CORUM: the comprehensive resource of mammalian protein complexes—2019

Madalina Giurgiu, Julian Reinhard, Barbara Brauner, Irmtraud Dunger-Kaltenbach, Gisela Fobo, Goar Frishman, Corinna Montrone and Andreas Ruepp ⓘ*

Institute for Bioinformatics and Systems Biology (IBIS), Helmholtz Zentrum München—German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

ABSTRACT

CORUM is a database that provides a manually curated repository of experimentally characterized protein complexes from mammalian organisms, mainly human (67%), mouse (15%) and rat (10%). Given the vital functions of these macromolecular machines, their identification and functional characterization is foundational to our understanding of normal and disease biology. The new CORUM 3.0 release encompasses 4274 protein complexes offering the largest and most comprehensive publicly available dataset of mammalian protein complexes. The CORUM dataset is built from 4473 different genes, representing 22% of the protein coding genes in humans. Protein complexes are described by a protein complex name, subunit composition, cellular functions as well as the literature references. Information about stoichiometry of subunits depends on availability of experimental data. Recent developments include a graphical tool displaying known interactions between subunits. This allows the prediction of structural interconnections within protein complexes of unknown structure. In addition, we present a set of 58 protein complexes with alternatively spliced subunits. Those were found to affect cellular functions such as regulation of apoptotic activity, protein complex assembly or define cellular localization. CORUM is freely accessible at http://mips.helmholtz-muenchen.de/corum/.

INTRODUCTION

Understanding biological processes at cellular and system levels is an important task in all living organisms. Protein complexes play critical roles in an array of biological processes, including protein synthesis, signaling and cellular degradation processes. To date, there are no reliable estimates about the total number of protein complexes in cells (complexome), but data from single cell organisms provide evidence that more than half of the gene products are involved in the formation of protein complexes (1). According to estimates from Berggård *et al.* (2), even more than 80% of proteins work in complexes. Many proteins are subunits of more than one complex, which extends the number of potential protein complexes. The RING-box protein 1 (RBX1), that was present in 35 protein complexes in the 2009 release of CORUM (3), is now found in 65 protein complexes.

Due to the importance of the topic, several endeavors were undertaken in order to unravel the cellular complexome. The first large-scale screens for protein complexes were performed in budding yeast (4,5) and discovered 491 and 547 complexes, respectively. Recent analyses of the interactome/complexome in human cells revealed a wealth of novel information (6–8). Integration of published datasets in the human protein complex map (hu.MAP) resulted in 4659 complexes composed from 7777 unique proteins (9).

For many years, the composition of thousands of protein complexes has been analysed in individual experiments and published in the scientific literature. In addition to the identification of subunit composition, which is the standard in high-throughput experiments, these complexes are also characterized with respect to their cellular function, association with diseases and sometimes stoichiometry. In order to provide a high-quality resource of information on mammalian protein complexes, we generated the comprehensive resource of mammalian protein complexes (CORUM) with 1750 complexes in the first release (10). The CORUM release 3.0 presents a significantly extended dataset that now consists of 4274 mammalian protein complexes. A graphical analysis tool was implemented that displays potential protein–protein interactions between the subunits of protein complexes. The tool is based on the Cytoscape javascript library and uses data of mammalian organisms from the IntAct database. At last, we present a collection of 58 protein complexes, where alternative splice variants result in altered complex function or affect diseases. CORUM

*To whom correspondence should be addressed. Tel: +49 89 3187 3189; Fax: +49 89 3187 3585; Email: andreas.ruepp@helmholtz-muenchen.de

is freely accessible at http://mips.helmholtz-muenchen.de/corum/.

