−100 ppm
Fragment ion mass tolerance	0.005−0.05 Da	0.005−0.05 Da
Positive or negative mode	Positive mode	Positive mode

$$w_i = \frac{\left(\sum_{j=1}^{m} (h_j - \bar{\mu})^2/(n-1)\right)^{1/2}}{h_i - \bar{\mu}}$$

where h_j is the intensity value of fragment ion signal j in reference spectra, h_i is the ion signal i intensity value in a query spectrum, is $\bar{\mu}$ is the mean of the background distribution, and m is the total number of fragment ion signals in reference spectra.

https://dx.doi.org/10.1021/acs.analchem.9b03355
Anal. Chem. 2020, 92, 5701−5707

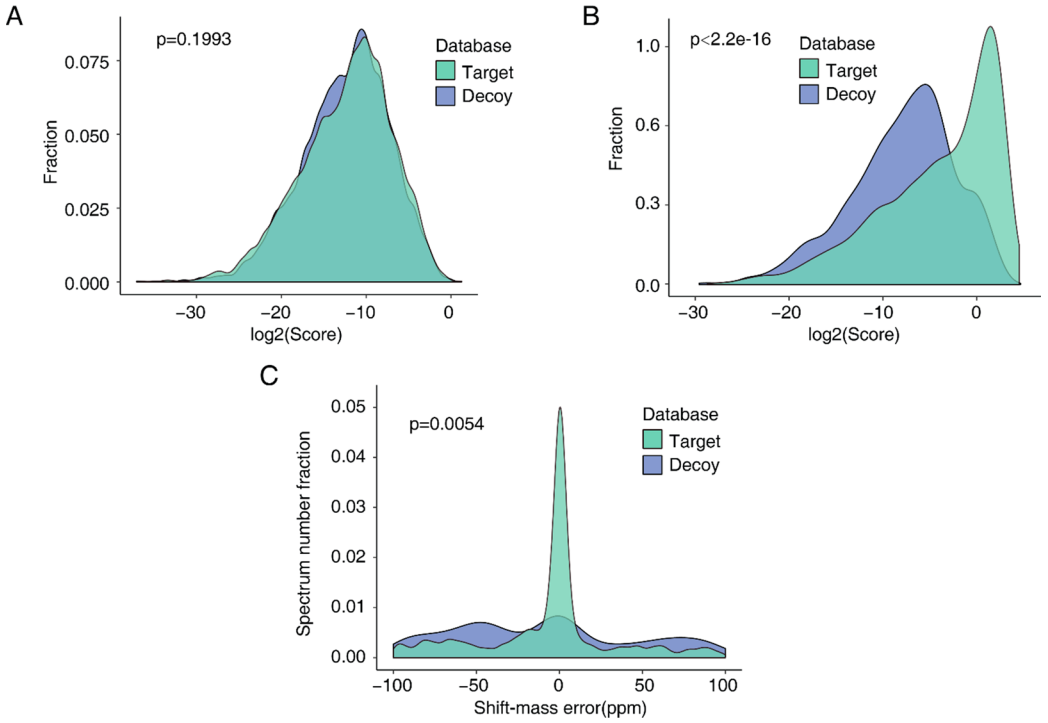


Figure 3. Evaluation of decoy database generated by XY-Meta. (A) Score distribution of the forged spectrum from Passatutto matches to the MoNA-Massbank spectrum library; corresponding decoy database was generated by XY-Meta. (B) Score distribution of the GNSP-NIST spectrum library matches to the MoNA-Massbank spectrum library; corresponding decoy database was generated by XY-Meta. (C) Mass shift error distribution of the query spectra matches to the target and decoy databases by using database search with 100 ppm precursor mass tolerance. All *p*-values are from KS-tests of the two distributions.

