Data in Brief 57 (2024) 111052

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of Panda sperm proteome

Siying Liu[a], Tao Wang[a], Yuliang Liu[b,c], Shenfei Wang[b,c], Feiping Li[b,c], Jiasong Chen[b,c], Xianbiao Hu[b,c], Mengshi Zhang[b,c], Juan Wang[b,c], Yan Li[b,c], Ayala James[b,c], Rong Hou[b,c], Kailai Cai[b,c,*]

[a] *Antibiotics Research and Re-evaluation Key Laboratory of Sichuan Province, Sichuan Industrial Institute of Antibiotics, School of Pharmacy, Chengdu University, 610106 Chengdu, China*
[b] *Chengdu Research Base of Giant Panda Breeding, Chengdu 610081, Sichuan, China*
[c] *Sichuan Key Laboratory of Conservation Biology for Endangered Wildlife, Chengdu 610081, Sichuan, China*

ARTICLE INFO

Article history:
Received 20 June 2024
Revised 18 September 2024
Accepted 14 October 2024
Available online 21 October 2024

Dataset link: Proteomic analysis of giant panda sperm (Original data)

Keywords:
Giant panda
Sperm
Proteome
Serpin
Ex-situ conservation

ABSTRACT

The giant panda (*Ailuropoda melanoleuca*), a vulnerable species and an icon of wildlife conservation, is still at risk due to habitat fragmentation and a low reproductive rate. To further safeguard the giant panda from extinction, a captive breeding program was established in the mid 1980's, however the growth of this population has been hindered by the poor reproductive ability of captive male giant pandas. To address this, we investigated the sperm proteome of the giant panda as detailed information on sperm proteome is unavailable, as it is a highly specialized area of study. A study of adult panda sperm proteome identified 1921 proteins with enriched domains, including EF-hand, AAA+ ATPase, and WD 40 repeat. A comparison with four other species revealed common sperm proteins related to metabolic processes, especially glycolysis and citrate cycle, which are crucial for sperm energy. Panda-specific proteins were mainly associated with cellular protein metabolism. Serpin domain-related proteins may play a key role in panda semen properties and liquefaction, involving prefoldin beta-like and heat shock

* Corresponding author at: Chengdu Research Base of Giant Panda Breeding, Chengdu 610081, Sichuan, China.
E-mail address: caikailai@panda.org.cn (K. Cai).

https://doi.org/10.1016/j.dib.2024.111052

chaperonin-binding. This research contributes to understanding giant panda reproduction and aids in conservation efforts.

Specifications Table

Subject	Biological Sciences
Specific subject area	Proteomics
Data format	Raw datasheet
Type of data	Raw and Filtered
Data collection	The six giant pandas' sperm samples were collected by electroejaculation, centrifuged at 900 × *g* for 30 min, collected sperm cells, transported to 1.5 mL tubes, and frozen at −80 °C. Sonicated the samples, add a lysis buffer, centrifuge, collect the supernatant, reduce and alkylated before being mixed with acetone and incubated. Then, it was centrifuged again, and the resulting precipitation was collected and washed. The pellet was dissolved in a dissolution buffer, and the protein concentration was measured using the Bradford protein assay. The supernatant was digested with Trypsin Gold, and the resulting peptides were desalted and dried before being analyzed using shotgun proteomics. The resulting spectra were searched against a specific database using Proteome Discoverer 2.2. Protein identification was performed using a false discovery rate (FDR) of less than 1.0 %, and precursor quantification was used for label-free quantification. Gene Ontology (GO) and Inter Pro (IPR) analysis were conducted using the InterProScan-5 program (EMBL-EBI) against the non-redundant protein database. In contrast, COG and KEGG were used to analyze protein families and pathways. Finally, the STRING-db server predicted probable interacting partners based on related species. Using an enrichment pipeline, GO, IPR, and KEGG enrichment analyses were performed.
Data source location	Institution: Chengdu Giant Panda Breeding Base City: Chengdu Country: China
Data accessibility	The mass spectrometry proteomics data have been deposited to the Proteome Xchange Consortium via the iPro X partner repository. (URL:http://proteomecentral.proteomexchange.org; Project ID: PXD051077)

