

AAGMarker 1.0: a resource of serological autoantigen biomarkers for clinical diagnosis and prognosis of various human diseases

Jianbo Pan¹, Sheng Liu¹, Heng Zhu² and Jiang Qian^{1,3,*}

¹Department of Ophthalmology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA, ²Department of Pharmacology and Molecular Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA and ³The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

Received July 10, 2017; Revised August 08, 2017; Editorial Decision August 21, 2017; Accepted August 29, 2017

ABSTRACT

Autoantibodies are produced to target an individual's own antigens (e.g. proteins). They can trigger autoimmune responses and inflammation, and thus, cause many types of diseases. Many high-throughput autoantibody profiling projects have been reported for unbiased identification of serological autoantigen-based biomarkers. However, a lack of centralized data portal for these published assays has been a major obstacle to further data mining and cross-evaluate the quality of these datasets generated from different diseases. Here, we introduce a user-friendly database, AAGMarker 1.0, which collects many published raw datasets obtained from serum profiling assays on the proteome microarrays, and provides a toolbox for mining these data. The current version of AAGMarker 1.0 contains 854 serum samples, involving 136 092 proteins. A total of 7803 (4470 non-redundant) candidate autoantigen biomarkers were identified and collected for 12 diseases, such as Alzheimer's disease, Bechet's disease and Parkinson's disease. Seven statistical parameters are introduced to quantitatively assess these biomarkers. Users can retrieve, analyse and compare the datasets through basic search, advanced search and browse. These biomarkers are also downloadable by disease terms. The AAGMarker 1.0 is now freely accessible at <http://bioinfo.wilmer.jhu.edu/AAGMarker/>. We believe this database will be a valuable resource for the community of both biomedical and clinical research.

INTRODUCTION

Biomarker, a portmanteau of 'biological marker', is defined as 'a characteristic that is objectively measured and evaluated as an indicator of normal biological processes,

or pharmacologic responses to a therapeutic intervention' (1). Until now, biomarkers of various flavors, such as proteins (e.g. enzyme, receptor, antigen or antibody) and nucleic acids (e.g. mRNAs, DNA mutations, or other non-coding RNAs), have been identified and routinely used for clinical diagnoses of different diseases, including cancers (2). Among them, autoantigens (self-antigens or individual's own antigens) are a major source of production of autoantibodies, which play a pivotal role in triggering autoimmune responses and inflammation. Autoantibodies could become the actual pathogenic agents of the autoimmune diseases or the secondary consequence of tissue damage, and thus markers of disease activity and severity (3,4). Hypothesis that ongoing diseases could result in specific perturbations of autoantibody profiles was also proven in some neurodegenerative disorders, such as Alzheimer's disease (AD) and Parkinson's disease (PD) (5–7). Furthermore, autoantibodies have been proposed to maintain homeostasis through auto-clearance of aged cells and dysfunctional dividing cells of healthy persons and patients with cancers (8,9). More and more studies on autoantigen or autoantibody have been reported recently. And some databases, such as AAGAtlas 1.0 (<http://biokb.ncpsb.org/aagatlas/>), HPTaa (<http://www.hptaa.org>) and CTdatabase (<http://www.cta.lncc.br/>), provide information of 1126, 3518 and 276 literature reported or predicted autoantigens, respectively (9–11). Moreover, autoantibodies of the IgG isotype are reported to be abundant and ubiquitous in human sera, and their serum diversity could be affected by multiple factors, such as age, gender and disease status (12). Especially, serum autoantigens, which could directly react with specific autoantibodies, have been investigated as non-invasive biomarkers for diagnosis and therapeutic intervention of autoimmune diseases, cancers and other diseases (5–7,13–20). For example, a FDA-approved autoantigen biomarker panel, comprised of Annexin I, CAGE, GBU4–5, NY-ESO-1, p53 and SOX2, was developed for early diagnosis of lung cancer with sensitivity of 46% and specificity of 83% (20).

*To whom correspondence should be addressed. Tel: +1 443 287 3882; Fax: +1 410 502 5382; Email: jiang.qian@jhmi.edu

Traditionally, autoantigen biomarkers could be discovered by enzyme-linked immunosorbent assays (ELISA), which permits the measurement of one protein at a time (21). Thanks to the rapid development of high-throughput technologies (i.e. proteome microarrays), unbiased, large-scale autoantibody profiling assays are now possible for the identification of serological autoantigen-based biomarkers (5–7,13–19). However, those raw datasets were deposited dispersedly in the public domain, as exemplified by GEO (<https://www.ncbi.nlm.nih.gov/geo/>), Array-Express (<https://www.ebi.ac.uk/arrayexpress/>), and PMD (<http://www.proteinmicroarray.cn/>) (22–24). Such a lack of centralized data portal for these published assays with quantitative evaluation of autoantigen biomarkers has been the major obstacle to cross-evaluate the quality of these works and prevented from further data mining.

