

# AAtlas 1.0: a human autoantigen database

Dan Wang<sup>1</sup>, Liuhui Yang<sup>1</sup>, Ping Zhang<sup>1</sup>, Joshua LaBaer<sup>2</sup>, Henning Hermjakob<sup>1,3,\*</sup>, Dong Li<sup>1,\*</sup> and Xiaobo Yu<sup>1,\*</sup>

<sup>1</sup>State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences-Beijing (PHOENIX Center), Beijing Institute of Radiation Medicine, Beijing 102206, China, <sup>2</sup>The Virginia G. Piper Center for Personalized Diagnostics, Biodesign Institute, Arizona State University, Tempe, AZ 85287, USA and <sup>3</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Received August 21, 2016; Revised September 22, 2016; Editorial Decision October 09, 2016; Accepted October 11, 2016

## ABSTRACT

**Autoantibodies refer to antibodies that target self-antigens, which can play pivotal roles in maintaining homeostasis, distinguishing normal from tumor tissue and trigger autoimmune diseases. In the last three decades, tremendous efforts have been devoted to elucidate the generation, evolution and functions of autoantibodies, as well as their target autoantigens. However, reports of these countless previously identified autoantigens are randomly dispersed in the literature. Here, we constructed an AAtlas database 1.0 using text-mining and manual curation. We extracted 45 830 autoantigen-related abstracts and 94 313 sentences from PubMed using the keywords of either ‘autoantigen’ or ‘autoantibody’ or their lexical variants, which were further refined to 25 520 abstracts, 43 253 sentences and 3984 candidates by our bio-entity recognizer based on the Protein Ontology. Finally, we identified 1126 genes as human autoantigens and 1071 related human diseases, with which we constructed a human autoantigen database (AAtlas database 1.0). The database provides a user-friendly interface to conveniently browse, retrieve and download human autoantigens as well as their associated diseases. The database is freely accessible at <http://biokb.ncpsb.org/aagatlas/>. We believe this database will be a valuable resource to track and understand human autoantigens as well as to investigate their functions in basic and translational research.**

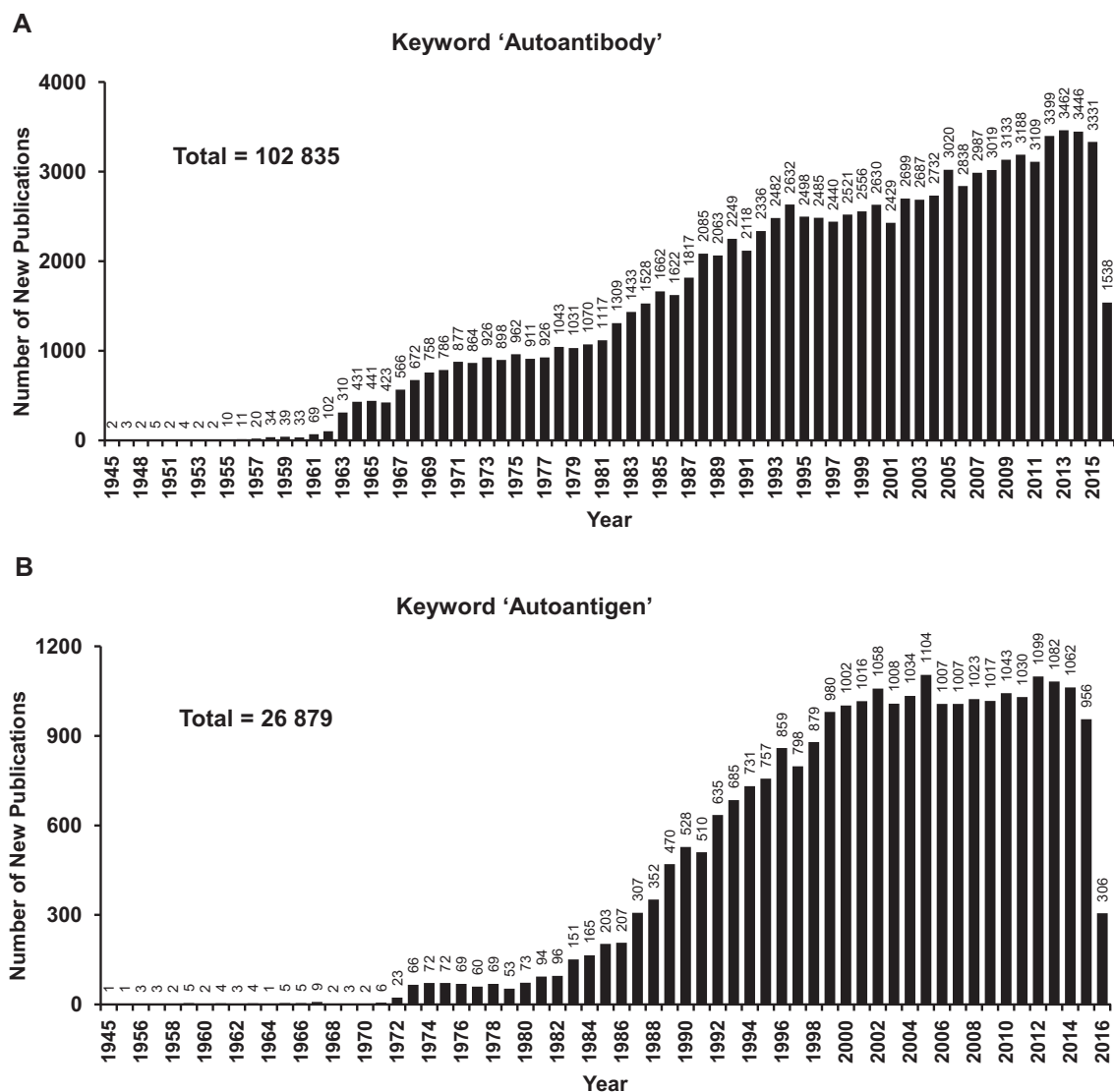
## INTRODUCTION

Autoantibodies (AAb) are antibodies that bind to individual's own antigens. Their production can be affected by ge-

