D588–D592 *Nucleic Acids Research, 2015, Vol. 43, Database issue* *Published online 12 September 2014*
doi: 10.1093/nar/gku830

VirusMentha: a new resource for virus-host protein interactions

Alberto Calderone[1,*], Luana Licata[1,*] and Gianni Cesareni[1,2,*]

[1]Department of Biology, University of Rome Tor Vergata, Rome, Italy and [2]Fondazione Santa Lucia Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Rome, Italy

Received July 31, 2014; Revised August 31, 2014; Accepted September 01, 2014

ABSTRACT

Viral infections often cause diseases by perturbing several cellular processes in the infected host. Viral proteins target host proteins and either form new complexes or modulate the formation of functional host complexes. Describing and understanding the perturbation of the host interactome following viral infection is essential for basic virology and for the development of antiviral therapies. In order to provide a general overview of such interactions, a few years ago we developed VirusMINT. We have now extended the scope and coverage of VirusMINT and established VirusMentha, a new virus–virus and virus–host interaction resource build on the detailed curation protocols of the IMEx consortium and on the integration strategies developed for *mentha*. VirusMentha is regularly and automatically updated every week by capturing, via the PSICQUIC protocol, interactions curated by five different databases that are part of the IMEx consortium. VirusMentha can be freely browsed at http://virusmentha.uniroma2.it/ and its complete data set is available for download.

INTRODUCTION

Viruses exploit the molecular machinery of the infected host to support their own life cycle and target host defense mechanisms to escape host resistance. This is achieved by establishing a virus-specific protein interaction network that perturbs cell processes, such as DNA replication, gene expression, growth and differentiation (1,2).

A prerequisite for a complete understanding of a virus life cycle is a proteome-wide description of the complexes formed by viral proteins, either on their own or in combination with host proteins. Over the past decades, a growing number of small- or large-scale studies reported evidence for interactions between viral and host proteins. Although this information can in principle be extracted from generic protein interaction databases, 5 years ago we developed VirusMINT (3), a protein-protein interaction (PPI) database focused on viral interactions. This offered two advantages: an increased focus of the MINT (4) curation team on the curation of viral interactions and the development of a dedicated website that could archive information useful for the community of experimental virologists. VirusMINT only offered data curated by the MINT curation team.

VirHostNet (5), a second resource dedicated to virus interactions, adopted the strategy of merging various data sources. This database contains re-curated data extracted from primary resources that use different curation strategies. In addition, VirHostNet complements this information with newly curated interactions. This approach has generated a network consisting of nearly 5000 non-redundant interactions between viral and host proteins. However, the integration strategy adopted by this resource implies substantial work making updates infrequent. As a consequence, the VirHostNet data set has not been updated since its publication.

As technologies to identify protein interactions evolve, generic or specialized journals regularly report new viral host protein interaction information (6,7). This abundance of data demands for a strategy to capture and integrate this information regularly, with minimal hands-on effort. In order to automate data merging, we recently implemented a resource called *mentha* that automatically integrates the content of five different PPI databases (8). By exploiting a similar automatic procedure we created VirusMentha, a resource that builds on the work started in VirusMINT by offering the curated data from our group combined with the data captured from the other IMEx (9) partners. Every week, VirusMentha integrates the virus–host interactions curated by the IMEx databases removing redundancy. Differently from other valuable resources, VirusMentha is not exclusively focused on specific organisms or viruses, such as HCVpro (http://cbrc.kaust.edu.sa/hcvpro/) (10) or HIV host-interaction map (http://www.ncbi.nlm.nih.gov/projects/RefSeq/HIVInteractions/) (11). VirusMentha captures all published virus–host interactions without limita-

*To whom correspondence should be addressed. Tel: +39 0672594315; Fax: +39 062023500; Email: sinnefa@gmail.com
Correspondence may also be addressed to Luana Licata. Tel: +39 0672594315; Fax: +39 062023500; Email: luana.licata@gmail.com
Correspondence may also be addressed to Gianni Cesareni. Tel: +39 0672594315; Fax: +39 062023500; Email: cesareni@uniroma2.it

tion with respect to virus strain or to host organism. VirusMentha currently contains interactions for 24 viral families. It also offers interactions between viral proteins and different host organisms, such as *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana* and so on.