Here, we introduce a user-friendly database, AAgMarker 1.0 (<http://bioinfo.wilmer.jhu.edu/AAgMarker/>), which collects many published raw datasets obtained from serum profiling assays on the proteome microarrays, and provides a toolbox for mining these data. Autoantigen-based serological biomarkers are defined as those human proteins that can distinguish patients from healthy controls with certain distinguish power. We used a series of statistical parameters to quantitatively evaluate the performance of these biomarkers. The current version of AAgMarker 1.0 deposits 7803 (4470 non-redundant) candidate biomarkers identified for 12 diseases, including acute myeloid leukemia, Alzheimer’s disease, basal-like breast cancer, Bechet’s disease, early-stage Parkinson’s disease, kidney transplantation, meningioma, myelodysplastic syndromes, Parkinson’s disease, primary biliary cirrhosis, type 1 diabetes mellitus and type 2 diabetes mellitus. The users can conveniently search, browse and download the list of autoantigen biomarkers and their related diseases from AAgMarker with a user-friendly interface.

THE DATA

The autoantibody profiling raw datasets (gpr files) and their protein annotation were mainly derived from the public microarray repositories NCBI GEO and EBI ArrayExpress (22,23). The datasets were selected using the following criteria. First, each assay contains at least 10 disease samples and 10 healthy controls to ensure the statistical significance of the derived biomarkers. Second, each assay screens relatively a large number of proteins (>5000). Third, the assays were performed on serum samples. As a result, a total of 12 high quality datasets were assembled in current version of database. Controls and proteins were removed from the datasets if they are not properly annotated.

QUANTITATIVE IDENTIFICATION OF AUTOANTIGEN BIOMARKERS

Each assay was analysed previously and published by the researchers who generated the data. Besides providing the biomarkers determined by the original studies, we also provide additional biomarker candidates as determined by our unified computational framework. Similarly, a user can adjust the criteria to identify customized candidates. In addition, seven commonly used statistical parameters were also

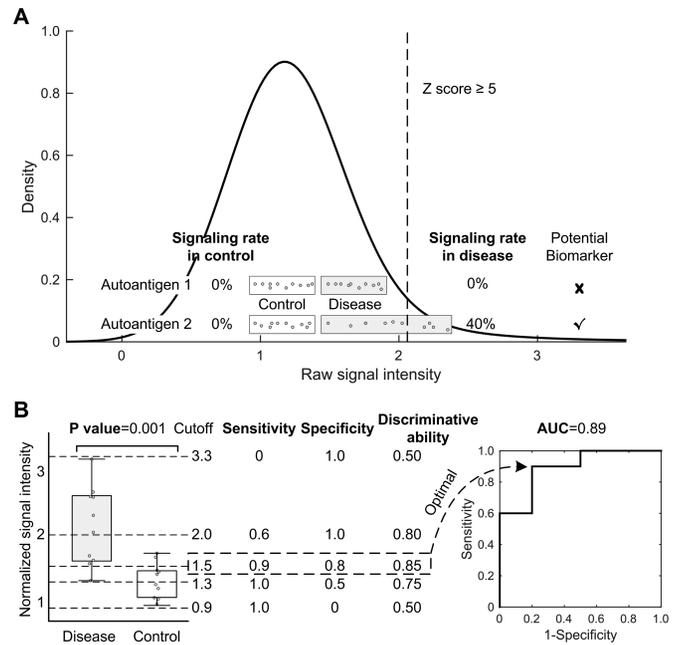


Figure 1. Quantitative identification of autoantigen biomarkers. (A) High signal intensities for easy detection. (B) Good performance to distinguish between disease and control.

generated for the candidate biomarkers, including signaling rate, *P* value, sensitivity, specificity, discriminative ability, AUC and optimal cutoff values (Figure 1). First, the median values of the foreground (F_{ij}) and background (B_{ij}) intensity at a given protein spot (i, j) on the protein arrays were extracted from the raw gpr files. The signal intensity (R_{ij}) of each protein spot was defined as F_{ij}/B_{ij} . Since each protein is often printed in duplicate on an array, R_{ij} was averaged for each protein as R_p . For each sample, we utilize the distribution of signal profile to calculate *Z*-score of each protein on protein arrays using the similar method described in our previous studies (25).

$$Z_p = \frac{R_p - \bar{N}}{SD} \tag{1}$$

where R_p is the intensity of protein on the microarray, \bar{N} and *SD* are mean value and standard deviation, respectively, of the noise distribution on the microarray. A stringent cutoff ($Z \geq 5$) was used to determine the positive hits in the assay.