In addition, the signal-to-noise ratio also affects the accuracy of the annotation result, and the quality of the spectra signal matching can be reflected by two variables: theoretical signal matching rate (i.e., R_t in the equation below) and experimental signal matching rate (i.e., R_e in the following equation). The matching quality of the spectrum signal is measured as follow:

$$Q_{Match} = R_t^x R_e^y$$

where the x and y are exponential weights of the theoretical signal matching rate and the experimental signal matching rate, respectively. The theoretical signal matching rate is the number of matched signals divided by the total number of signals in the theoretical spectrum. The experimental signal matching rate is the number of matched signals divided by the total number of signals in the experimental spectrum.

The final score of the spectra match is the product of the spectra dot product score and the spectra matching quality

$$S_{Final} = S_{Match} \cdot Q_{Match}$$

Construction of a Decoy Database. A high-quality decoy database should have spectra with similar characteristics to those in the target database. This could not be achieved by fully random signals. One idea to overcome this problem is to randomly obtain the ion signal from the reference database to construct decoy spectra, which keeps the characteristics of various real compounds. The decoy database generated in this way can thus simulate the random matching of the true spectrum. However, the study by Scheubert et al. showed that only the noise-filtered reference database is proper for this method to generate decoy database.[12]

On the basis of the random selection method mentioned above, here we proposed an improved one by randomly selecting signals from the similar parent ions to the target spectrum (Figure 1). After the following two steps, we can generate a decoy database. First, in a target database, for a spectrum p with a precursor mass M, we build a signal warehouse S by merging all the fragment ion signals which are smaller than M from selected spectra which have more than one fragment where the mass ratio equals M. Second, we remove a certain proportion of the fragment ion signal in spectrum p and fill in some fragment ion signals which are randomly selected from the signal warehouse S, so that the number of the fragment ion signals in the new spectrum is equal to the original one. Finally, 30% of the fragment ion signals in the new spectrum is randomly selected to shift $\pm M/200{,}000$, thereby obtaining a final new decoy spectrum q. This decoy spectrum can be effectively simulated in XY-Meta without the help of other spectra generation tools and can be used for large-scale metabolome annotation.

Target−Decoy Strategy and FDR Estimation. The target−decoy strategy is the mainstream FDR control strategy for proteomics. There are two main types of database searches based on the target−decoy strategy: concatenated search and separated search.[16] The concatenated mode searches one reference database which consists of a target and decoy database given a query spectrum (Figure 2). Here, T is the number of annotation results which were hit by target spectra, and D is the number of annotation results which were hit by decoy spectra. The FDR calculation formula for the concatenated search mode is as follows