DATABASE DESCRIPTION AND NEW DEVELOPMENTS

Design and application of CORUM

Our goal was the generation of a reference dataset of protein complex information from mammalian organisms. In order to obtain a high quality and reliability of the data, we only include protein complexes that have been isolated and characterized in individual experiments. Although protein complex information from high-throughput approaches provides an invaluable resource of novel information, we do not include it in CORUM as these data are usually not corroborated by experiments unraveling the biological function of complexes. As of note, our novel core set from CORUM 3.0 (3512 protein complexes) and the hu.MAP dataset share only 29 identical protein complexes (9). Experienced biocurators critically extract information from the scientific literature and transfer it into CORUM using established vocabularies and stable identifiers from well-known resources such as UniProt and Gene Ontology. The major focus for the application of CORUM lies on network biology and network medicine. Hence, the dataset is biased toward protein complexes which consist of at least two different subunits. According to Teichmann (11), the majority of (known) protein complexes in PDB are homomers. Based on recent statistics from PDB, 9206 homomers and 2677 heteromeric protein structures have been determined using X-ray crystallography. Of note, these data cover all groups of organisms. In the UniProt release 06_2018, 2344 proteins from the taxon Homo sapiens were annotated either as homodimer or homotetramer (12). In contrast, CORUM includes only 126 homomers. Apart from this exception, CORUM aims to offer a representation of the complexome of mammalian cells.

Since the previous CORUM release, basic annotation topics such as protein complex name, identification of subunits based on UniProt identifiers, references of used articles and comments remained unchanged. For functional annotation of protein complexes, we offer a mapping to Gene Ontology (GO) terms since CORUM 2.0. In the meantime, we switched manual functional annotation of novel complexes completely to GO terms (Figure 1).

Compared to the previous CORUM publication (3), the content of the dataset has increased from 2837 to 4274 protein complexes (Figure 2). In particular, the core set has grown considerably. The core set is a subset of CORUM that is reduced by eliminating redundant information such as protein complexes that were characterized from different mammalian species or complexes which were isolated by different methods. The core set now includes 3512 protein complexes, which is a gain of 70% compared to CORUM 2.0. Regarding the organisms that were used to characterize protein complexes, there are only minor changes comparing to the previous CORUM publication. The largest fraction of complexes was isolated from Homo sapiens (67%) followed by mouse (15%) and rat (10%). Other complexes were isolated from organisms such as cattle, pig, rabbit or are composed of subunits from different organisms. The growth

of the CORUM dataset is also accompanied by a higher total number of different gene products that are used as protein complex subunits. While in CORUM 2.0 complexes were composed of 3189 different proteins, the number now increased to 4473. This represents 22% of all known protein-coding genes according to the Ensembl database version 92.38 (13).

In recent years, the CORUM dataset was used in a large number of analyses as reference dataset, for benchmarking computational models and high-throughput experimental data. Examples are results from three pioneering high-throughput investigations that were combined in order to present a comprehensive complexome of human cells (9). In addition, protein complex information from CORUM is increasingly applied for the analyses of disease mechanisms. Use cases are network-based in silico drug efficacy screening (14), pituitary hormone deficiency in inherited gingival fibromatosis (15) or the architecture of human protein communities and disease networks (16). In the area of diseases, the CORUM dataset is particularly often applied in cancer analyses, demonstrated by more than ten citations during the past 2 years. These include protein-interaction network associated analyses of MLL(KMT2A)-fusion proteins in leukemia (17) and the detection of dysregulated protein-association networks in breast cancer cell lines (18).

Beside the analysis of experimental data, CORUM is also used by other data resources such as 'Interactome IN-SIDER' (19) or the 'MouseNet v2' database of gene networks (20). At last, CORUM is cross-referenced by the UCSC-Genome browser (21) and UniProt (12).

Annotation of splice variant complexes and their functional implications

To generate a high number of different gene products from comparably small genomes, mammals make use of processes such as alternative splicing to produce more than 82 335 distinct mRNAs e.g. in humans, according to the Gencode release 28 (22). Accordingly, mammalian protein complexes also exist in different isoforms, with the composition varying across cell types and conditions (23). In our new CORUM release, we provide a dataset of 58 protein complexes, which contain alternatively spliced subunits. Those subunits were found to alter cellular functions such as methylation of nucleosomal histone H3-lysine 27 by the PCM3 complex (with EED isoform 3) whereas the PRC2 complex (with EED isoform 1) preferentially methylates nucleosomal histone H1-lysine 26 (24). Examples, where alternative splicing affects protein complex functions via altered protein binding are the caspase-2 gene products. The short isoform, CASP-2S, inhibits apoptosis, whereas the long isoform, CASP-2L, promotes apoptosis. The CASP-2S–fodrin complex inhibits DNA damage-induced cytoplasmic fodrin cleavage. This process inhibits membrane blebbing and phosphatidylserine externalization that are indicative of apoptosis in cancer cells. The molecular basis for the different activities of the two CASP variants is that, in contrast to CASP-2S, the long isoform CASP-2L does not interact with fodrin (25).