IMPLEMENTATION

Molecular interaction evidence is reported in the scientific literature in natural language format, thus making retrieval and processing a difficult task (12,13). The data contained in VirusMINT is manually curated by our curation team in compliance with IMEx standards. In order to further expand the coverage of the virus–virus and virus–host interactome, we decided to integrate our data with data curated and stored in the databases of the IMEx partners. IMEx is an international consortium whose purpose is to make protein interaction curation more efficient by distributing the curation load and by limiting the duplication of effort. IMEx standards represent a seal of quality among the several PPIs repositories (14).

The databases that we chose to integrate are MINT, IntAct (15), DIP (16), MatrixDB (17) and BioGRID (18). VirHostNet data were not imported because the database does not fully conform to PSI-MI standards and does not provide enough experimental details.

The procedure used for data merging is the same as the one developed for *mentha*. *mentha* retrieves data from external resources using the PSICQUIC (Proteomics Standard Initiative Common QUery InterfaCe) protocol (19) which provides standard programmatic access to molecular interaction repositories. Using the ontology trees defined by IMEx, the procedure identifies whether two pieces of evidence are actually the same piece of evidence described at different levels of curation depth. VirusMentha exploits the same procedure and offers direct access to evidence of viral protein interactions.

The *mentha* procedure detects redundancy using the PSI-MI TAB file provided by the five source databases listed above. The key information to identify redundancy are supported by every database of the IMEx consortium, i.e. interaction type, experimental method, publication ID. However, these databases do not offer the same version of PSI-MI TAB file. MINT and IntAct are currently the two major databases in the IMEx consortium that offer PSI-MI TAB 2.7 format defined in PSICQUIC. The PSI-MI TAB file format has been updated to version 2.7 to allow curators to add extra information, such as for instance the polypeptide fragment (of viral polyproteins) involved in the interaction. In order to preserve the richness of such format, VirusMentha implements a detailed view if the evidence is curated by databases supporting PSI-MI TAB 2.7.

TAXONOMY ORGANIZATION

VirusMentha organizes the interactions according to viral families (Figure 1). The classification is based on the Baltimore classification which is based on the nucleic acid content of the virion (20,21). This classification is generally used in conjunction with the ICTV (The International Committee on Taxonomy of Viruses) classification system in order to create a tree of viral species (22).

In order to give access to this hierarchy of viral strains and families, we have reconstructed the taxonomy tree of viruses using the classification information provided by UniProt (23). In this taxonomy tree, as show in Figure 1, the first level is defined as per the Baltimore classification while families are defined according to the ICTV classification system. In order to handle this taxonomic hierarchy, the retrieval and integration procedure of VirusMentha analyses each interaction and for each viral protein in the interaction it extracts its taxonomy ID (organism). The mapping procedure is illustrated in Figure 1. Using the taxonomy tree, each virus strain (taxonomy number) is mapped to the most general viral family, without dropping the original strain identifier.

DATA GROWTH AND STATISTICS

Over the past years, the MINT curation team has curated a large number of articles using a detailed curation policy. The integration of VirusMINT data with other databases on a weekly basis resulted in the most comprehensive and self-updating resource available so far. On August 2014, VirusMentha contained up to 8084 non-redundant interactions supported by 8450 publications (Figure 2).

WEB INTERFACE

Searches in the VirusMentha data set can be carried out directly from the home page by entering in the search field UniProt IDs, gene names, polypeptide names, keywords or single PMID (PubMed publication ID). To offer direct access to the VirusMentha data set we implemented the possibility of searching and downloading viral interaction evidence from a virus or host perspective. The user can decide to search the entire database or to restrict the search to specific organisms or viral families.