$$FDR = \frac{2 \times D}{T + D}$$

https://dx.doi.org/10.1021/acs.analchem.9b03355
Anal. Chem. 2020, 92, 5701−5707

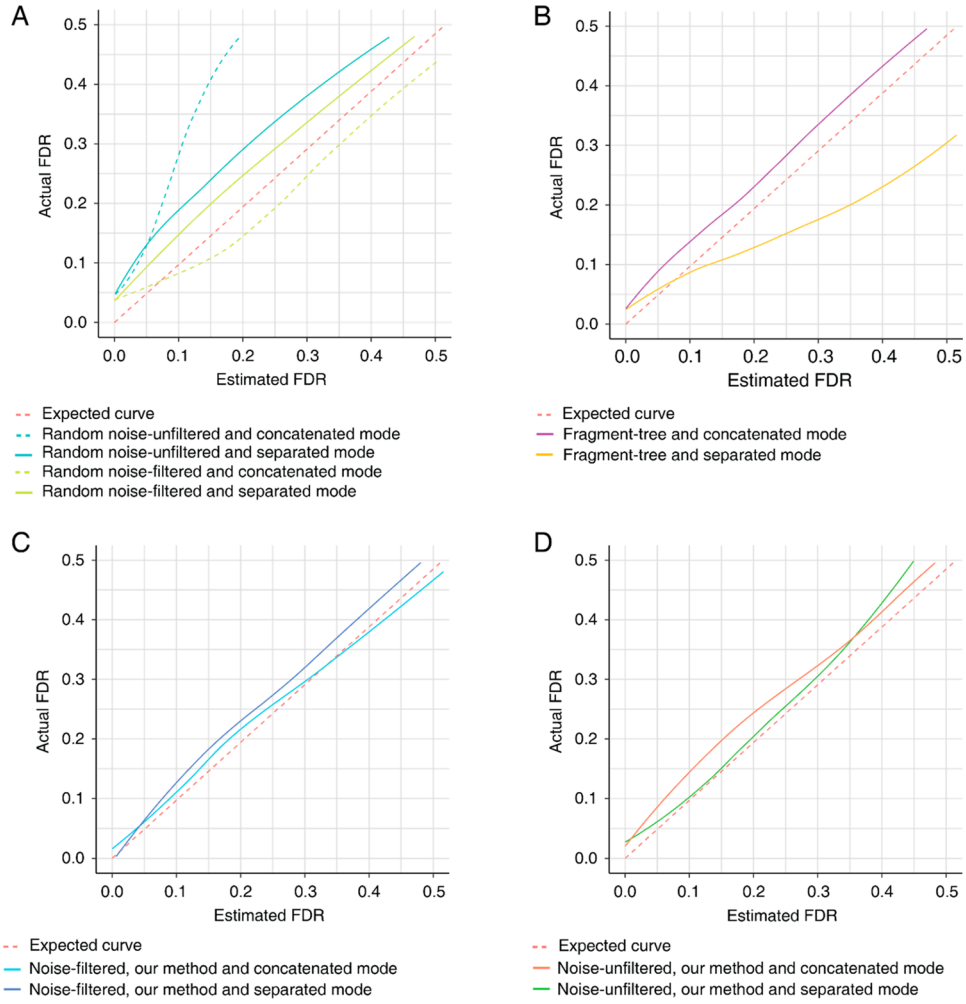


Figure 4. Comparison of the FDR evaluation performance of different decoy database. (A) FDR evaluation performance of random decoy database. (B) FDR evaluation performance of fragment-tree decoy database. (C) FDR evaluation performance of decoy database from our method when using noise-filtered reference. (D) FDR evaluation performance of decoy database from our method when using noise-unfiltered reference.

The separated search mode is to search the query spectra separately from the target database and decoy database (Figure 2). Only matching results from the target database and decoy database above the scoring threshold are used for FDR estimation which is measured as

$$\mathrm{FDR} = \frac{D}{T + D}$$

Database Search. We used XY-Meta for metabolome annotation. The search parameters involved mainly include ionization mode, precursor ion mass tolerance, fragment ion mass tolerance, and adduct type. The parameter settings refer to Table 1. The workflow of XY-Meta is shown in Figure S1.

RESULTS AND DISCUSSION

Evaluation on a Decoy Database. For a target−decoy strategy, the decoy database should have the similar probability of random matching to the target database. To test this point on our method, we took the entire Massbank mass spectrum library from MoNA (MoNA-Massbank) as the target database and used XY-Meta to generate the corresponding decoy database. To evaluate the decoy database, a randomized query spectrum library obtained by random signal generation in Passatutto was matched to the target and decoy databases of MoNA-Massbank, respectively. The matching result showed that both the number of matched spectra of the randomized query spectra and the distribution of matching scores were similar for the target and decoy databases (Figure 3A, p = 0.1993, KS-test), indicating that the decoy spectra had similar characteristics to the spectra of target database. We then used the GNPS-NIST spectrum library as the query and matched them to the target database and decoy database of MoNA-Massbank, respectively. The score distributions vary remarkably: the score matching to target is higher than that to decoy (Figure 3B, $p < 2.2 \times 10^{-16}$, KS-test), demonstrating the effectiveness of our strategy. Furthermore, we extended the tolerance of the parent ion to 100 ppm. The mass shift between the query and target spectra was largely concentrated in ±20 ppm, similar to the resolution of the high-precision mass spectrometer. In contrast, the mass error between the query and decoy databases is broadly distributed, which perfectly reflected its randomness (Figure 3C, p = 0.0054, KS-test). This result was consistent with the conclusion in previous studies.[17] These results confirmed that our method generated effectively the decoy database and can be used in target−decoy strategy implementation.