1. Value of the Data

- These data provide comprehensive and standardized Sperm protein enrichment domains and protein identification results.
- The dataset also includes the interactions between proteins within a biological system and reveals the function of the unique Panda protein.
- Other researchers can use these data to investigate protein interactions in spermatids further.
- The samples used for this study were obtained during routine semen collection and were conducted through the cryogenic storage processes

2. Background

The giant panda (*Ailuropoda melanoleuca*) is a charismatic species recognized globally as a symbol of conservation. However, the panda's survival is still threatened by various factors, including habitat loss and low reproductive rates. To safeguard the giant panda from extinction, a captive breeding program was established in the mid 1980's. While considered a success, the growth of this *ex-situ* population has been hindered by the poor reproductive ability of cap-

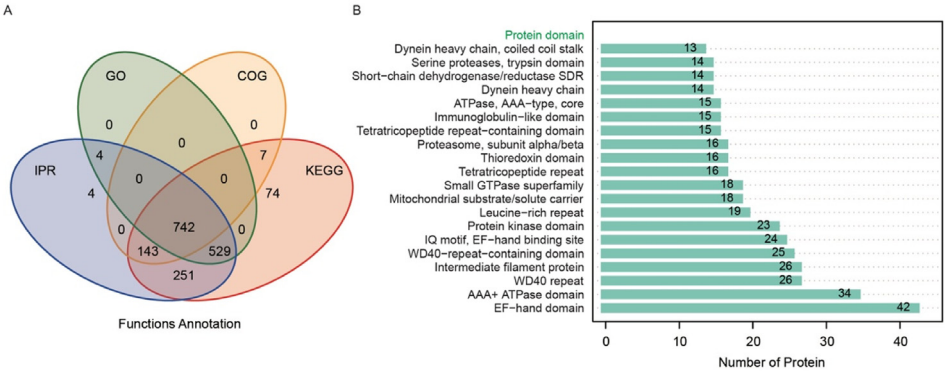

Fig. 1. The Venn diagram displays the annotation results of the identified proteins in four databases (A) and shows IPR annotations (B). The bar represents the number of proteins.

tive male giant pandas. As the giant panda is particularly vulnerable to stochastic processes, a comprehensive understanding of their genetic health status is crucial for the continued conservation of this vulnerable species [1]. One of the critical aspects of understanding and improving panda health and reproduction in captivity is the study of their sperm proteome. Sperm cells are specialized reproductive cells that play a crucial role in fertilization. The proteome of sperm encompasses a diverse array of proteins that are involved in various cellular processes, including sperm maturation, motility, capacitation [2,3], and interaction with the female reproductive tract [4]. The sperm proteome of several species have previously been studied, including mice [5], zebra finches [6], oysters [7], horses [8], and humans [9]), using techniques such as mass spectrometry to identify and quantify the proteins with previous research focusing on issues related to sperm development, quality, and infertility, though deep comprehensive, former research is lack of system and depth, in-depth study of sperm protein group helps further to study the relationship between sperm and animal breeding, look for ways to improve animal reproduction rate, it is especially important for endangered species.

3. Data Description

A label-free shotgun proteomic approach was used to identify 1921 proteins in the sperm cells. Functional annotations were performed on these proteins using several databases, including GO, IPR, KEGG, and IPR. The majority of the proteins were successfully annotated, as shown in Fig. 1A. The results of the IPR annotation show these proteins mainly enriched EF-hand domain, AAA+ ATPase domain, and WD 40 repeat domain (Fig. 1B).

Gene names corresponding to proteins were used consistently across different species to facilitate comparative analysis. By comparing the seminal plasma proteins identified in pandas with those of human [10], pig[11], sheep[12], and horse [13], it was found that giant pandas share 43 genes (corresponding to 43 proteins) with these species, while also having 596 unique genes. Functional enrichment analysis revealed that the shared proteins were enriched in the metabolic process and single organism process/metabolic process (Fig. 2A), meanwhile primarily participating in carbon metabolism, glycolysis/gluconeogenesis, the citrate cycle (TCA cycle), and the glucagon signaling pathway (Fig. 2B). The interactions between proteins within a biological system were studied using PPI analysis. This analysis revealed several node proteins Fig. 2C, including l-lactate dehydrogenase (D2HU89), citrate synthase (G1LAU6), alpha-enolase (G1LB18), pyruvate dehydrogenase E1 subunit beta (G1M029), chaperonin containing TCP1 subunit 2 (G1LZI4), glucose-6-phosphate isomerase (D2HRQ7), malate dehydrogenase (G1LWG5), and phosphoglycerate kinase (D2H1P3 (Fig. 2C). These proteins are involved in various metabolic processes such as glycolysis, the citric acid cycle, and protein folding.