The search results are presented as a list of proteins whose gene name, UniProt ID, description or PMID matches the query. The user can select the proteins of interest by adding them to the 'protein bag'. All the interactions related to the selected proteins can be browsed either as a list in a table format or in a network view by starting a graphical applet. The user who wants to compare the same protein from different isolates can use the 'Align' button, to see their similarities; the alignment is performed on the fly using BLOSUM62 as substitution matrix, with the following penalties: open: 10; extend: 0.5. Experimental details linked to each interaction can be accessed by clicking the 'show evidence' button. Together with the link to the original paper, the interaction detection method and the interaction type of the experiment are displayed. Moreover, by clicking to the magnifier icon, a pop-up window opens showing the curation details as per PSI-MI TAB 2.7. In the binary interaction result page it is possible to filter relevant interactions using keywords. For instance, the user can search a specific human protein that interacts with several types of Herpes simplex virus strains and, using the field provided, restrict the list only to the 'Patton' strain.

The VirusMentha interface has been developed to facilitate the assembly and modification of networks of interactions. The graphical applet displays viral proteins in cyan.

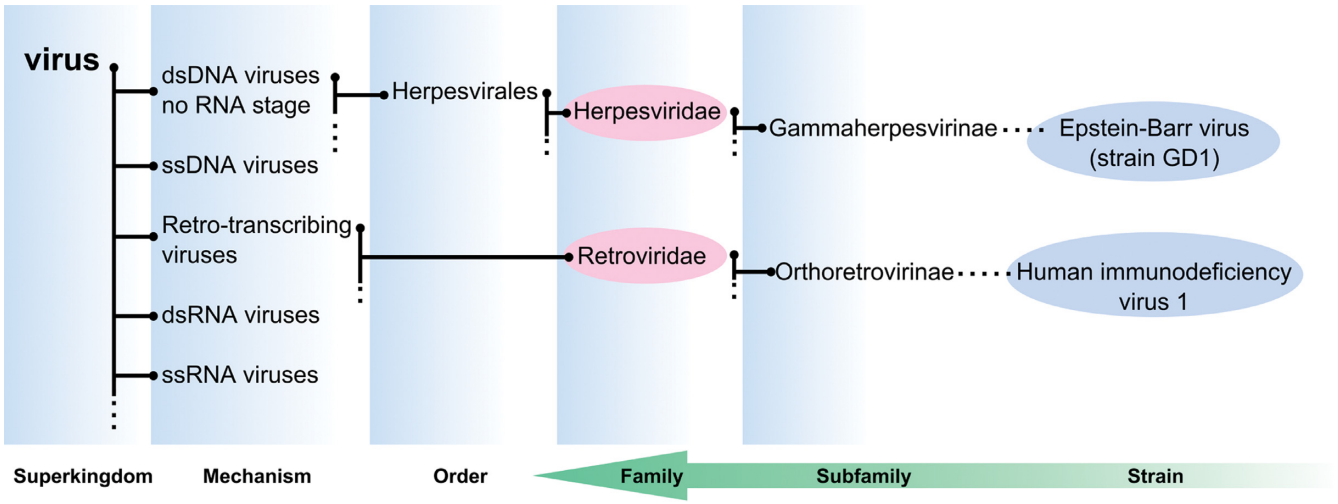

Figure 1. Mapping strategy. This figure shows the strategy used by VirusMentha to group all viruses into families. On the right, the diagram shows the strain annotated in the experimental evidence. The taxonomy tree is used to map back to the viral family, highlighted by red ovals.