Signaling rate

For a given protein, we calculated the number of samples in which the protein was scored positive in both the disease and control groups. The signaling (or positive) rate in both disease and control groups were then calculated accordingly (Figure 1A). A high signaling rate indicates that many samples show higher signal intensities for easy detection.

$$\begin{aligned} &\text{Signaling rate of disease} \\ &= \text{number of positive hits} / \text{number of samples in disease group} \end{aligned} \tag{2}$$

$$\begin{aligned} &\text{Signaling rate of control} \\ &= \text{number of positive hits} / \text{number of samples in control group} \end{aligned} \tag{3}$$

P value

To compare the protein signals across samples, quantile normalization among samples was performed. For each protein, normalized signal values were compared between disease group and control group by *t*-test or paired *t*-test. *P* value and FDR-adjusted *P* value was calculated for each protein.

Sensitivity, specificity & discriminative ability

For the normalized signals of all samples, given a signal criterion (i.e. '> intensity cutoff' or '< intensity cutoff'), when the protein meets the criterion, the sample will be scored positive. The sensitivity and specificity will be calculated for each protein using the following formula,

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (5)$$

$$\text{Discriminative ability} = (\text{Sensitivity} + \text{Specificity})/2 \quad (6)$$

where the true negatives (TN) are the healthy persons correctly classified as negatives, and the true positives (TP) refer to patients with disease correctly classified as positives. The false negatives (FN) are patients incorrectly classified as negatives, and the false positives (FP) are healthy persons incorrectly classified as positives. For a given protein, the value of sensitivity and specificity vary with different intensity cutoff values used for calling positives. We provide the sensitivity and specificity at an optimal cutoff value at which the highest discriminative ability is achieved (Figure 1B).

AUC

The receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold varies. The curve is created by plotting the true positive rate (sensitivity) against the false positive rate (1 – specificity) at various intensity cutoffs. The area under the ROC curve (AUC) was calculated for each protein.

The seven parameters could be applied alone or in combination to quantitative evaluation of autoantigen biomarkers. We used a combination of the criteria to determine the biomarker with good distinguish ability and high detection intensity (AUC \geq 0.6 and *P* value \leq 0.05 and Signaling rate of disease or control \geq 0.1). Users can obtain biomarker candidates using their own defined criteria (see Advanced Search below).

THE ACCESS OF THE DATABASE

The AAgMarker 1.0 is available online at <http://bioinfo.wilmer.jhu.edu/AAgMarker/>. The AAgMarker 1.0 was developed using MySQL server on Ubuntu Linux operating system. A user-friendly web interface designed using JSP

technology was provided to facilitate searching, downloading and updating. The web interface comprises five sections including 'HOME', 'SEARCH' (Figure 2A and B), 'BROWSE' (Figure 2C), 'DOWNLOAD' (Figure 2D) and 'HELP'. Bulk data are available in 'DOWNLOAD' page (Figure 2D), which also provides statistics of datasets. User manual and documentation are available on the 'HELP' page. Three main data retrieval methods were developed for the AAgMarker access. They are briefly described as following:

Basic search

Biomarkers defined by the default cutoff (i.e. AUC \geq 0.6, *P* value \leq 0.05 and Signaling rate of disease or control \geq 0.1) can be retrieved using basic search method (Figure 2A). A user can input names of a protein and disease, and those that meet the criteria of biomarkers will be searched and reported. When a protein name is not provided, all biomarkers associated with the disease will be shown. If multiple biomarkers are searched, the search page will show a table listing all autoantigen biomarkers ranked by their AUC values (Figure 2E). Each biomarker is also listed with its associated information, including disease type, the statistical parameters, and source, which indicates whether the biomarker was identified from AAgMarker, literature or both. Clicking on the protein name will then lead to the detailed information page (Figure 2F), where three sections of information are given: (i) 'Protein information', which provides basic information of the protein, such as description of its functions, chromosomal location and Gene Ontology, as well as external links to reference databases, such as UniProtKB (26), NCBI Gene (27), AAgAtlas (9) and neXtProt (28), when available; (ii) 'Dataset information', which provides information of datasets, sample number and clinical characteristics of patients and controls (e.g. age, sex and duration). (iii) 'Biomarker performance', which provides the signal distribution of all proteins for reference and signalling rate of the biomarker, AUC, *P* value, fold change of mean value between disease group and healthy group, optimal cutoff and corresponding sensitivity and specificity, the rank by AUC value in the biomarker list of the disease and the validation information in the literature when available, with the visualization of signal profile of the biomarker and ROC curve.

Advanced search

The advanced search method offers users with customized criteria (Figure 2B). One or multiple parameters with cutoffs ranging from 0 to 1 could be set for query.

Browse

The AAgMarker offers the browse method for direct retrieval of information from the database (Figure 2C). Users can browse autoantigen biomarker list by clicking initial of protein name or disease term. Like the quick search method, the browse method will automatically adopt the default query criteria for serum autoantigen biomarkers.