netic predisposition, environment, viruses and pathogens (via epitope mimicking) and chemicals, etc. In the last three decades numerous studies have been devoted to elucidating the generation, evolution and functions of AAbs as well as their target autoantigens (AAGs) (Figure 1) (1–3). The accumulating evidence indicates that AAbs play pivotal roles in maintaining homeostasis of healthy individuals and cancer patients through auto-clearance of aged cells and dysfunctional dividing cells, respectively (2,4). Moreover, disorders of the immune system may lead to autoimmune diseases. Multiple AAbs that are specific to autoimmune disease have been found, including thyroiditis, type 1 diabetes mellitus and primary biliary cirrhosis, indicating the contribution of AAbs to pathology by inflammatory stimulation (5–7).

Circulating AAbs can serve as biomarkers for the early diagnostics, prognosis and therapeutic treatment of human diseases (8–11). For example, an AAg biomarker panel (p53, NY-ESO-1, CAGE, GBU4-5, Annexin1 and SOX2), called EarlyCDT<sup>®</sup>Lung test, has been approved by FDA for the early detection of lung cancer in high risk and asymptomatic patients prior to standard CT imaging test. The panel showed sensitivity and specificity of 46% (12/26) and 83% (599/726) for the early stage of lung cancer (12). Using protein microarrays with 4988 human proteins, Anderson *et al.* found a panel of 28 AAGs for the detection of early stage breast cancer with sensitivity of 80.8% and specificity of 61.6%, respectively (10). These AAbs have been integrated with several protein biomarkers and become the first protein-based blood test, Videssa<sup>®</sup> Breast, provided by Provista Diagnostics for early breast cancer detection (13). Using bead-based arrays, Ayoglu *et al.* profiled the sera from multiple sclerosis patients with 11 520 proteins and identified Anoctamin 2 as an autoimmune target in multiple sclerosis (14). In addition, some tumor AAGs might be ideal targets for targeted immunotherapy such as p53, HER2 and NY-ESO-1 (15,16). For example, HER2 is highly expressed on the surface of some cancer cells, and several tar-

\*To whom correspondence should be addressed. Tel: +86 10 61777093; Fax: +86 10 61777050; Email: xiaobo.yu@hotmail.com  
Correspondence may also be addressed to Henning Hermjakob. Tel: +44 1223 494 671; Fax: +44 1223 494 468; Email: hhe@ebi.ac.uk  
Correspondence may also be addressed to Dong Li. Tel: +86 10 61777057; Fax: +86 10 61777050; Email: lidong.bprc@foxmail.com



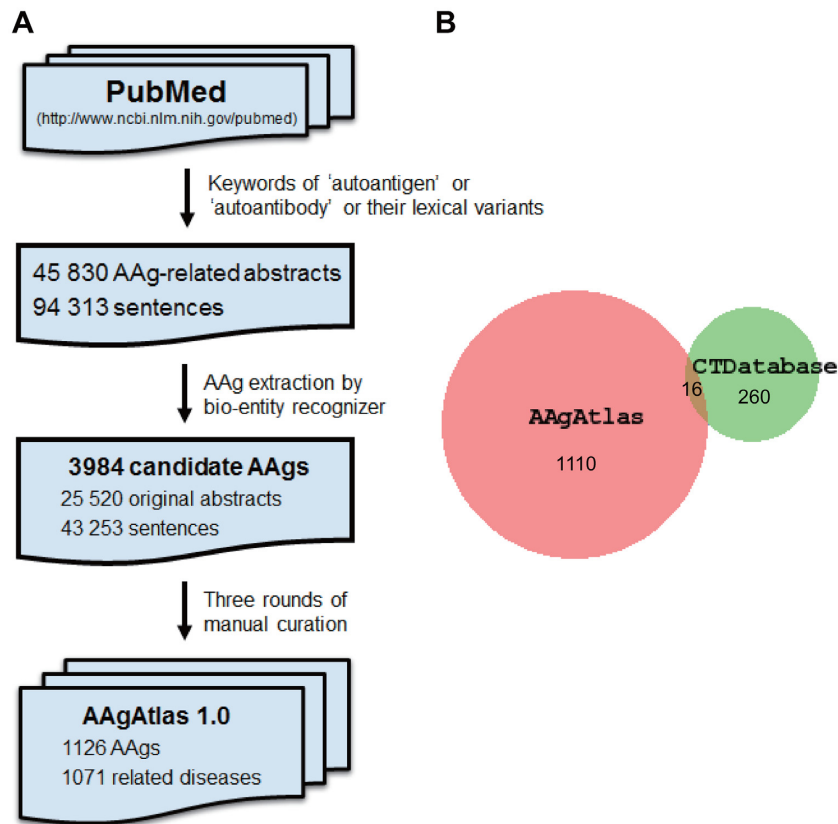
**Figure 1.** Number of new publications for autoantigen and autoantibodies from 1945 to 2016. The search was performed using the keyword of either (A) autoantigen or (B) autoantibody in PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) on July 9, 2016; the total number was calculated using the sum of publications for autoantigen and autoantibody, respectively.

geted therapies directly against HER2 have been employed to treat breast and stomach cancers, i.e. trastuzumab (Herceptin®) (17).