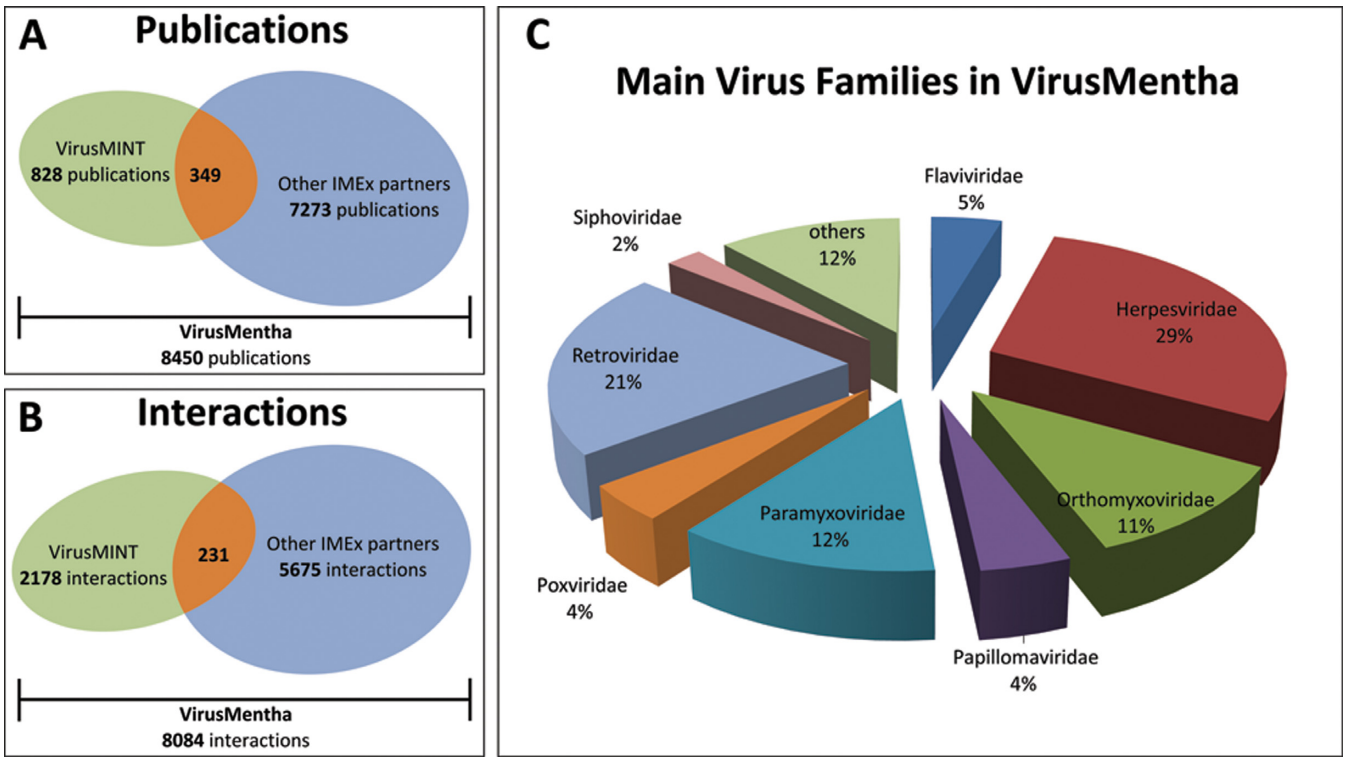

Figure 2. Comparison of VirusMINT and VirusMentha. (A and B) Venn diagrams to show the relative contribution to the VirusMentha data set of the curation effort of VirusMINT and that of the others IMEx databases. (B) The pie chart represents the number of interactions of the main virus families as annotated in VirusMentha.

The networks built by the user through various searches and/or network manipulations can be downloaded as a plain text file. Exported networks always report strain number and family number to facilitate clustering by family.

Finally, for each protein VirusMentha reports the associated Gene Ontology terms (24), links to antibodies through Antibodypedia (25), cleaved polypeptides and, for each interaction, the MINT score and links to papers supporting the interaction are reported (Figure 3).

CONCLUSION AND PERSPECTIVES

VirusMint was developed as a public repository to capture and organize manually curated information about interactions between viral and host proteins. VirusMentha extends this concept to offer the most comprehensive and