The effect of alternatively spliced subunits of cytoskeletal protein dystrophin can be illustrated for DAPCs

| 1 | 7524 | Ubiquitin E3 ligase (CUL3, KLHL20, RBX1) | | | | Human |

Entry information
Complex ID: 7524
Date modified: 2018-06-15 14:15:51

Protein complex name and species
Complex Name: Ubiquitin E3 ligase (CUL3, KLHL20, RBX1)
Organism: Human
Tissue/Cell line: HEK 293T cells

Subunits

Protein description	Gene name	Gene name synonyms	Organism	UniProt ID	Entrez-ID
Cullin-3	CUL3	KIAA0617	Homo sapiens (Human)	Q13618	8452
E3 ubiquitin-protein ligase RBX1	RBX1	RNF75, ROC1	Homo sapiens (Human)	P62877	9978
Kelch-like protein 20	KLHL20	KLEIP,KLHLX	Homo sapiens (Human)	Q9Y2M5	27252

Subunits comment

Purification method
MI:0007-anti tag coimmunoprecipitation

Functional annotation (GO vs. Funcat)

GO characterization

GO	Description	Evi	PMID
GO:0031461	cullin-RING ubiquitin ligase complex	exp	20389280
GO:1904093	negative regulation of autophagic cell death	exp	20389280
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	exp	20389280

Functional characterization

Funcat
14.13.01.01 proteasomal degradation (ubiquitin/proteasomal pathway)
14.13.01 cytoplasmic and nuclear protein degradation
14.13 protein/peptide degradation
14 PROTEIN FATE (folding, modification, destination)

Reference
PubMed: 20389280

Comment
Death-associated protein kinase (DAPK) is a mediator of interferon-induced cell death. The Ubiquitin E3 ligase (CUL3, KLHL20, RBX1) complex promotes DAPK polyubiquitination, thereby inducing the proteasomal degradation of DAPK and attenuation of autophagy. In interferon-treated cells, the enrichment and sequestration of KLHL20 in PML nuclear bodies disrupts KLHL20–DAPK complex, thus impairing DAPK ubiquitination by KLHL20.

Disease comment

Subunits interaction
See interactions

Figure 1. Annotation of protein complexes in CORUM. Presentation of the Ubiquitin E3 ligase (CUL3, KLHL20, RBX1) complex in CORUM. Information about functional annotation using term from Gene Ontology is automatically translated into respective terms from the FunCat annotation scheme.

(dystrophin-associated protein complexes). Duchenne muscular dystrophy is caused by the absence of dystrophin (26). The DAPC is destabilized when dystrophin is absent, which leads to downregulated levels of the member proteins (27). Two dystrophin 71 isoforms (Dp71d and Dp71f) form multi-protein complexes in the hippocampal neurons. Dp71d-DAPC is mainly localized in bipolar GABAergic and Dp71f–DAPC in multipolar glutamatergic hippocampal neurons. The subunit composition of these protein complexes seems to affect their neuronal phenotype (28). It has been described that Dp71d–DAP complexes were present only in the nuclei of non-neuronal cells (29). However, recently it was demonstrated for the first time that isoform-containing complexes Dp71f–DAP and Dp71fd–DAP were localized in the nucleus of primary hippocampal neurons (28). This example shows that alternative splicing variants as subunits may regulate not only specific activities but also the tissue-specific localization of protein complexes.

The splice variant dataset illustrates that functional properties of particular protein complexes can only be represented by isoform-specific variant annotation. The splice variant complexes can be downloaded as a separate dataset.