Comparison of FDR Estimation Performance to Other Decoy Databases. To evaluate the performance of FDR

https://dx.doi.org/10.1021/acs.analchem.9b03355
Anal. Chem. 2020, 92, 5701−5707

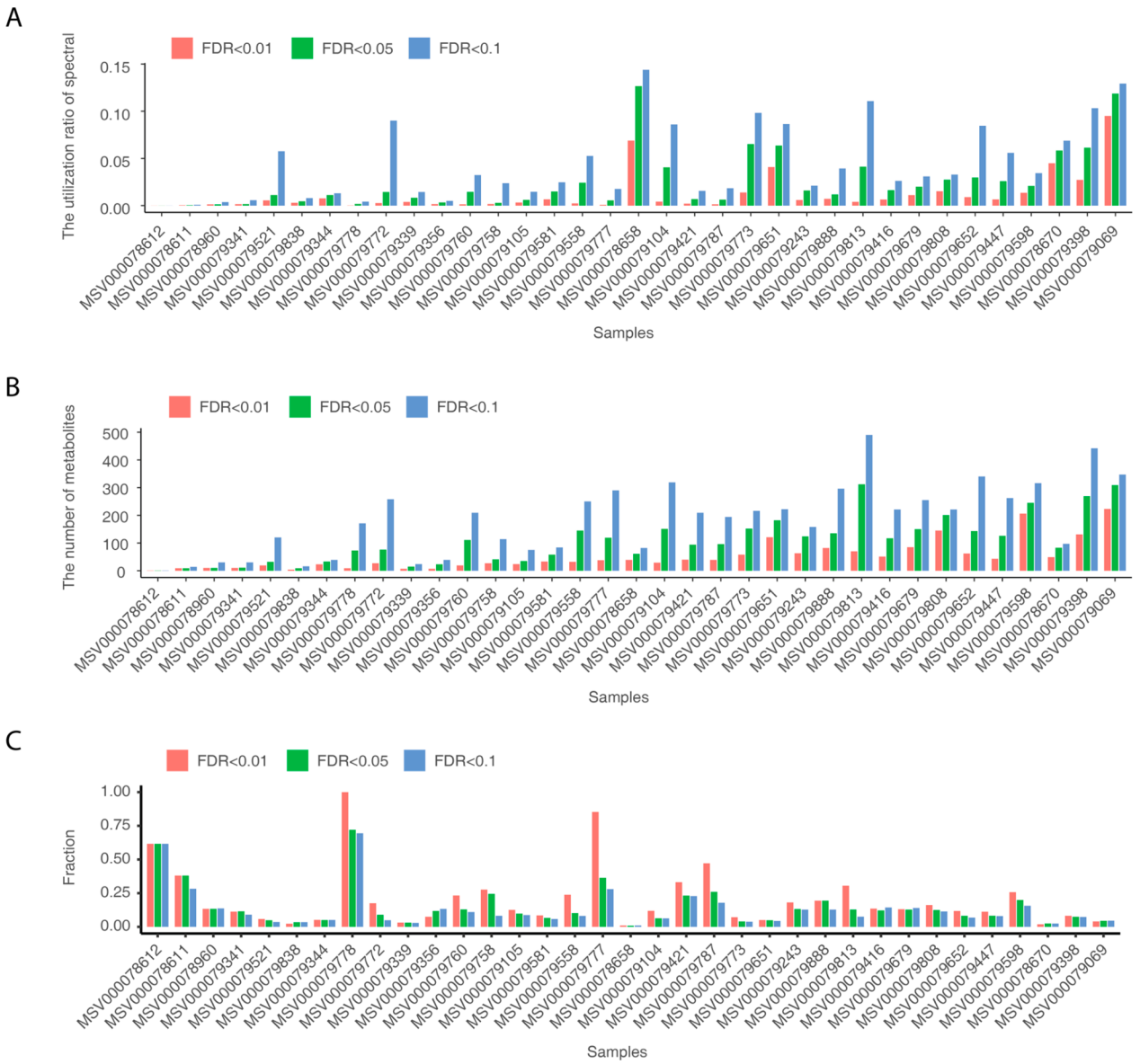


Figure 5. Results of metabolome annotation of 35 data sets in GNPS using XY-Meta. (A) Spectral utilization. (B) Number of metabolites identified. (C) Spectral efficiency, defined as the number of metabolites annotated per utilized spectral.

estimation on metabolite annotations, the predicted FDR from XY-Meta was compared to the actual FDR. The decoy database can be constructed with or without noise filtering, where the noise unfiltered decoy database is larger and thus provides a higher probability of random matches. Scheubert et al. showed that the fragment-tree method is better than random signal methods using traditional algorithms.[12] Therefore, in our comparison, we included three types of decoy databases in Passatutto including the random noise-unfiltered decoy database, random noise-filtered decoy database, and fragment-tree decoy database with the decoys generated by XY-Meta. Since concatenated and separated search modes were both used in the literatures, we evaluated these decoy databases in both modes. In the performance comparison, we used the structural accuracy of the recognized metabolites as the criterion for true positive or false positive to deduce the actual FDR. If the predicted structure of a metabolite from its spectrum is consistent with its labeled structure, this is a true positive prediction and vice versa.[18] The comparison showed that the accuracy of XY-Meta in using a different decoy database and different target−decoy search mode was better than all other decoy generation strategies considering all aspects. The AUC of the ROC curve exceeded 0.93 (Figure S2A). In both the concentrated and separated modes, FDR estimated by the random noise-filtered decoy database is closer to the actual FDR than other strategies. In addition, the random noise-filtered decoy database showed better FDR accuracy in the concatenated mode than the separated mode (Figure 4A). Furthermore, in the concatenated mode, the FDR evaluated using the fragment-tree decoy database is generally superior to the random noise-filtered decoy database. Although the FDR assessed