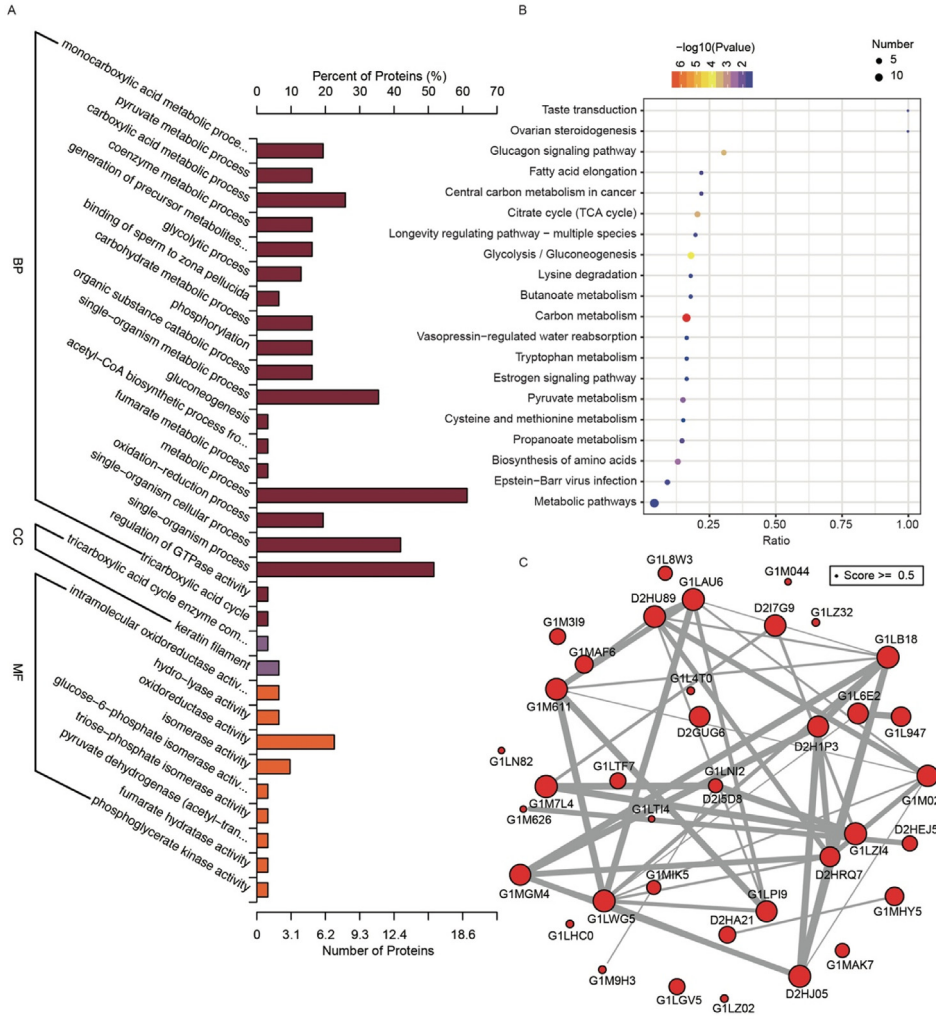

Fig. 2. The GO (A) and Pathway (B) analysis of common sperm cell proteins in giant pandas and other four species, then the PPI interactions were analyzed (C).

In the comparative analysis, it was observed that the majority of proteins were specific to giant pandas. This could be due to the unclear annotation of proteins in this species. Functional analysis was performed on these unique panda proteins, revealing their involvement in binding functions. These proteins are associated with cellular protein metabolic processes, cellular macromolecule metabolic processes, and biosynthetic processes. These biological processes primarily occur within intracellular organelles (Fig. 3A). We further found that these proteins are the serpin family or contain serpin domain through IPR analysis, prefoldin beta-like, and heat shock chaperonin-binding proteins exhibit more prominent features (Fig. 3B).