However, the information of previously identified AAGs is dispersed in thousands of published papers, and the systematic investigation of human AAGs is hampered due to unavailability of a comprehensive human AAG collection. To address this need, Wang *et al.* constructed a Human Potential Tumor Associated Antigen (HPtaa) database that collected 3518 potential tumor antigens identified by *in silico* computing (<http://www.hptaa.org>) based on tumor-specific expression, coding capacity, chromosomal location, subcellular location and the knowledge of gene function (18). Almeida *et al.* constructed a cancer testis antigen database, CTdatabase (<http://www.cta.lncc.br>), which contains 276 genes that were compiled manually from the literature and computational prediction (19). However, these

two databases only cover a small fraction of the published AAGs and their inclusion of predicted but not documented AAGs can be confusing (Figure 2B).

We have built an AAGAtlas database using text-mining and manual curation with the aim to provide a comprehensive AAG repository that supports basic and translational studies. We extracted 45 830 AAG-related abstracts and 94 313 sentences from PubMed by the keywords of either 'autoantigen' or 'autoantibody' or their lexical variants, which were further refined to 25 520 abstracts, 43 253 sentences and 3984 candidates by bio-entity recognizer based on the Protein Ontology (20,21). A total of 1126 genes were selected as human AAGs after three rounds of manual curation, and 1071 associated human diseases were identified based on the Human Disease Oncology (22) accordingly. AAGAtlas database 1.0 provides a user-friendly interface to



**Figure 2.** Construction of AAgAtlas 1.0 database. (A) The workflow of AAgAtlas database construction; (B) Comparison of human antigens between AAgAtlas and CTDatabase.

conveniently browse, retrieve and download the list of human AAg and their related diseases.

## DATA COLLECTION AND MANUAL CURATION

Our data collection relied on the text mining of AAg-related PubMed abstracts. Bio-entity recognition and extraction from these abstracts for human AAg candidates were performed by our custom developed ontology-based bio-entity recognizer. After evaluation against the CRAFT corpus for gene/protein recognition based on Protein Ontology (PR) (20,21), the F-measure, the point where precision and recall are equal (23), of our recognition tool is 0.845, which is on a par with that of current state-of-the-art biomedical annotation system, BeCAS, 0.76 (24).

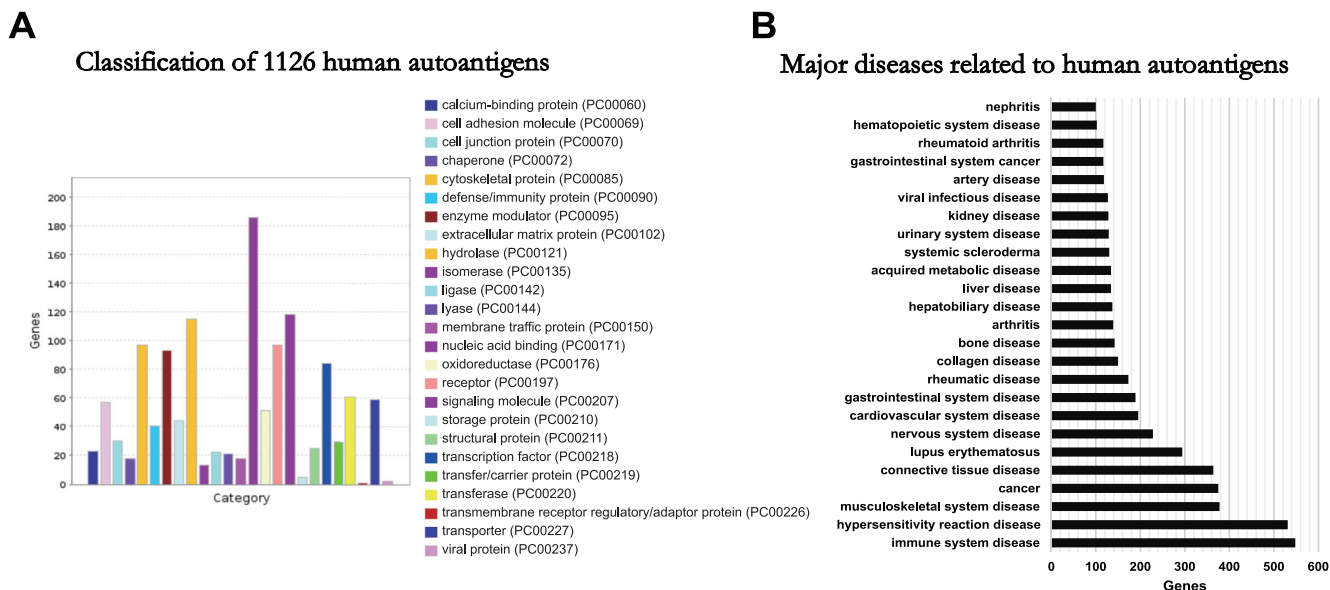
We compiled a list of human AAg together with their related diseases as well as the evidence from PubMed abstract using the following three steps. All PubMed abstracts were fetched through the NCBI E-utilities API (25) and the AAg-related abstracts were extracted by our bio-entity recognizer with the keywords of either 'autoantigen' or 'autoantibody' or their lexical variants like 'auto-antigen', 'auto antigen', 'autoantigens', 'auto-antigens', 'auto antigens', 'auto-antibody', 'auto antibody', 'autoantibodies', 'auto-antibodies' or 'autoantibodies'. 45,830 AAg-related abstracts and 94,313 sentences were obtained.