using the fragment-tree decoy database is too strict in the separated mode in the high FDR range (0.1−0.5), this method estimates FDR more accurately in the range of 0.05−0.1 than the concatenated mode (Figure 4B). To be noted, our decoy generation method gives the best FDR estimation in the range of 0−0.05, where all other methods fail to estimate false discoveries even when setting FDR = 0.01

https://dx.doi.org/10.1021/acs.analchem.9b03355
Anal. Chem. 2020, 92, 5701−5707

(Figure 4). When using the noise-filtered database, the decoy from our method database still performs well; no matter which search mode is used, its performance is better than the fragment-tree decoy database. Although the decoy database from our method evaluates FDR more accurately, this also leads to a decrease in the AUC of the P−R curve (Figure S2B).

FDR Control Applied to Real-World Metabolome Annotation. We used the GNSP-NIST spectral database as a reference to perform metabolite annotation and FDR quality control on 35 data sets in the GNPS library.[19] Three FDR thresholds (0.01, 0.05, and 0.1) were used. The spectral utilization of most data sets was below 10%. When the FDR thresholds were set lower, the spectra utilization also decreases remarkably (Figure 5A). The median value of the number of metabolites identified under the FDR thresholds of 0.01, 0.05, and 0.1 were 112, 155, 174, respectively (Figure 5B). We also identified metabolites with using the library search tool on GNPS. The precursor mass tolerance and fragment ion mass tolerance were set identical as XY-Meta. Most of the identification results from XY-Meta were contained in the GNPS identification results. When the FDR threshold was set to 0.1, 0.05, and 0.01, this fraction of coincidence was 0.53, 0.63, and 0.76, respectively (Figure S3). Additionally, in most data sets, reducing the FDR threshold led to a dramatic decrease (even more than 10 times in some cases) of utilized spectra, indicating that most of the raw annotations are false positives. However, the spectral efficiency, i.e., the number of metabolites annotated per utilized spectra, increases in most cases when using strict FDR, indicating a better quality and confidence (Figure 6C). This emphasized the necessity of quality control using strict FDR in metabolomics like in proteomics.