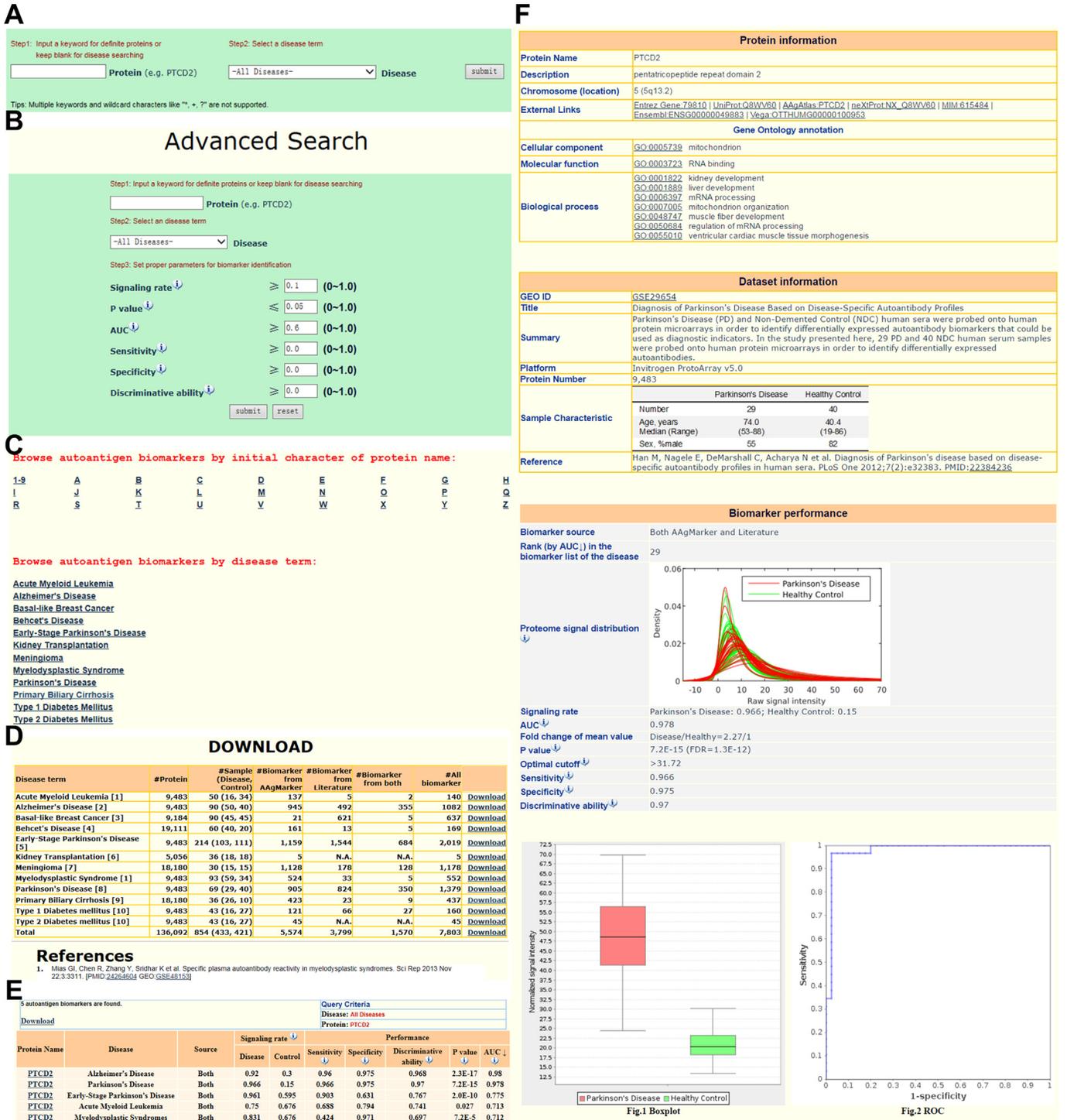


Figure 2. The web pages of AAgMarker 1.0. (A) Basic search; (B) advanced search; (C) browse; (D) download; (E) autoantigen biomarker list page; (F) detailed information page.

STATISTICS & COMPARISON

Currently, the AAgMarker 1.0 has collected high-throughput autoantibody (IgG) profiling of 12 diseases, including acute myeloid leukemia, Alzheimer's disease, Basal-like Breast Cancer, Bechet's disease, Early-Stage Parkinson's Disease, Kidney Transplantation, Meningioma, Myelodysplastic Syndromes, Parkinson's Disease, Primary Biliary Cirrhosis, Type 1 Diabetes Mellitus and Type 2 Diabetes Mellitus (Table 1; page of 'DOWNLOAD', Figure 2D). It contains 854 samples (433 for disease and 421 for healthy control), involving 136 092 proteins (i.e. autoantigens). A total of 7803 (4470 non-redundant) candidate biomarkers were identified for the 12 diseases. Among them, 5574 were identified using our unified methods, while 3799 were extracted from the literature. Our approach recovered 41.3% (1570/3799) of the biomarker candidates reported in the literature, a highly significant recover rate ($P = 0$) (Table 1). In other words, 20.1% (1570/7803) of all the biomarker candidates were identified with both ours and previously published approach (Figure 3A). Compared with the current methods, using one or two parameters would increase 6–9% of recovery rate from the reported biomarker candidates (identified literature biomarkers/all literature biomarkers), while decrease 10% of recovery rate from all biomarkers (identified literature biomarkers/all biomarkers) (Figure 3A), which could be easily analysed with the advanced search method. In addition, 316 and 44 biomarkers were also found in two autoantigen databases, text-mining-based AAgAtlas 1.0 and prediction-based CTdatabase, respectively (Figure 3B) (9,10).