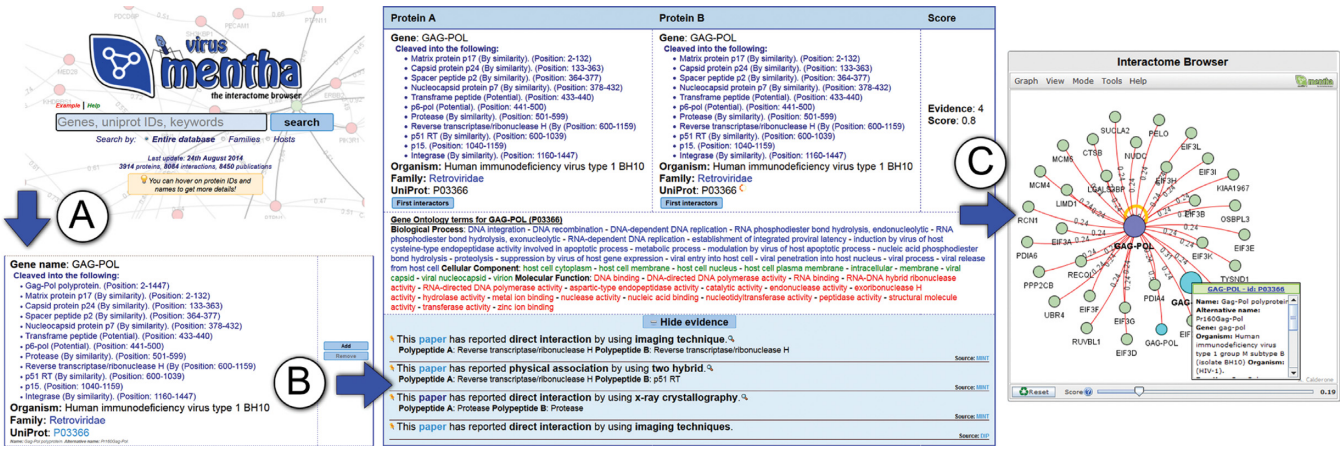

Figure 3. VirusMentha web interface. (A) VirusMentha search page. By typing the gene name GAGPOL the user is presented with a list a GAG-POL genes from several HIV isolates. Each gene is accompanied with essential information extracted from UniProt including the polypeptide fragments processed inside the virion. (B) If the user loads one of these proteins into the 'protein bag' and clicks 'list', all binary interactions are presented in a table-like view, together with details about papers supporting the interaction. (C) Alternatively, the collected proteins can be browsed by displaying the interactions with a graphic applet.

self-updating resource for viral interactions available so far. In addition, VirusMentha exploits the curation potential of five different database curation teams by integrating data curated by databases adhering to the IMEx consortium. In addition, VirusMentha continues to benefit from the VirusMINT curation team that regularly captures data published in peer-reviewed scientific journals. Finally, VirusMentha has the advantage of being regularly and automatically updated every week.

According to Google Scholar, VirusMINT has been cited in ∼100 papers (99 in August 2014) testifying the interest of the community of virologists for such a dedicated protein interaction repository. As more relevant information is reported in the scientific literature, we expect the higher coverage and curation depth that characterize VirusMentha to become a valuable resource for more comprehensive analysis of viral mechanisms and interactions.

FUNDING

Oncodiet project of the Italian Ministry of Education [RBAP11LP2W to G.C.]; Italian Association for Cancer Research [N IG 2013 N.14135 to G.C.]. Funding for open access charge: Oncodiet project of the Italian Ministry of Education [RBAP11LP2W to G.C.]; Italian Association for Cancer Research [N IG 2013 N.14135 to G.C.].
Conflict of interest statement. None declared.