Then, a dictionary of human gene/protein names and synonyms was built by integrating all terms and their cor-

respondent synonyms from the Protein Ontology. These were mapped to all the official symbols of human genes provide by HUGO Gene Nomenclature Committee ([www.genenames.org/](http://www.genenames.org/)) (26). The human genes that were recognized and extracted from the previously generated set of sentences were considered as candidate AAg.

All 3984 candidates from 25 520 original abstracts and 43 253 evidence sentences in which the gene and keyword co-occurred were validated through three rounds of manual curation. First, all sentences with AAg names were checked and selected by two experienced researchers independently; second, these selected sentences were submitted to an internal review, in which all AAg names were manually reviewed and approved one by one by a reviewer panel consisting of three experts; third, we asked all co-authors to randomly check AAg from our website to make sure that all genes loaded into our database are *bona fide* AAg.

Disease terms were extracted from the abstracts by our recognition tool based on the Human Disease Ontology (22,27). The logic relation type between child disease term node and parent node in the ontology has been taken into consideration as in NCBO Annotator (28), i.e. if a child disease term is associated with an AAg, then its Human Disease Ontology parent disease terms are supposed to be associated as well. Associations between AAg genes/proteins and human disease terms were generated by sentence-level co-occurrence and scored by co-occurrence count. All extracted AAg genes/proteins, human disease terms as well



**Figure 3.** Analysis of human autoantigen's protein classes and major related diseases. (A) Classification of all human autoantigens using PANTHER (<http://pantherdb.org/>) (42). (B) Major diseases associated with human autoantigens. The diseases containing more than 100 human autoantigens were selected.

as their associations and evidence abstracts with biomedical keywords highlighted were stored in the our database for user query and navigation (Figure 2A).

We obtained 1126 human AAGs and 1071 related diseases and constructed the human AAG database (AAGAtlas 1.0 database) (Figures 2-4). All evidence sentences that have been manually curated for AAGs are regarded as validated evidence and included on the top of each AAG page (Supplementary Figure S1). Considering all AAGs and their related diseases were documented from previous publications that might not be fully validated, we added a simple community curation function to the evidence phrases with which the users can easily provide their feedback by simply clicking the green 'Yes' or red 'No' button after login as registered users (Supplementary Figure S1 and Figure S2). With this function, the users can confirm or reject that the evidence phrase has the correct AAG and related disease.

## DATA SEARCH AND NAVIGATION

We provide a user-friendly web interface that facilitates searching, exploring information and updating the AAGAtlas 1.0 database (<http://biokb.ncpsb.org/aagatlas/>) (Figure 4). The website interface comprises six sections including 'Home', 'Browse & Download', 'FAQ', 'Contact', 'Feedback' and 'Log in'. From the 'Home' page, the users can see the introduction to AAGAtlas 1.0 database and search for their gene or disease of interest. By selecting 'Browse & Download' in the navigation bar, the complete list of human AAGs, their related diseases and supporting evidence can be browsed and downloaded in one page. From the 'FAQ' page, the users can see the general questions and answers. From the 'Feedback' page, the users can submit new genes to our database manually. From the 'Log in' page, the users can make the registration and start do the manual curation

on our website. Our database will be updated periodically in future according to these information provided by the users.

Two query approaches are provided for searching AAG: query by gene symbol and query by disease term.

### Query by gene symbol

For the query by gene symbol, the users can enter an NCBI Entrez gene symbol in the 'Gene Symbol' search box, and a drop down menu will provide auto-completed gene symbols present in the AAGAtlas (Supplementary Figure S1). After selecting one of them and clicking the 'Search' button, the search engine will return the queried results containing a table showing the gene name, related diseases as well as the supporting literature evidence. By clicking the hyperlink of the gene symbol for an individual AAG in the 'Gene' column, users can obtain the basic information for the AAG gene and the cross references to external databases (i.e. Ensembl (29), NCBI Gene (30), UniProtKB (31), neXtProt (32) and Antibodypedia (33)).

For example, p53 is perhaps one of the most famous AAG whose antibodies are frequently up-regulated in serum or plasma of cancer patients (34). Here, we searched our database with 'TP53' and the results revealed that p53 might be involved into more than one hundred different diseases such as liver cancer, neck cancer, lymphoma and rheumatic disease, *etc* (Supplementary Table S1). This result emphasizes the role of p53 in the development of different human diseases.