Overall, the giant panda sperm proteome is particularly important in the context of assisted reproductive techniques and ex-situ conservation efforts. It helps in identifying potential biomarkers for sperm quality and fertility, as well as understanding the impact of various factors on sperm function and viability. It's important to note that the information provided is based on the available research and may not represent the most current findings. Further studies and re-

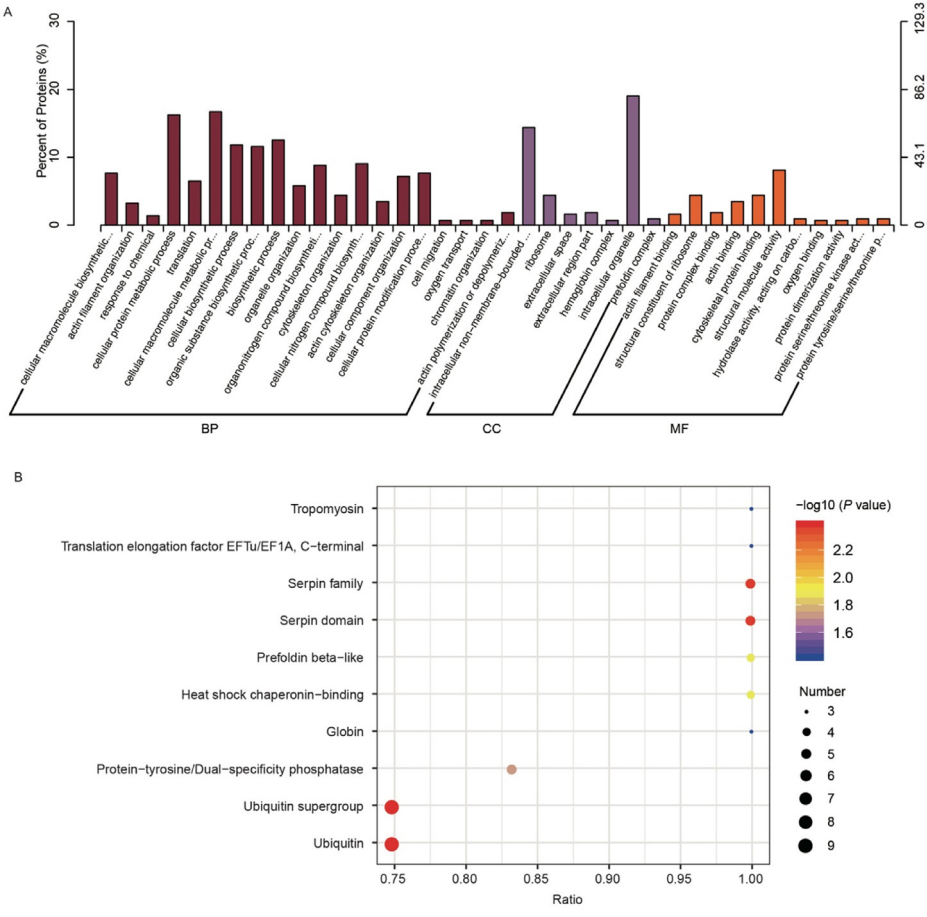

Fig. 3. The GO (A) and Pathway (B) analysis of pandas-specific sperm cell proteins compared to the other four species.

search are necessary to gain a more comprehensive understanding of the factors affecting giant panda fertility and develop strategies to improve their reproductive success in captivity.

4. Experimental Design, Materials and Methods

To address the lack of knowledge of the giant panda sperm proteome, the present study aimed to identify whole proteome profiles of the giant panda seminal plasma using a gel-free, label-free shotgun proteomics approach. Sperm cells from six sexually mature giant pandas (aged between 9 and 16 years) were simultaneously collected by electroejaculation during the giant panda breeding season according to previous methodology, and the six giant pandas' sperm samples were mixed into a sample [14]. Semen was collected into a plastic container, and centrifugation was immediately performed; an aliquot of 0.5 mL of fresh semen was centrifuged at 900 $\times$ *g* for 30 min at 4 °C, seminal plasma was discarded, and sperm cells were transported to 1.5 mL tubes and frozen at −80 °C until further use. The samples used for this study were obtained during routine semen collection for artificial insemination, and the cryogenic storage processes were conducted at the Sichuan Key Laboratory of Conservation Biology for Endangered

Wildlife, Chengdu Research Base of Giant Panda Breeding. Specifically, an aliquot of stored sperm intended for insemination was selected for metabolomics investigation.