Prediction of protein–protein interactions within protein complexes

Although there is a growing number of protein complexes with structure information, characterization of complexes
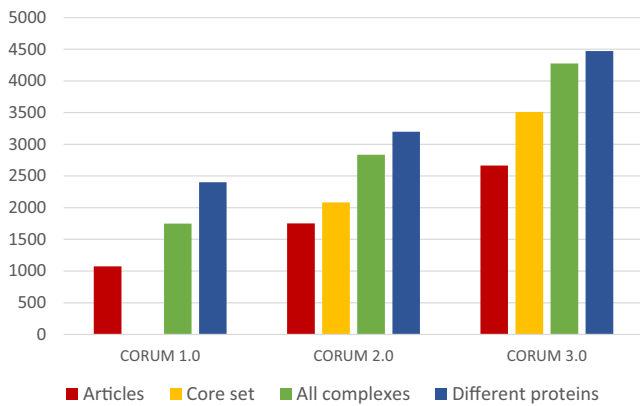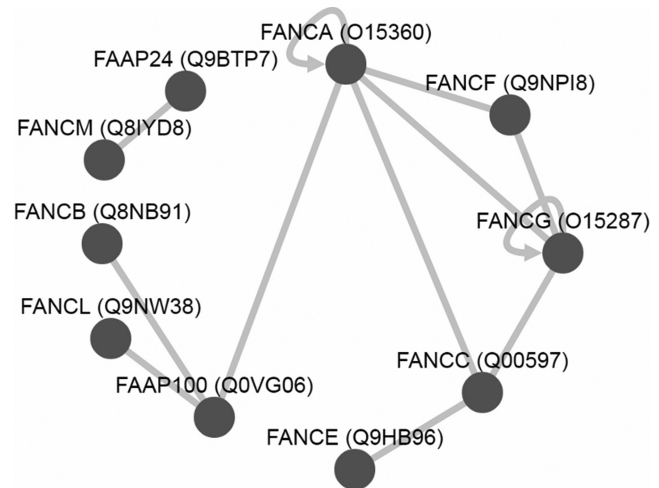
Figure 2. Data growth in CORUM. The plot compares the data content of CORUM versions 1.0, 2.0 and 3.0. It includes the number of articles that were used to create the datasets, the number of core set complexes, the total number of protein complexes as well as the total number of different proteins that are found in the dataset. As we did not provide a core set in CORUM 1.0, a respective number is missing.

Figure 3. Protein–protein interactions between protein complex subunits. Based on data from the IntAct database, validated protein–protein interactions of the Fanconi anemia FAAP100 complex (complex 6884) are displayed with Cytoscape.

is usually restricted to the identification of the subunits and biomedical information such as cellular function or association with diseases. Structural information such as stoichiometry of individual subunits is rarely discovered. In articles used for CORUM, we found only 288 protein complexes with stoichiometry data. For detailed structural information discovered by x-ray crystallography or nuclear magnetic resonance, information is even more sparse. In the new CORUM dataset 3.0, we have annotated 109 protein complexes with the PSI–MI interaction method, MI:0114 x-ray crystallography'. Preliminary information about protein complex structure with respect to neighboring proteins can be performed by serial protein interaction experiments. This was successfully applied for elucidation of the chaperonin Tric/CCT (30).

In order to provide users with the option to inspect putative interactions between complex subunits, the new CORUM release provides a graphical tool that integrates known PPI information. A large publicly available collection of protein–protein interactions is provided by the IntAct database which includes data from different resources (31). For the CORUM tool, we used the 'physical association'-interactions of IntAct, release 05 November 2018, from the organisms *Homo sapiens*, *Rattus norvegicus* and *Mus musculus*. The largest fraction of PPI information was obtained from Homo sapiens (238 841 entries), followed by mouse (15 138) and rat (3962).

For visualization of interactions between protein complex subunits we use Cytoscape. Cytoscape.js is a javascript-based graph-visualization library that we embedded in version 3.2 on our website. For each protein complex that contains at least one interaction we show all protein–protein interactions according to the IntAct dataset in a graph below the list of search results. For Fanconi anemia FAAP100 complex (complex 6884) for example, nine interactions between subunits and two self-associations were found (Figure 3). Five interactions between FANCA, FANCC, FANCF and FANCG give rise to the speculation that these four proteins build a core of the complex. This is corrobo-

rated by analyses of four forms of the Fanconi anaemia core complexes from different subcellular compartments (32). All forms of the complex contained at least these four proteins.

The fact, that a protein–protein interaction was discovered in an experiment is no proof that it also exists in a protein complex. However, it is tempting to assume that for the majority of protein complexes, PPI data provide a reliable prediction of intramolecular associations.