### Query by disease term

For the query by disease term, the user can enter a disease term in the 'Disease Term' search box, and a drop down menu will provide disease terms that were extracted

Search AAgAtlas

**Gene Symbol?**

E.g. BRCA1

**Disease Term?**

E.g. Breast carcinoma

Introduction

Autoantibodies are antibodies that are produced by the B-cell immune system against an individual's own antigens and play pivotal roles in the maintenance of healthy individuals' homeostasis, as well as in tumors and autoimmune diseases. In the last two decades, tremendous efforts have been devoted to elucidate the generation, evolution and function of autoantibodies and their targets, autoantigens [1-9]. However, the previously identified autoantigens are randomly dispersed in the literature. Here we present a database containing 1,126 well-annotated autoantigens, determined by text-mining and manual curation [10]. AAgAtlas database 1.0 provides a user-friendly interface to conveniently browse, retrieve and download the list of autoantigens and their associated diseases. This database is freely accessible at <http://biokb.ncpsb.org/aagatlas>.

References

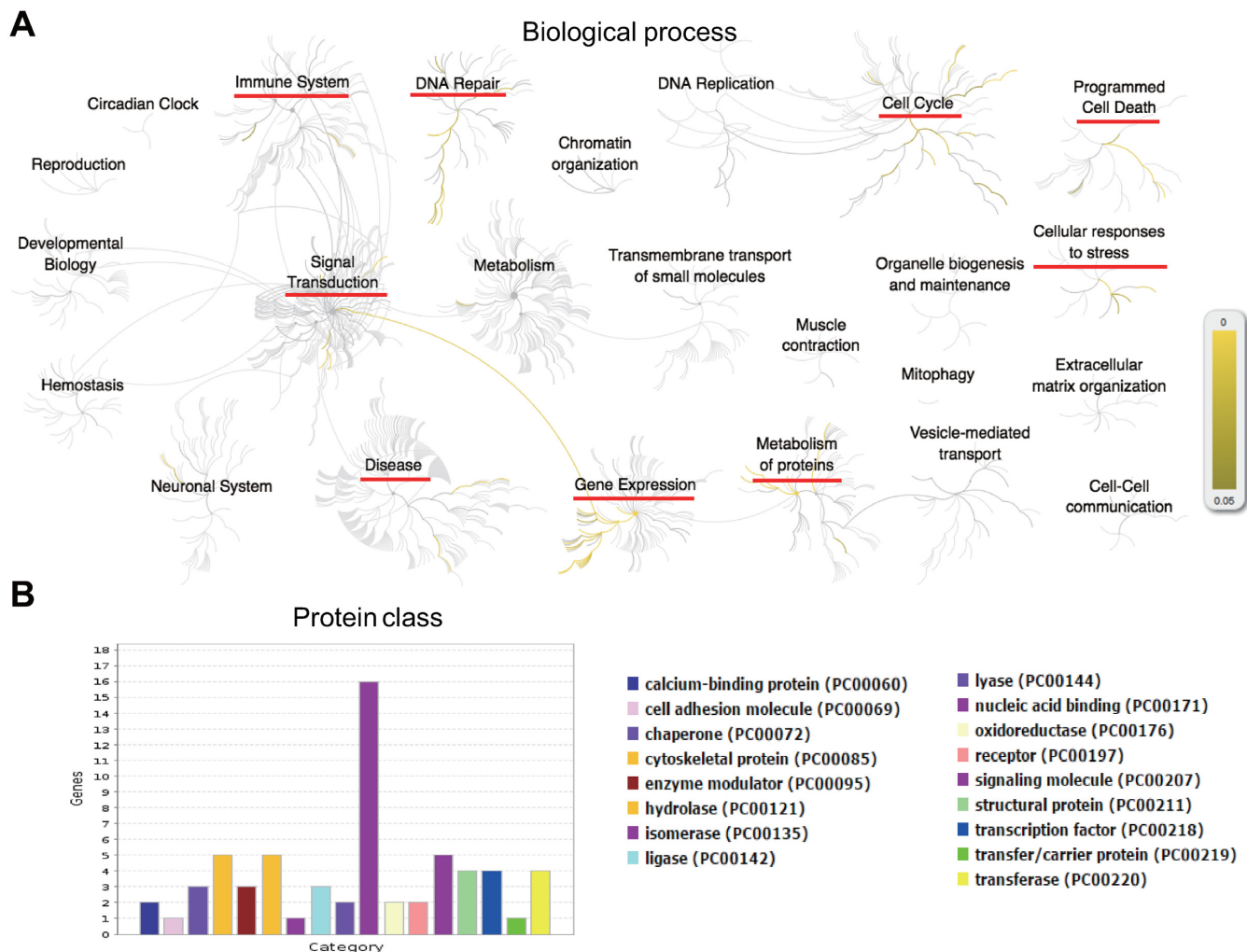
1. Plotz PH. The autoantibody repertoire: searching for order. *Nat Rev Immunol.* 2003 3(1):73-8.
2. Elkon K, Casali P. Nature and functions of autoantibodies. *Nat Clin Pract Rheumatol.* 2008 4(9):491-8.
3. Madi A, Hecht I, Bransburg-Zabary S, et al. Organization of the autoantibody repertoire in healthy newborns and adults revealed by system level informatics of antigen microarray data. *Proc Natl Acad Sci U S A.* 2009 106(34):14484-9.
4. Anderson KS, LaBaer J. The sentinel within: exploiting the immune system for cancer biomarkers. *J Proteome Res.* 2005 4(4):1123-33.
5. Olsen NJ, Karp DR. Autoantibodies and SLE—the threshold for disease. *Nat Rev Rheumatol.* 2014 10(3):181-6.
6. Wang X, Yu J, Sreekumar A, Varambally S, et al. Autoantibody signatures in prostate cancer. *N Engl J Med.* 2005 353(12):1224-35.
7. Wang J, Figueroa JD, Wallstrom G, et al. Plasma Autoantibodies Associated with Basal-like Breast Cancers. *Cancer Epidemiol Biomarkers Prev.* 2015 24(9):1332-40.
8. Mellman I, Coukos G, Dranoff G. Cancer immunotherapy comes of age. *Nature.* 2011 480(7378):480-9.
9. Jackson HJ, Rafiq S, Brentjens RJ, Driving CAR T-cells forward. *Nat Rev Clin Oncol.* 2016 13(6):370-83.
10. Nunes T, Campos D, Matos S, et al. BeCAS: biomedical concept recognition services and visualization. *Bioinformatics.* 2013, 29(15): 1915-1916.