Sonication was used to break samples down, and a lysis buffer was added to prepare them for analysis. The lysate was then centrifuged and the supernatant was transferred to a clean tube. Next, the extracts were reduced and alkylated before being mixed with acetone and incubated. The samples were then centrifuged again and the resulting precipitation was collected and washed. Afterward, the pellet was dissolved in a dissolution buffer and the protein concentration was measured using the Bradford protein assay. The supernatant from each sample was digested with Trypsin Gold and the resulting peptides were desalted and dried before being analyzed using shotgun proteomics. This involved injecting the peptides into a column and separating them using a linear gradient. The resulting spectra were then searched against a specific database using Proteome Discoverer 2.2 (PD 2.2, Thermo Fisher Scientific-CN).

Protein identification was performed using a false discovery rate (FDR) of less than 1.0 %, and precursor quantification was used for label-free quantification. Gene Ontology (GO) and InterPro (IPR) analysis were conducted using the InterProScan-5 program (EMBL-EBI) against the non-redundant protein database (including Pfam, PRINTS, ProDom, SMART, ProSiteProfiles, PANTHER) [15]. In contrast, COG and KEGG were used to analyze protein families and pathways. Finally, the STRING-db server predicted probable interacting partners based on related species [16]. GO, IPR and KEGG enrichment analyses were performed using an enrichment pipeline [17].

Limitations

Due to the rarity and special status of panda species, this project concerns biological sample collection; a mix of six individuals provides enough sperm samples to test the protein group, but it lacks biological reproducibility.

Ethics Statement

The Institutional Animal Experimental Ethics Committee of the Chengdu Research Base of Giant Panda Breeding (Approval no CRBGPB2018008).

CRediT Author Statement

Siying Liu: Writing, Original draft preparation, Visualization. **Tao Wang:** Analyzed the data, Writing-Reviewing and Editing, Project administration. **Yuliang Liu:** Methodology, Conceptualization. **Shenfei Wang:** Methodology, Investigation. **Feiping Li:** Methodology, Formal analysis. **Jiasong Chen**: Validation, Methodology. **Xianbiao Hu:** Conceptualization, Formal analysis. **Mengshi Zhang**: Validation, Formal analysis. **Juan Wang:** Data curation. **Yan Li:** Data curation. **Ayala James**: Methodology. **Rong Hou:** Data curation, Visualization. **Kailai Cai:** Methodology, Writing-Reviewing and Editing, Funding acquisition.

Data Availability

Proteomic analysis of giant panda sperm (Original data) (iPro X).

Acknowledgments

Funding: This work was supported by the Chengdu Giant Panda Breeding Research Foundation (CPF2017-17).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Li, C. Zhou, M. Cheng, H. Tu, G. Wang, Y. Mao, Y. Huang, M. Chen, M. Price, Y. Meng, B. Yue, X. Zhang, Large-scale genetic surveys for main extant population of wild giant panda (Ailuropoda melanoleuca) reveals an urgent need of human management, Evol. Appl. 16 (3) (2023) 738–749, doi:10.1111/eva.13532.

[2] G. Wang, Y. Guo, T. Zhou, X. Shi, J. Yu, Y. Yang, Y. Wu, J. Wang, M. Liu, X. Chen, W. Tu, Y. Zeng, M. Jiang, S. Li, P. Zhang, Q. Zhou, B. Zheng, C. Yu, Z. Zhou, X. Guo, J. Sha, In-depth proteomic analysis of the human sperm reveals complex protein compositions, J. Proteom. 79 (2013) 114–122, doi:10.1016/j.jprot.2012.12.008.

[3] C.J. Li, D. Wang, X. Zhou, Sperm proteome and reproductive technologies in mammals, Anim. Reprod. Sci. 173 (2016) 1–7, doi:10.1016/j.anireprosci.2016.08.008.

[4] C. Luongo, L. Gonzalez-Brusi, P. Cots-Rodriguez, M.J. Izquierdo-Rico, M. Aviles, F.A. Garcia-Vazquez, Sperm proteome after interaction with reproductive fluids in porcine: from the ejaculation to the fertilization site, Int. J. Mol. Sci. 21 (17) (2020) 6060, doi:10.3390/ijms21176060.