Disease-specific or shared biomarkers

59.8% (2674/4470) biomarkers are exclusively associated with a single disease, and only 4.9% (226/4470) are found in more than four diseases (Figure 3C). The biomarker-disease network is shown in Figure 3D. Some diseases share more biomarker candidates than others and as a result, they are closer to each other in the network. These diseases are exemplified by Type 1 and Type 2 Diabetes Mellitus, and the neurological diseases, including Alzheimer's disease, Early-Stage Parkinson's disease, and Parkinson's disease.

DISCUSSION & CONCLUSION

Non-invasive biomarkers are in urgent need for diagnosis, prognosis and timely treatment of various diseases. Human serum autoantigens are a great resource of generating non-invasive biomarkers with well-established detection methods in the clinics, including the ELISA tests and Luminex. Here, we collected 12 proteome datasets of 12 diseases, and quantitatively identified 7803 serum autoantigen biomarkers including those identified in the corresponding literature. To date, many high-throughput projects of discovery and validation of serological biomarkers have been published and more are coming. Therefore, we strongly believe that this database will expand rapidly with a growing impact on the biomarker community.

We applied commonly used parameters to identify potential clinical biomarkers with good discriminative ability and

high detection signals, which enables users to easily understand and explore the database. We have found that some biomarkers could be used for distinguish of multiple diseases from healthy controls. For example, from both AAgMarker and literature, pentatricopeptide repeat domain 2 (PTCD2) could be serologically detected in five types of diseases, including Alzheimer's disease, Parkinson's disease, early-stage Parkinson's disease, acute myeloid leukemia and myelodysplastic syndromes (Figures 2E and 3D), which indicates PTCD2 could be highly immunogenic in patients suffering from many different types of diseases; however, they may not be very useful in distinguishing these diseases. In contrast, kelch like family member 7 (KLHL7) was an autoantigen biomarker exclusively for primary biliary cirrhosis with a specificity of 100.0%, but was not found significantly associated with the other 11 types of diseases (Figure 3D). Therefore, disease specificity can be an important attribute for a good biomarker, which can now be analysed by AAgMarker. Those 2647 single-disease related biomarkers could be potential biomarkers for specific disease (Figure 3C).

A particular issue should be noted is the consistence between autoantigen biomarkers produced by AAgMarker and those reported in the literature. As shown in Figure 3A, AAgMarker can recover only 41.3% of biomarkers from the literature, and 20.1% of all biomarkers from both sources. Many reasons, including different methods of data process and biomarker identification, could contribute to this discrepancy. Researchers may use different signal calling methods, such as Foreground-Background (17); they may use one or more different parameters, such as signalling rate of 15% in diseases and 0% in controls (18), sensitivity at 95% specificity (14); they may use different cutoff values, namely P values of <0.01 versus <0.05 (5). In addition, some studies only provided an incomplete list of biomarkers, or focused on biomarkers showing significantly high reactivity in patient groups compared to the healthy controls (13). Studies may use different reference groups from AAgMarker. For example, study on primary biliary cirrhosis compared the performance of the biomarkers in disease control groups in addition to the healthy controls, which is not currently available for all 12 diseases with AAgMarker (18). Although a uniform method could not make highly consistent results with other studies, a uniform method is needed for better comparison and most intersections are extremely significant (Table 1). Finally, AAgMarker provides advanced search enabling users to set their parameters for identification of biomarkers, and alternatively for users, AAgMarker included biomarkers identified from literature.