**Figure 4.** Screenshot of the AAgAtlas database website. The website comprises 'Home', 'Browse & Download', 'FAQ', 'Contact', 'Feedback' and 'Log in' sections. From the 'Home' page, the users can see the introduction of AAgAtlas 1.0 database and search for the gene or disease of their interest. From the 'Browse & Download', the users can see and download the human AAg, their related diseases and supporting evidences. The 'FAQ' page shows the general questions and answers. The 'Contact' page shows the emails of web managers that are used for further communications. With the 'Feedback' page, the users can submit new genes to our database manually. From the 'Log in' page, the users can make the registration and start do the manual curation on our website. Our database will be updated periodically according to these information provided by the users.

with AAg from literature, i.e. Breast carcinoma (Supplementary Figure S2). After selecting one of them and clicking the 'Search' button, the search engine will return the queried results containing a table showing the disease name, related AAg genes, and the supporting literature evidence. When the user clicks the small triangle on the head of each table column, the results in the table will be resorted in ascending/descending order for that column. The users can also input words or select a term in the drop list of the text box to filter listed genes or diseases from the results by substring match. In addition, when the users click the number

of the evidence abstracts or sentences, a table containing gene, disease, PubMed ID, evidence and manual validation information will be displayed. The users can click the hyperlink of evidence to see the original abstract in which the key words, i.e. gene and disease names, are also highlighted.

For example, as the sixth most common cancer and the second leading cause of cancer death in the world, liver cancer is closely associated with the body's immune system that could either evacuate the abnormally growing cells or attack healthy liver cells causing inflammation and liver damage (35,36). We searched our database with 'Liver cancer'



**Figure 5.** Bioinformatics analysis of the human autoantigens associated with liver cancer. (A) Biological and pathway analysis using Reactome (<http://www.reactome.org/>). (B) Protein class analysis using PANTHER (<http://pantherdb.org/>) (42).

and found 53 AAg genes. The analysis with Reactome (<http://www.reactome.org/>) (37) reveals that these genes participate in the DNA repair, signal transduction, gene expression, immune system, metabolism of proteins, cell cycle, programmed cell death and cellular responses to stress (Figure 5A, Supplementary Table S2 and S3). The protein class analysis indicates that these AAGs are the proteins belonging to nucleic acid binding, enzymes, cytoskeletal protein, receptor, signaling molecule and transcription factors, etc. (Figure 5B). The results align well with previously reported results on liver cancer studies (38) and suggest that AAGs might be active proteins in the development of human disease.

## DATABASE IMPLEMENTATION AND DESIGN

AAGAtlas database runs entirely on open software (MySQL database server, Apache web server with the interface written in PHP, Ubuntu Linux operating system). Python scripts were written for data collection and processing and text mining. The web interface is available online at <http://biokb.ncpsb.org/aagatlas/>.

## DISCUSSION AND CONCLUSION

Since the first AAg paper appeared in 1945, more than 120 000 papers have been published and 1000 AAGs have been identified. Some of them have been shown to be important regulators or biomarkers in both healthy people and patients of different diseases. Reports of these AAGs occur in thousands of paper scattered through the literature (Figure 1) and construction of a systematic database for AAGs would be valuable to study the generation, evolution and function of AAG as well as to elucidate the mechanism of our human immune system in maintaining homeostasis and evacuating abnormal growing cells.

In this work, we searched and extracted the abstracts of all publications from Pubmed with 'Autoantibody', 'Autoantigen' and their synonyms using text mining and a self-developed ontology-based bio-entity recognizer. We evaluated our data set against the CRAFT corpus for gene/protein recognition based on Protein Ontology (PR) (20,21), and the results indicate that the F-measure (0.835) of our recognition tool is comparable to that of the state-of-art biomedical annotation system, BeCAS (0.76) (24). We

compiled a list of 1126 human AAgS followed after a series of manual curation and constructed the AAgAtlas 1.0 database (Figures 2-4). We compared AAgAtlas 1.0 (1126 antigens) database to CTdatabase (276 antigens) and found that only 26 antigens overlapped (Figure 2B). This might be due to that many antigens from CTdatabase were selected based on computational prediction (19). The website of HP-taa database is not available and thus the comparison to AAgAtlas was not performed. The results here suggest that AAgAtlas 1.0 database is the largest collection of human AAgS to date.