CONCLUSION

We proposed a decoy-generation method, which constructs a decoy spectrum by combining retained original signals of its target spectrum and randomly selected ion signals from the whole reference database without knowing the chemical formula of a metabolite or using a spectroscopic prediction tool. This method was implemented in the tool XY-Meta. In addition, XY-Meta can implement a target−decoy strategy not only by using the above-mentioned decoy database but also by utilizing an external decoy database from other methods. We exhibited its accuracy and robustness using standard and real-world data sets, demonstrating an FDR control method for metabolomics.

Due to the complexity of the sample and MS noise, random matches will result in false positives. Therefore, FDR quality control of the annotation results is critical to the application of large-scale metabolomics. Researchers applied the target−decoy strategy for FDR estimation in the metabolome annotation.[11−13] In the MS-based proteomics, FDR control has been introduced more than a decade ago to ensure the confidence of annotations and is explicitly required in the guidelines[20] as the community standard (peptide FDR < 0.01, protein FDR < 0.01). This largely facilitated the boom of proteomics. However, the previous decoy generation methods in metabolomics could not effectively control the real FDR in the range of 0−0.05. Our decoy generation method solved this problem and thus can provide confident metabolome annotations like proteomics. Therefore, we believe that our XY-Meta tool will bring shotgun metabolomics into a more confident era and will facilitate its applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.9b03355.

Workflow of XY-Meta, ROC curve, and P−R curve of different decoys generated by different methods and matching ratios of metabolite identification results from XY-Meta in the identification results from GNPS of 17 data sets (PDF)

AUTHOR INFORMATION

Corresponding Authors

Gong Zhang − *Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes and MOE Key Laboratory of Tumor Molecular Biology, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, Guangzhou 510632, China;* orcid.org/0000-0003-0418-3433; Email: zhanggong-uni@qq.com

Dehua Li − *Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes and MOE Key Laboratory of Tumor Molecular Biology, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, Guangzhou 510632, China; Shenzhen Digital Life Institute, Shenzhen 518110, China;* orcid.org/0000-0003-0289-6798; Email: dehualiay@qq.com

Authors

Binghang Liu − *Shenzhen Digital Life Institute, Shenzhen 518110, China*

Hancheng Zheng − *Shenzhen Digital Life Institute, Shenzhen 518110, China; iCarbonX (Shenzhen) Co., Ltd., Shenzhen 518053, China*

Xu Xiao − *Shenzhen Digital Life Institute, Shenzhen 518110, China*

Zhenyu Li − *Shenzhen Digital Life Institute, Shenzhen 518110, China;* orcid.org/0000-0002-6455-2117

Enhui Luan − *Shenzhen Digital Life Institute, Shenzhen 518110, China*

Wei Li − *Shenzhen Webio Co., Ltd., Shenzhen 518110, China*

Yaling Yang − *Shenzhen Digital Life Institute, Shenzhen 518110, China*

Yalan Wang − *Shenzhen Webio Co., Ltd., Shenzhen 518110, China*

Qiaoyun Long − *Shenzhen Webio Co., Ltd., Shenzhen 518110, China;* orcid.org/0000-0003-3559-8594

Jiaping Song − *Shenzhen Digital Life Institute, Shenzhen 518110, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.9b03355

Author Contributions

D.L. conceived the project, supervised the whole study, analyzed the data, and wrote the manuscript. G.Z. supervised the whole study and wrote the manuscript. B.L., H.Z., J.S., Z.L., and X.X. analyzed the data and wrote the manuscript. E.L., W.L., Y.W., Y.Y., and Q.L. analyzed the data.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program (2018YFC0910200/

2017YFA0505001) and the Guangdong Key R&D Program (2019B020226001). We thank the Webio Lab members and the Digital Life Institute colleagues for the discussion and suggestions. D.L. wants to thank his wife Meiyun Zhang for her understanding and support for his research work.