The AAgMarker will be updated regularly. Above all, qualified and relevant datasets will be standardized and incorporated into the database continuously when they become available. We are aware that researchers are performing antibody microarray profiling of sera from patients with various diseases including lung cancer, Takayasu Arteritis, Moyamoya disease and so on. We will include these datasets in our database when they become publically available. Moreover, we also plan to design a submission page to allow users to submit their own datasets. Finally, we are also considering to include the datasets for disease progression

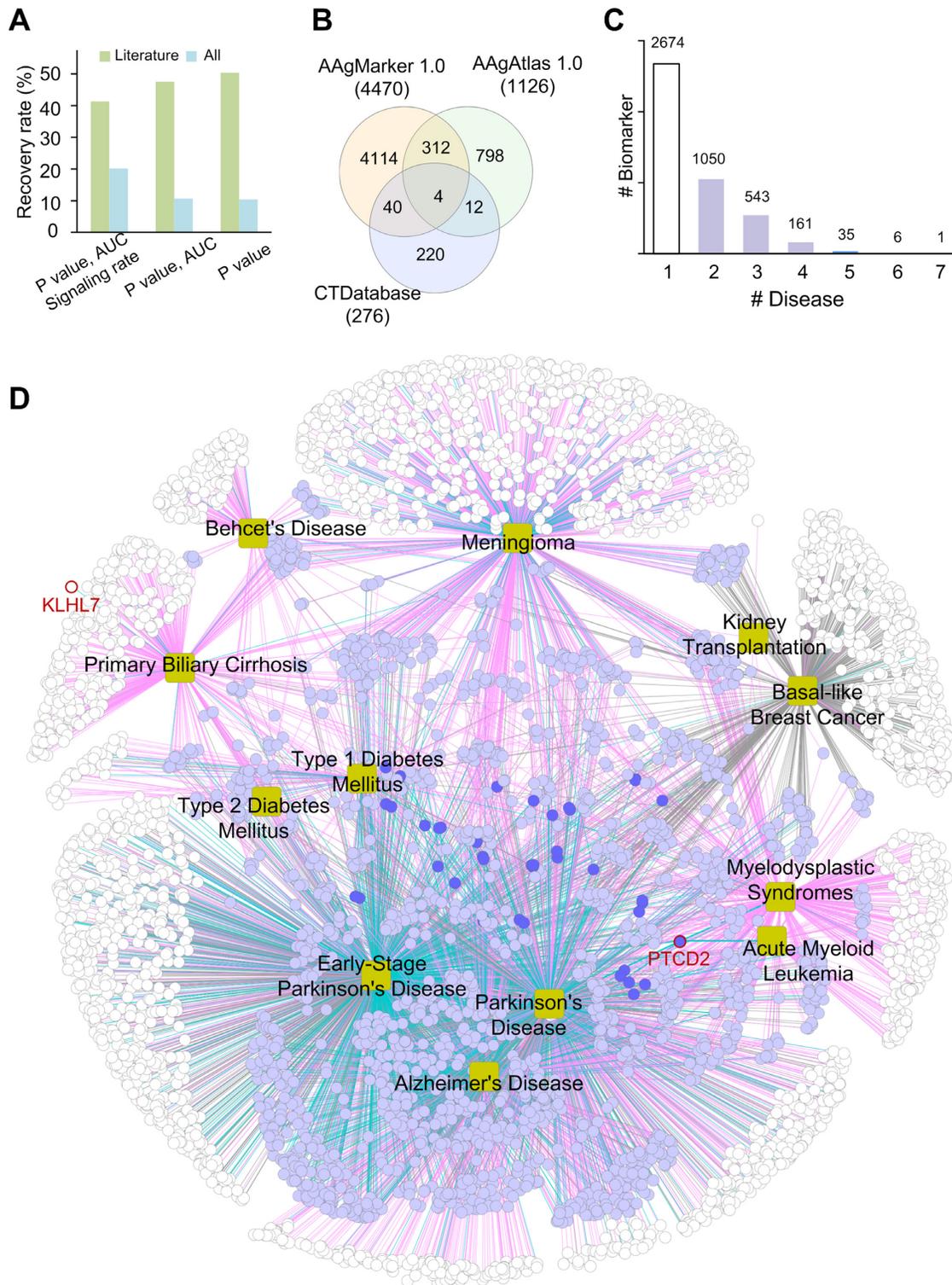


Figure 3. The statistics and comparison of AAgMarker 1.0. (A) Recovery rate of autoantigen biomarkers in literatures from different parameter combinations; (B) comparison of autoantigen biomarkers/autoantigens among AAgMarker 1.0, AAgAtlas 1.0 and CTDatabase; (C) statistics of biomarkers related to the number of diseases; (D) autoantigen biomarker-disease association network.