In addition, the comprehensive collection of AAgAtlas 1.0 database allows us to have an overview of human AAgS' protein classes and their related diseases. The gene ontology analysis reveals that the largest class of autoantigen proteins is nucleic acid binding, followed by signalling molecule, hydrolase, cytoskeletal protein, receptor, enzyme modulator and transcription factor (Figure 3A). The disease ontology analysis results suggest that human AAg proteins are actively involved in major human diseases including immune system, hypersensitivity reaction, musculoskeletal system and cancer (Figure 3B). All these results demonstrate the importance of our AAg collection and show the value of AAgAtlas 1.0 database in human AAg studies.

However, several issues should be considered for AAgAtlas 1.0 database. First, some AAgS in the full texts of the literature are not shown in abstracts and might not be included in our database. Currently there is no universal or centralized system capable of screening full text articles (39), especially for the journals from most publishers that are not freely available. The published abstracts are still the main resource for text-mining to date (40). To obtain AAgS from full text, we have to figure out a systematic plan to obtain as many of these full-text papers as possible and extract AAgS from them, which will be executed in the next version of AAgAtlas database. Second, some AAbs recognize the post-translational modifications instead of protein epitopes. In this work, all AAgS were selected based on the proteins of AAb targeted regardless of post-translational modifications (3). This may be due to the lack of high throughput AAb screening platforms that display proteins with modifications, although one was recently published (41). Third, AAbs that recognize non protein biomolecules such as lipids, DNA or small molecules (3), were not considered in our database. Fourth, our curation process only confirmed that authors reported them as AAg and did not validate the supporting data from the experiments. To address this issue, we added a manual curation function to the end of each evidence phase with which the users can simply click the green 'Yes' or red 'No' button to confirm or deny our evidence (Supplementary Figure S1 and Figure S2). Our database system will be updated periodically.

In conclusion, we identified 1126 AAgS, covering ~5% of the human genome, and 1071 related human diseases by text mining and manual curation. We built the AAgAtlas 1.0 database and website (<http://biokb.ncpsb.org/aagatlas/>) that is freely available to the scientific community to browse, search and download. It also provides a simple community curation interface for future enhancements. We believe it will be a valuable resource for scientific community to in-

vestigate the functions of human AAgS and their related diseases in the future.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors would like to thank Beijing Genestone Technology Ltd for their support on the website development.

## FUNDING

National Natural Science Foundation of China [81673040 to X.Y.]; Beijing Municipal Human Resources and Social Security Bureau [grant for Overseas Scholars to X.Y.]; Guangdong Science and Technology Department [2016A020215004 to X.Y.]; International S&T Cooperation Program [2014DFB30020 to D.L.]; Program of Precision Medicine [2016YFC0901905 to D.L.]; Early Detection Research Network [5U01CA117374 to J.L.]; Funding for open access charge: National Natural Science Foundation of China [81673040 to X.Y.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Madi,A., Hecht,I., Bransburg-Zabary,S., Merbl,Y., Pick,A., Zucker-Toledano,M., Quintana,F.J., Tauber,A.I., Cohen,I.R. and Ben-Jacob,E. (2009) Organization of the autoantibody repertoire in healthy newborns and adults revealed by system level informatics of antigen microarray data. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 14484–14489.
- Elkon,K. and Casali,P. (2008) Nature and functions of autoantibodies. *Nat. Clin. Pract. Rheumatol.*, **4**, 491–498.
- Plotz,P.H. (2003) The autoantibody repertoire: searching for order. *Nat. Rev. Immunol.*, **3**, 73–78.
- Hanash,S.M., Baik,C.S. and Kallioniemi,O. (2011) Emerging molecular biomarkers—blood-based strategies to detect and monitor cancer. *Nat. Rev. Clin. Oncol.*, **8**, 142–150.
- Flach,A.C., Litke,T., Strauss,J., Haberl,M., Gomez,C.C., Reindl,M., Saiz,A., Fehling,H.J., Wienands,J., Odoardi,F. *et al.* (2016) Autoantibody-boosted T-cell reactivation in the target organ triggers manifestation of autoimmune CNS disease. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 3323–3328.
- Olsen,N.J. and Karp,D.R. (2014) Autoantibodies and SLE: the threshold for disease. *Nat. Rev. Rheumatol.*, **10**, 181–186.
- Burbelo,P.D. and O'Hanlon,T.P. (2014) New autoantibody detection technologies yield novel insights into autoimmune disease. *Curr. Opin. Rheumatol.*, **26**, 717–723.
- Yu,X., Petritis,B. and LaBaer,J. (2016) Advancing translational research with next generation protein microarrays. *Proteomics*, **16**, 1238–1250.
- 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.
- Anderson,K.S., Sibani,S., Wallstrom,G., Qiu,J., Mendoza,E.A., Raphael,J., Hainsworth,E., Montor,W.R., Wong,J., Park,J.G. *et al.* (2011) Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *J. Proteome Res.*, **10**, 85–96.
- Wang,X., Yu,J., Sreekumar,A., Varambally,S., Shen,R., Giacherio,D., Mehra,R., Montie,J.E., Pienta,K.J., Sanda,M.G. *et al.* (2005) Autoantibody signatures in prostate cancer. *N. Engl. J. Med.*, **353**, 1224–1235.
- 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.