REFERENCES

1. Chisari,F.V (2005) Unscrambling hepatitis C virus-host interactions. *Nature*, **436**, 930–932.
2. Lanford,R.E., Guerra,B., Lee,H., Averett,D.R., Pfeiffer,B., Chavez,D., Notvall,L. and Bigger,C. (2003) Antiviral effect and virus-host interactions in response to alpha interferon, gamma interferon, poly(i)-poly(c), tumor necrosis factor alpha, and ribavirin in hepatitis C virus subgenomic replicons. *J. Virol.*, **77**, 1092–1104.
3. Chatr-aryamontri,A., Ceol,A., Peluso,D., Nardozza,A., Panni,S., Sacco,F., Tinti,M., Smolyar,A., Castagnoli,L., Vidal,M. *et al.* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res.*, **37**, D669–D673.
4. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
5. Navratil,V., de Chassey,B., Meyniel,L., Delmotte,S., Gautier,C., André,P., Lotteau,V. and Rabourdin-Combe,C. (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.*, **37**, D661–D668.
6. Mendez-Rios,J. and Uetz,P. (2010) Global approaches to study protein–protein interactions among viruses and hosts. *Future Microbiol.*, **5**, 289–301.
7. Rozenblatt-Rosen,O., Deo,R.C., Padi,M., Adelmant,G., Calderwood,M.A., Rolland,T., Grace,M., Dricot,A., Askenazi,M., Tavares,M. *et al.* (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature*, **487**, 491–495.
8. Calderone,A., Castagnoli,L. and Cesareni,G. (2013) Mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods*, **10**, 690–691.
9. Orchard,S., Kerrien,S., Abbani,S., Aranda,B., Bhate,J., Bidwell,S., Bridge,A., Briganti,L., Brinkman,F.S.L., Cesareni,G. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.
10. Kwofie,S.K., Schaefer,U., Sundararajan,V.S., Bajic,V.B. and Christoffels,A. (2011) HCVpro: hepatitis C virus protein interaction database. *Infect. Genet. Evol.*, **11**, 1971–1977.
11. Pinney,J.W., Dickerson,J.E., Fu,W., Sanders-Beer,B.E., Ptak,R.G. and Robertson,D.L. (2009) HIV-host interactions: a map of viral perturbation of the host system. *AIDS*, **23**, 549–554.
12. Lehne,B. and Schlitt,T. (2009) Protein-protein interaction databases: keeping up with growing interactomes. *Hum. Genom.*, **3**, 291–297.
13. Leitner,F., Chatr-aryamontri,A., Mardis,S.A., Ceol,A., Krallinger,M., Licata,L., Hirschman,L., Cesareni,G. and Valencia,A. (2010) The FEBS Letters/BioCreative II.5 experiment: making biological information accessible. *Nat. Biotechnol.*, **28**, 897–899.
14. Mosca,R., Pons,T., Céol,A., Valencia,A. and Aloy,P. (2013) Towards a detailed atlas of protein-protein interactions. *Curr. Opin. Struct. Biol.*, **23**, 929–940.
15. Kerrien,S., Aranda,B., Breuza,L., Bridge,A., Broackes-Carter,F., Chen,C., Duesbury,M., Dumousseau,M., Feuermann,M., Hinz,U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, **40**, D841–D846.
16. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.

17. Chautard,E., Fatoux-Ardore,M., Ballut,L., Thierry-Mieg,N. and Ricard-Blum,S. (2011) MatrixDB, the extracellular matrix interaction database. *Nucleic Acids Res.*, **39**, D235–D240.
18. Chatr-Aryamontri,A., Breitkreutz,B.-J., Heinicke,S., Boucher,L., Winter,A., Stark,C., Nixon,J., Ramage,L., Kolas,N., O'Donnell,L. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D816–D823.
19. Aranda,B., Blankenburg,H., Kerrien,S., Brinkman,F.S.L., Ceol,A., Chautard,E., Dana,J.M., De Las Rivas,J., Dumousseau,M., Galeota,E. *et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods*, **8**, 528–529.
20. Baltimore,D. (1974) The strategy of RNA viruses. *Harvey Lect.*, **70**, 57–74.
21. Temin,H.M. and Baltimore,D. (1972) RNA-directed DNA synthesis and RNA tumor viruses. *Adv. Virus Res.*, **17**, 129–186.
22. Andrew,M.Q.K., Elliot,L., Michael,J.A. and Eric,B.C. (2011) *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier, Amsterdam.
23. The UniProt Consortium. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
24. Blake,J.A., Dolan,M., Drabkin,H., Hill,D.P., Li,N., Sitnikov,D., Bridges,S., Burgess,S., Buza,T., McCarthy,F. *et al.* (2013) Gene Ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
25. Björling,E. and Uhlén,M. (2008) Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol. Cell. Proteom.*, **7**, 2028–2037.