Table 1. The statistics of AAgMarker 1.0

| Disease | #Protein | #Disease | #Control | #Biomarkers by AAgMarker | #Biomarkers from Literature | #Biomarkers from both sources (hypergeometric P value) | #Total biomarkers |
|-------------------------------------|----------|----------|----------|--------------------------|-----------------------------|--|-------------------|
| Acute Myeloid Leukemia (13) | 9483 | 16 | 34 | 137 | 5 | 2 (0.002) | 140 |
| Alzheimer's Disease (5) | 9483 | 50 | 40 | 945 | 492 | 355 (1.0e-265) | 1082 |
| Basal-like Breast Cancer (14) | 9184 | 45 | 45 | 21 | 621 | 5 (0.011) | 637 |
| Behcet's Disease (15) | 19 111 | 40 | 20 | 161 | 13 | 5 (4.9e-08) | 169 |
| Early-Stage Parkinson's Disease (7) | 9483 | 103 | 111 | 1159 | 1544 | 684 (4.4e-290) | 2019 |
| Kidney Transplantation (16) | 5056 | 18 | 18 | 5 | N.A. | N.A. | 5 |
| Meningioma (17) | 18 180 | 15 | 15 | 1128 | 178 | 128 (8.0e-115) | 1178 |
| Myelodysplastic Syndromes (13) | 9483 | 59 | 34 | 524 | 33 | 5 (0.033) | 552 |
| Parkinson's Disease (6) | 9483 | 29 | 40 | 905 | 824 | 350 (7.9e-160) | 1379 |
| Primary Biliary Cirrhosis (18) | 18 180 | 26 | 10 | 423 | 23 | 9 (1.1e-09) | 437 |
| Type 1 Diabetes Mellitus (19) | 9483 | 16 | 27 | 121 | 66 | 27 (5.5e-35) | 160 |
| Type 2 Diabetes Mellitus (19) | 9483 | 16 | 27 | 45 | N.A. | N.A. | 45 |
| Total | 136 092 | 433 | 421 | 5574 | 3799 | 1570 (0) | 7803 |

or disease subtype identification. These datasets will greatly expand the content of the AAgMarker.

In conclusion, we have established a user-friendly database AAgMarker that deposits serum autoantigen biomarker candidates identified from high-throughput autoantibody profiling projects. To our knowledge, it is the most comprehensive public repository for quantitative evaluation of autoantigen-based serological biomarkers. We believe this database will be a valuable resource for the community of both biomedical and clinical research. It will not only aid autoantigen-based biomarker discovery for clinical diagnosis and prognosis of various diseases, but also promote protein annotation and exploring immunological mechanisms of diseases.

FUNDING

National Institutes of Health [R01GM111514 to J.Q., R01EY024580 to J.Q.]. Funding for open access charge: National Institutes of Health [R01GM111514 to J.Q., R01EY024580 to J.Q.].

Conflict of interest statement. None declared.