CONCLUSION

CORUM is a publicly available, centralized database for mammalian protein complexes based on manually curated information from scientific literature. Basic and translational researchers are provided with extensive search options to look for complexes containing their genes of interest, exhibiting a specific biological function or displaying other features. Computational biologists can download entire datasets in different formats for advanced studies. The importance of protein complex data is demonstrated by hundreds of studies that used the CORUM dataset during the last decade. These include basic research such as the inventory of the mammalian complexome as well as applied biomedical research, in particular of cancer. The CORUM 3.0 release provides a substantially enlarged dataset of mammalian complexes which is accompanied by a wider coverage of gene products that serve as subunits of protein complexes.

In recent years, a growing number of studies has demonstrated that variants of gene products may have substantial effects on protein complex function. Here, we present for the first time a dataset that covers the impact of splice variants on cellular processes and diseases.

In addition to the larger dataset, the CORUM 3.0 release also presents a Cytoscape-based tool for the graphical representation of known protein–protein interactions of subunits. As structural information about large protein com-

plexes is sparse, this approach allows the prediction of structurally adjacent proteins within these large complexes.

An important goal of CORUM for the future will be to obtain an even more complete representation of experimentally characterized protein complexes. To achieve this goal, we welcome the input of other researchers sending us information about novel published protein complexes not yet included in the dataset. Please contact us at andreas.ruepp@helmholtz-muenchen.de.