[5] D.A. Skerrett-Byrne, A.L. Anderson, E.G. Bromfield, I.R. Bernstein, J.E. Mulhall, J.E. Schjenken, M.D. Dun, S.J. Humphrey, B. Nixon, Global profiling of the proteomic changes associated with the post-testicular maturation of mouse spermatozoa, Cell Rep. 41 (7) (2022) 111655, doi:10.1016/j.celrep.2022.111655.

[6] M. Rowe, S. Skerget, M.A. Rosenow, T.L. Karr, Identification and characterization of the zebra finch (Taeniopygia guttata) sperm proteome, J. Proteomics 193 (2019) 192–204, doi:10.1016/j.jprot.2018.10.009.

[7] H. Mu, S. Ke, D. Zhang, Y. Zhang, X. Song, Z. Yu, Y. Zhang, J.W. Qiu, The sperm proteome of the oyster Crassostrea hongkongensis, Proteomics 20 (19–20) (2020) e2000167, doi:10.1002/pmic.202000167.

[8] R.A. Griffin, A. Swegen, M. Baker, R.J. Aitken, D.A. Skerrett-Byrne, A.S. Rodriguez, F.E. Martin-Cano, B. Nixon, F.J. Pena, M. Delehedde, N. Sergeant, Z. Gibb, Mass spectrometry reveals distinct proteomic profiles in high- and low-quality stallion spermatozoa, Reproduction 160 (5) (2020) 695–707, doi:10.1530/REP-20-0284.

[9] T. Greither, M. Dejung, H.M. Behre, F. Butter, H. Herlyn, The human sperm proteome-Toward a panel for male fertility testing, Andrology 11 (7) (2023) 1418–1436, doi:10.1111/andr.13431.

[10] M. Saraswat, S. Joenvaara, T. Jain, A.K. Tomar, A. Sinha, S. Singh, S. Yadav, R. Renkonen, Human spermatozoa quantitative proteomic signature classifies normo- and asthenozoospermia, Mol. Cell Proteom. 16 (1) (2017) 57–72, doi:10.1074/mcp.M116.061028.

[11] C. Perez-Patino, I. Barranco, J. Li, L. Padilla, E.A. Martinez, H. Rodriguez-Martinez, J. Roca, I. Parrilla, Cryopreservation differentially alters the proteome of epididymal and ejaculated pig spermatozoa, Int. J. Mol. Sci. 20 (7) (2019) 1791, doi:10.3390/ijms20071791.

[12] W. Zhu, X. Cheng, C. Ren, J. Chen, Y. Zhang, Y. Chen, X. Jia, S. Wang, Z. Sun, R. Zhang, Z. Zhang, Proteomic characterization and comparison of ram (Ovis aries) and buck (Capra hircus) spermatozoa proteome using a data independent acquisition mass spectometry (DIA-MS) approach, PLoS One 15 (2) (2020) e0228656, doi:10.1371/journal.pone.0228656.

[13] F.E. Martin-Cano, G. Gaitskell-Phillips, J.M. Ortiz-Rodriguez, A. Silva, C. Gil, C. Ortega-Ferrusola, F.J. Pena, Data set of the proteome of fresh and frozen thawed stallion spermatozoa, Data Br. 31 (2020) 105887, doi:10.1016/j.dib.2020.105887.

[14] Z.G. Cai, J.H. An, Y.L. Liu, S.M. Yie, Y. Zhang, F.P. Li, J.S. Chen, X. Wang, J.M. Morrell, R. Hou, Single layer centrifugation improves the quality of frozen-thawed sperm of giant panda (Ailuropoda melanoleuca), Anim. Reprod. Sci. 195 (2018) 58–64, doi:10.1016/j.anireprosci.2018.05.006.

[15] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification, Bioinformatics 30 (9) (2014) 1236–1240, doi:10.1093/bioinformatics/btu031.

[16] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, L.J. Jensen, STRING v9.1: protein-protein interaction networks, with increased coverage and integration, Nucleic Acids Res. 41 (2013) D808–D815 (Database issue), doi:10.1093/nar/gks1094.

[17] W. Huang da, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic Acids Res. 37 (1) (2009) 1–13, doi:10.1093/nar/gkn923.