13. Henderson, M.C., Hollingsworth, A.B., Gordon, K., Silver, M., Mulpuri, R., Letsios, E. and Reese, D.E. (2016) Integration of serum protein biomarker and tumor associated autoantibody expression data increases the ability of a blood-based proteomic assay to identify breast cancer. *PLoS One*, **11**, e0157692.
14. Ayoglu, B., Mitsios, N., Kockum, I., Khademi, M., Zandian, A., Sjoberg, R., Forsstrom, B., Bredenberg, J., Lima Bomfim, I., Holmgren, E. *et al.* (2016) Anoctamin 2 identified as an autoimmune target in multiple sclerosis. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 2188–2193.
15. Jackson, H.J., Rafiq, S. and Brentjens, R.J. (2016) Driving CAR T-cells forward. *Nat. Rev. Clin. Oncol.*, **13**, 370–383.
16. Scott, A.M., Wolchok, J.D. and Old, L.J. (2012) Antibody therapy of cancer. *Nat. Rev. Cancer*, **12**, 278–287.
17. Ouyang, Q., Tian, C., Gao, J., Huang, J., Fu, H., He, J. and Yang, J. (2016) HER2-positive double primary tumor of gastric and breast cancer occur synchronously in a patient: A case report. *Mol. Clin. Oncol.*, **4**, 719–722.
18. 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.
19. 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.
20. Natale, D.A., Arighi, C.N., Blake, J.A., Bult, C.J., Christie, K.R., Cowart, J., D'Eustachio, P., Diehl, A.D., Drabkin, H.J., Helfer, O. *et al.* (2014) Protein Ontology: a controlled structured network of protein entities. *Nucleic Acids Res.*, **42**, D415–D421.
21. Natale, D.A., Arighi, C.N., Barker, W.C., Blake, J.A., Bult, C.J., Caudy, M., Drabkin, H.J., D'Eustachio, P., Evsikov, A.V., Huang, H. *et al.* (2011) The Protein Ontology: a structured representation of protein forms and complexes. *Nucleic Acids Res.*, **39**, D539–D545.
22. Kibbe, W.A., Arze, C., Felix, V., Mitra, E., Bolton, E., Fu, G., Mungall, C.J., Binder, J.X., Malone, J., Vasant, D. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.
23. Cohen, A.M. and Hersh, W.R. (2005) A survey of current work in biomedical text mining. *Brief. Bioinformatics*, **6**, 57–71.
24. Nunes, T., Campos, D., Matos, S. and Oliveira, J.L. (2013) BeCAS: biomedical concept recognition services and visualization. *Bioinformatics*, **29**, 1915–1916.
25. Sayers, E. (2008) *E-utilities Quick Start*. National Center for Biotechnology Information, Bethesda.
26. Gray, K.A., Yates, B., Seal, R.L., Wright, M.W. and Bruford, E.A. (2015) Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.*, **43**, D1079–D1085.
27. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.W., Mazaitis, M., Felix, V., Feng, G. and Kibbe, W.A. (2012) Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.*, **40**, D940–D946.
28. Jonquet, C., Shah, N. and Youn, C. (2009) *NCBO Annotator: Semantic Annotation of Biomedical Data*. International Semantic Web Conference, Poster and Demo session.
29. Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., Garcia Giron, C., Hourlier, T. *et al.* (2016) The Ensembl gene annotation system. *Database (Oxford)*, **2016**, baw093.
30. 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.
31. 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.
32. 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.
33. Alm, T., von Feilitzen, K., Lundberg, E., Sivertsson, A. and Uhlen, M. (2014) A chromosome-centric analysis of antibodies directed toward the human proteome using Antibodypedia. *J. Proteome Res.*, **13**, 1669–1676.
34. Katchman, B.A., Barderas, R., Alam, R., Chowell, D., Field, M.S., Esserman, L.J., Wallstrom, G., LaBaer, J., Cramer, D.W., Hollingsworth, M.A. *et al.* (2016) Proteomic mapping of p53 immunogenicity in pancreatic, ovarian, and breast cancers. *Proteomics Clin Appl.*, **10**, 720–731.
35. Yu, H., Wang, F., Lin, L., Cao, W., Liu, Y., Qin, L., Lu, H., He, F., Shen, H. and Yang, P. (2016) Mapping and analyzing the human liver proteome: progress and potential. *Expert Rev. Proteomics*, **13**, 1–11.
36. Bernard, W.S. and Christopher, P.W. (2014) *World Cancer Report 2014*. International Agency for Research on Cancer, World Health Organization, Lyon.
37. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S. *et al.* (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
38. Aravalli, R.N., Steer, C.J. and Cressman, E.N. (2008) Molecular mechanisms of hepatocellular carcinoma. *Hepatology*, **48**, 2047–2063.
39. Ongenaert, M., Van Neste, L., De Meyer, T., Menschaert, G., Bekaert, S. and Van Criekinge, W. (2008) PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res.*, **36**, D842–D846.
40. Noorden, R.V. (2012) Trouble at the text mine. *Nature*, **483**, 134–135.
41. Karthikeyan, K., Barker, K., Tang, Y., Kahn, P., Wiktor, P., Brunner, A., Knabben, V., Takulapalli, B., Buckner, J., Nepom, G. *et al.* (2016) A contra capture protein array platform for studying post-translationally modified auto-antigenomes. *Mol. Cell Proteomics*, **15**, 2324–2337.
42. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T. and Thomas, P.D. (2016) PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.*, **44**, D336–D342.