REFERENCES

(1) Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B. *Curr. Protoc Mol. Biol.* **2012**, *98*, 30.2.1−30.2.24.
(2) Vinayavekhin, N.; Saghatelian, A. *Curr. Protoc Mol. Biol.* **2010**, *1*, 1−24.
(3) Blazenovic, I.; Kind, T.; Sa, M. R.; Ji, J.; Vaniya, A.; Wancewicz, B.; Roberts, B. S.; Torbasinovic, H.; Lee, T.; Mehta, S. S.; Showalter, M. R.; Song, H.; Kwok, J.; Jahn, D.; Kim, J.; Fiehn, O. *Anal. Chem.* **2019**, *91* (3), 2155−2162.
(4) Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C. E. *Curr. Opin. Biotechnol.* **2019**, *55*, 44−50.
(5) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78* (3), 779−87.
(6) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.
(7) Wang, Y.; Kora, G.; Bowen, B. P.; Pan, C. *Anal. Chem.* **2014**, *86* (19), 9496−503.
(8) Fernandez-Albert, F.; Llorach, R.; Andres-Lacueva, C.; Perera, A. *Bioinformatics* **2014**, *30* (13), 1937−9.
(9) Duhrkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Bocker, S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12580−5.
(10) Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D. S. *Metabolites* **2019**, *9* (4), 72.
(11) Palmer, A.; Phapale, P.; Chernyavsky, I.; Lavigne, R.; Fay, D.; Tarasov, A.; Kovalev, V.; Fuchser, J.; Nikolenko, S.; Pineau, C.; Becker, M.; Alexandrov, T. *Nat. Methods* **2017**, *14* (1), 57−60.
(12) Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L. F.; Duhrkop, K.; Bandeira, N.; Dorrestein, P. C.; Bocker, S. *Nat. Commun.* **2017**, *8* (1), 1494.
(13) Wang, X.; Jones, D. R.; Shaw, T. I.; Cho, J. H.; Wang, Y.; Tan, H.; Xie, B.; Zhou, S.; Li, Y.; Peng, J. *J. Proteome Res.* **2018**, *17* (7), 2328−2334.
(14) Huan, T.; Tang, C.; Li, R.; Shi, Y.; Lin, G.; Li, L. *Anal. Chem.* **2015**, *87* (20), 10619−26.
(15) Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O. *Mass Spectrom. Rev.* **2018**, *37* (4), 513−532.
(16) Elias, J. E.; Gygi, S. P. *Nat. Methods* **2007**, *4* (3), 207−14.
(17) Everley, P. A.; Bakalarski, C. E.; Elias, J. E.; Waghorne, C. G.; Beausoleil, S. A.; Gerber, S. A.; Faherty, B. K.; Zetter, B. R.; Gygi, S. P. *J. Proteome Res.* **2006**, *5* (5), 1224−31.
(18) Rathahao-Paris, E.; Alves, S.; Junot, C.; Tabet, J.-C. *Metabolomics* **2016**, *12* (1), na DOI: 10.1007/s11306-015-0882-8.
(19) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W. T.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C. C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C. C.; Yang, Y. L.; Humpf, H. U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya, C. A. P.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.; Alexandrov, T.; Litaudon, M.; Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. *Nat. Biotechnol.* **2016**, *34* (8), 828−837.
(20) Deutsch, E. W.; Lane, L.; Overall, C. M.; Bandeira, N.; Baker, M. S.; Pineau, C.; Moritz, R. L.; Corrales, F.; Orchard, S.; Van Eyk, J. E.; Paik, Y. K.; Weintraub, S. T.; Vandenbrouck, Y.; Omenn, G. S. *J. Proteome Res.* **2019**, *18*, 4108.

https://dx.doi.org/10.1021/acs.analchem.9b03355
Anal. Chem. 2020, 92, 5701−5707