REFERENCES

1. Guldener,U., Munsterkotter,M., Kastenmuller,G., Strack,N., van Helden,J., Lemer,C., Richelles,J., Wodak,S.J., Garcia-Martinez,J., Perez-Ortin,J.E. *et al.* (2005) CYGD: the comprehensive yeast genome database. *Nucleic Acids Res.*, **33**, D364–D368.
2. Berggard,T., Linse,S. and James,P. (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics*, **7**, 2833–2842.
3. Ruepp,A., Waegele,B., Lechner,M., Brauner,B., Dunger-Kaltenbach,I., Fobo,G., Frishman,G., Montrone,C. and Mewes,H.W. (2010) CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res.*, **38**, D497–D501.
4. Gavin,A.C., Aloy,P., Grandi,P., Krause,R., Boesche,M., Marzioch,M., Rau,C., Jensen,L.J., Bastuck,S., Dumpelfeld,B. *et al.* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature*, **440**, 631–636.
5. Krogan,N.J., Cagney,G., Yu,H., Zhong,G., Guo,X., Ignatchenko,A., Li,J., Pu,S., Datta,N., Tikuisis,A.P. *et al.* (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*, **440**, 637–643.
6. Hein,M.Y., Hubner,N.C., Poser,I., Cox,J., Nagaraj,N., Toyoda,Y., Gak,I.A., Weisswange,I., Mansfeld,J., Buchholz,F. *et al.* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell*, **163**, 712–723.
7. Huttlin,E.L., Ting,L., Bruckner,R.J., Gebreab,F., Gygi,M.P., Szpyt,J., Tam,S., Zarraga,G., Colby,G., Baltier,K. *et al.* (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell*, **162**, 425–440.
8. Wan,C., Borgeson,B., Phanse,S., Tu,F., Drew,K., Clark,G., Xiong,X., Kagan,O., Kwan,J., Bezginov,A. *et al.* (2015) Panorama of ancient metazoan macromolecular complexes. *Nature*, **525**, 339–344.
9. Drew,K., Lee,C., Huizar,R.L., Tu,F., Borgeson,B., McWhite,C.D., Ma,Y., Wallingford,J.B. and Marcotte,E.M. (2017) Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol. Syst. Biol.*, **13**, 932.
10. Ruepp,A., Brauner,B., Dunger-Kaltenbach,I., Frishman,G., Montrone,C., Stransky,M., Waegele,B., Schmidt,T., Doudieu,O.N., Stumpflen,V. *et al.* (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.*, **36**, D646–D650.
11. Marsh,J.A. and Teichmann,S.A. (2015) Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.*, **84**, 551–575.
12. UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **46**, 2699.
13. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Giron,C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
14. Guney,E., Menche,J., Vidal,M. and Barabasi,A.L. (2016) Network-based in silico drug efficacy screening. *Nat. Commun.*, **7**, 10331.
15. Tommiska,J., Kansakoski,J., Skibsbye,L., Vaaralahti,K., Liu,X., Lodge,E.J., Tang,C., Yuan,L., Fagerholm,R., Kanters,J.K. *et al.* (2017) Two missense mutations in KCNQ1 cause pituitary hormone deficiency and maternally inherited gingival fibromatosis. *Nat. Commun.*, **8**, 1289.
16. Huttlin,E.L., Bruckner,R.J., Paulo,J.A., Cannon,J.R., Ting,L., Baltier,K., Colby,G., Gebreab,F., Gygi,M.P., Parzen,H. *et al.* (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature*, **545**, 505–509.
17. Skucha,A., Ebner,J., Schmollerl,J., Roth,M., Eder,T., Cesar-Razquin,A., Stukalov,A., Vittori,S., Muhar,M., Lu,B. *et al.* (2018) MLL-fusion-driven leukemia requires SETD2 to safeguard genomic integrity. *Nat. Commun.*, **9**, 1983.
18. Lapek,J.D. Jr, Greninger,P., Morris,R., Amzallag,A., Pruteanu-Malinici,I., Benes,C.H. and Haas,W. (2017) Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. *Nat. Biotechnol.*, **35**, 983–989.
19. Meyer,M.J., Beltran,J.F., Liang,S., Fragoza,R., Rumack,A., Liang,J., Wei,X. and Yu,H. (2018) Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods*, **15**, 107–114.
20. Kim,E., Hwang,S., Kim,H., Shim,H., Kang,B., Yang,S., Shim,J.H., Shin,S.Y., Marcotte,E.M. and Lee,I. (2016) MouseNet v2: a database of gene networks for studying the laboratory mouse and eight other model vertebrates. *Nucleic Acids Res.*, **44**, D848–D854.
21. Casper,J., Zweig,A.S., Villarreal,C., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Karolchik,D. *et al.* (2018) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.*, **46**, D762–D769.
22. Harrow,J., Frankish,A., Gonzalez,J.M., Tapanari,E., Diekhans,M., Kokocinski,F., Aken,B.L., Barrell,D., Zadissa,A., Searle,S. *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.*, **22**, 1760–1774.
23. Ryan,C.J., Kennedy,S., Bajrami,I., Matallanas,D. and Lord,C.J. (2017) A compendium of Co-regulated protein complexes in breast cancer reveals collateral loss events. *Cell Syst.*, **5**, 399–409.
24. Kuzmichev,A., Jenuwein,T., Tempst,P. and Reinberg,D. (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell*, **14**, 183–193.
25. Han,C., Zhao,R., Kroger,J., Qu,M., Wani,A.A. and Wang,Q.E. (2013) Caspase-2 short isoform interacts with membrane-associated cytoskeleton proteins to inhibit apoptosis. *PLoS One*, **8**, e67033.
26. Hoffman,E.P., Brown,R.H. Jr and Kunkel,L.M. (1987) Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell*, **51**, 919–928.
27. Straub,V. and Campbell,K.P. (1997) Muscular dystrophies and the dystrophin-glycoprotein complex. *Curr. Opin. Neurol.*, **10**, 168–175.
28. Rodriguez-Munoz,R., Cardenas-Aguayo Mdel,C., Aleman,V., Osorio,B., Chavez-Gonzalez,O., Rendon,A., Martinez-Rojas,D. and Meraz-Rios,M.A. (2015) Novel nuclear protein complexes of dystrophin 71 isoforms in rat cultured hippocampal GABAergic and glutamatergic neurons. *PLoS One*, **10**, e0137328.
29. Gonzalez-Ramirez,R., Morales-Lazaro,S.L., Tapia-Ramirez,V., Mornet,D. and Cisneros,B. (2008) Nuclear and nuclear envelope localization of dystrophin Dp71 and dystrophin-associated proteins (DAPs) in the C2C12 muscle cells: DAPs nuclear localization is modulated during myogenesis. *J. Cell. Biochem.*, **105**, 735–745.
30. Liou,A.K. and Willison,K.R. (1997) Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *EMBO J.*, **16**, 4311–4316.
31. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
32. Thomashevski,A., High,A.A., Drozd,M., Shabanowitz,J., Hunt,D.F., Grant,P.A. and Kupfer,G.M. (2004) The Fanconi anemia core complex forms four complexes of different sizes in different subcellular compartments. *J. Biol. Chem.*, **279**, 26201–26209.