REFERENCES

- Biomarkers Definitions Working Group (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.*, **69**, 89–95.
- Subramanyam, M. and Goyal, J. (2016) Translational biomarkers: From discovery and development to clinical practice. *Drug Discov. Today Technol.*, **21–22**, 3–10.
- Naparstek, Y. and Plotz, P.H. (1993) The role of autoantibodies in autoimmune disease. *Annu. Rev. Immunol.*, **11**, 79–104.
- Scofield, R.H. (2004) Autoantibodies as predictors of disease. *Lancet*, **363**, 1544–1546.
- Nagele, E., Han, M., DeMarshall, C., Belinka, B. and Nagele, R. (2011) Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. *PLoS One*, **6**, e23112.
- Han, M., Nagele, E., DeMarshall, C., Acharya, N. and Nagele, R. (2012) Diagnosis of Parkinson's disease based on disease-specific autoantibody profiles in human sera. *PLoS One*, **7**, e32383.
- DeMarshall, C.A., Han, M., Nagele, E.P., Sarkar, A., Acharya, N.K., Godsey, G., Goldwasser, E.L., Kosciuk, M., Thayasivam, U., Belinka, B. *et al.* (2015) Potential utility of autoantibodies as blood-based biomarkers for early detection and diagnosis of Parkinson's disease. *Immunol. Lett.*, **168**, 80–88.
- Elkon, K. and Casali, P. (2008) Nature and functions of autoantibodies. *Nat. Clin. Pract. Rheumatol.*, **4**, 491–498.
- Wang, D., Yang, L., Zhang, P., LaBaer, J., Hermjakob, H., Li, D. and Yu, X. (2017) AAgAtlas 1.0: A human autoantigen database. *Nucleic Acids Res.*, **45**, D769–D776.
- Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., Chen, Y.T., Chua, R., Gurung, S., Gnjjatic, S. *et al.* (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res.*, **37**, D816–D819.
- Wang, X., Zhao, H., Xu, Q., Jin, W., Liu, C., Zhang, H., Huang, Z., Zhang, X., Zhang, Y., Xin, D. *et al.* (2006) HPTaa database-potential target genes for clinical diagnosis and immunotherapy of human carcinoma. *Nucleic Acids Res.*, **34**, D607–D612.
- Nagele, E.P., Han, M., Acharya, N.K., DeMarshall, C., Kosciuk, M.C. and Nagele, R.G. (2013) Natural IgG autoantibodies are abundant and ubiquitous in human sera, and their number is influenced by age, gender, and disease. *PLoS One*, **8**, e60726.
- Mias, G.I., Chen, R., Zhang, Y., Sridhar, K., Sharon, D., Xiao, L., Im, H., Snyder, M.P. and Greenberg, P.L. (2013) Specific plasma autoantibody reactivity in myelodysplastic syndromes. *Sci. Rep.*, **3**, 3311.
- Wang, J., Figueroa, J.D., Wallstrom, G., Barker, K., Park, J.G., Demirkan, G., Lissowska, J., Anderson, K.S., Qiu, J. and LaBaer, J. (2015) Plasma autoantibodies associated with basal-like breast cancers. *Cancer Epidemiol. Biomarkers Prev.*, **24**, 1332–1340.
- Hu, C.J., Pan, J.B., Song, G., Wen, X.T., Wu, Z.Y., Chen, S., Mo, W.X., Zhang, F.C., Qian, J., Zhu, H. *et al.* (2017) Identification of novel biomarkers for Behcet disease diagnosis using human proteome microarray approach. *Mol. Cell. Proteomics*, **16**, 147–156.
- Li, L., Wadia, P., Chen, R., Kambham, N., Naesens, M., Sigdel, T.K., Miklos, D.B., Sarwal, M.M. and Butte, A.J. (2009) Identifying compartment-specific non-HLA targets after renal transplantation by integrating transcriptome and “antibodyome” measures. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4148–4153.
- Gupta, S., Mukherjee, S., Syed, P., Pandala, N.G., Choudhary, S., Singh, V.A., Singh, N., Zhu, H., Epari, S., Noronha, S.B. *et al.* (2017) Evaluation of autoantibody signatures in meningioma patients using human proteome arrays. *Oncotarget*, doi:10.18632/oncotarget.16997.
- Hu, C.J., Song, G., Huang, W., Liu, G.Z., Deng, C.W., Zeng, H.P., Wang, L., Zhang, F.C., Zhang, X., Jeong, J.S. *et al.* (2012) Identification of new autoantigens for primary biliary cirrhosis using human proteome microarrays. *Mol. Cell. Proteomics*, **11**, 669–680.
- Koo, B.K., Chae, S., Kim, K.M., Kang, M.J., Kim, E.G., Kwak, S.H., Jung, H.S., Cho, Y.M., Choi, S.H., Park, Y.J. *et al.* (2014) Identification of novel autoantibodies in type 1 diabetic patients using a high-density protein microarray. *Diabetes*, **63**, 3022–3032.
- Jett, J.R., Peek, L.J., Fredericks, L., Jewell, W., Pingleton, W.W. and Robertson, J.F. (2014) Audit of the autoantibody test, EarlyCDT(R)-lung, in 1600 patients: an evaluation of its performance in routine clinical practice. *Lung Cancer*, **83**, 51–55.
- Kokkonen, H., Mullazehi, M., Berglin, E., Hallmans, G., Wadell, G., Ronnelid, J. and Rantapaa-Dahlqvist, S. (2011) Antibodies of IgG, IgA and IgM isotypes against cyclic citrullinated peptide precede the development of rheumatoid arthritis. *Arthritis Res. Ther.*, **13**, R13.

22. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
23. Kolesnikov,N., Hastings,E., Keays,M., Melnichuk,O., Tang,Y.A., Williams,E., Dylag,M., Kurbatova,N., Brandizi,M., Burdett,T. *et al.* (2015) ArrayExpress update–simplifying data submissions. *Nucleic Acids Res.*, **43**, D1113–D1116.
24. Xu,Z., Huang,L., Zhang,H., Li,Y., Guo,S., Wang,N., Wang,S.H., Chen,Z., Wang,J. and Tao,S.C. (2016) PMD: A resource for archiving and analyzing protein microarray data. *Sci. Rep.*, **6**, 19956.
25. Hu,S., Wan,J., Su,Y., Song,Q., Zeng,Y., Nguyen,H.N., Shin,J., Cox,E., Rho,H.S., Woodard,C. *et al.* (2013) DNA methylation presents distinct binding sites for human transcription factors. *Elife*, **2**, e00726.
26. Boutet,E., Lieberherr,D., Tognolli,M., Schneider,M., Bansal,P., Bridge,A.J., Poux,S., Bougueleret,L. and Xenarios,I. (2016) UniProtKB/swiss-prot, the manually annotated section of the UniProt KnowledgeBase: How to use the entry view. *Methods Mol. Biol.*, **1374**, 23–54.
27. Brown,G.R., Hem,V., Katz,K.S., Ovetsky,M., Wallin,C., Ermolaeva,O., Tolstoy,I., Tatusova,T., Pruitt,K.D., Maglott,D.R. *et al.* (2015) Gene: A gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
28. Gaudet,P., Michel,P.A., Zahn-Zabal,M., Cusin,I., Duek,P.D., Evalet,O., Gateau,A., Gleizes,A., Pereira,M., Teixeira,D. *et al.* (2015) The neXtProt knowledgebase on human proteins: Current status. *Nucleic Acids Res.*, **43**, D764–D770.