Computational and Structural Biotechnology Journal 20 (2022) 2551–2557

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Data article

DisPhaseDB: An integrative database of diseases related variations in liquid–liquid phase separation proteins

Alvaro M. Navarro, Fernando Orti, Elizabeth Martínez-Pérez, Macarena Alonso, Franco L. Simonetti, Javier A. Iserte *, Cristina Marino-Buslje *

Fundación Instituto Leloir. Av. Patricias Argentinas 435, C1405BWE Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 11 March 2022
Received in revised form 3 May 2022
Accepted 3 May 2022
Available online 12 May 2022

Keywords:
Liquid–liquid phase separation proteins
LLPS
Disease variations
Membrane-less organelles
MLO
Diseases
Database
Web server

ABSTRACT

Motivation: Proteins involved in liquid–liquid phase separation (LLPS) and membraneless organelles (MLOs) are recognized to be decisive for many biological processes and also responsible for several diseases. The recent explosion of research in the area still lacks tools for the analysis and data integration among different repositories. Currently, there is not a comprehensive and dedicated database that collects all disease-related variations in combination with the protein location, biological role in the MLO, and all the metadata available for each protein and disease. Disease-related protein variants and additional features are dispersed and the user has to navigate many databases, with a different focus, formats, and often not user friendly.

Results: We present DisPhaseDB, a database dedicated to disease-related variants of liquid–liquid phase separation proteins. It integrates 10 databases, contains 5,741 proteins, 1,660,059 variants, and 4,051 disease terms. It also offers intuitive navigation and an informative display. It constitutes a pivotal starting point for further analysis, encouraging the development of new computational tools.

The database is freely available at http://disphasedb.leloir.org.ar.

1. Introduction

Cells compartmentalize biological processes to achieve spatial and temporal control over biochemical reactions. This is accomplished through both membrane bound and membraneless organelles (MLOs). MLOs are formed through the process of Liquid–liquid phase separation (LLPS) in which a liquid demixes in two phases where one phase is enriched in particular macromolecules, while depleted in others [1–3].

Examples are Nucleolus, Cajal bodies, and nuclear speckles in the nucleus and stress granules, P granules and P-bodies in the cytoplasm [1,3], among others. These structures play diverse roles in various biological processes such as organization of the cytoplasm and nucleoplasm, regulation of gene expression, signaling, transport and compartmentalization [4,5]. However, they are also increasingly implicated in several complex human diseases [5–7]. Examples of abnormal LLPS have been implicated in cancer, neurodegenerative and infectious diseases among others [8–12]. Therefore, it is not surprising that a perturbation in proteins that undergo LLPS, like a single nucleotide variant (SNV), gene copy number variation (CNV), protein mutation and post-translational modifications (PTMs) can upset the fine tuned process of MLO formation, stability and dynamics [6,13–20].

Proteins that undergo LLPS are often intrinsically disordered or have disordered regions (IDPs and IDRs, respectively), they might also have a biased amino acid composition or low-complexity regions (LCRs), and are therefore highly dynamic [21–23]. To a large extent, these regions are responsible for the separation in phases, although other types of regions or domains can also be found in proteins that separate into phases [21,24,25]. There are several molecular interaction types contributing to LLPS, such as multivalent protein–protein and protein-DNA/RNA interactions. Also, dynamically transient interacting regions as IDRs, LCR and prion-like, aggregation, coacervation, electrostatic, cation-π and π-π interactions, among others [26]. Mapping mutations to structural features could help to understand mechanisms involved in the formation of pathological aggregates. As an example, it was shown that mutations in the prion-like domains (PLDs) of several proteins are involved in neurodegenerative diseases such as

* Corresponding authors.
E-mail addresses: jiserte@leloir.org.ar (J.A. Iserte), cmb@leloir.org.ar (C. Marino-Buslje).

https://doi.org/10.1016/j.csbj.2022.05.004

amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and multisystem proteinopathy [11].

Numerous experimental methods have been developed or repurposed to study the LLPS process and proteins involved, such as fluorescence recovery after photobleaching (FRAP), nuclear magnetic resonance spectroscopy (NMR), immunofluorescence, fluorescence correlation spectroscopy (FCS), and many others [27–30]. However, there are still not enough bioinformatics tools and databases to study them, much less in the context of human diseases. We hypothesize that this is in part due to the lack of centralized data repositories, the low agreement among existing ones, the scarcity of dedicated cross-referenced databases and, the poor scalability offered for large integrative analysis of phase separating proteins.

It was shown that the agreement between 4 dedicated databases of LLPS proteins [31–34] is rather poor, sharing 42 human proteins out of 4,367, proving that none of the four databases taken alone provides enough data to enable a meaningful analysis [35], added to the fact that they do not focus on protein variations in diseases.

To cover this gap, we present **DisPhaseDB**, an integrative database focused on disease variations in LLPS proteins. The database encompasses all known phase separating proteins, including Drivers, Clients, Regulators and other MLOs experimentally associated proteins together with their disease related variations. We expect our database to be of interest for researchers studying MLOs, LLPS proteins, diseases, proteins for targeting therapies, specific MLO components in a disease and also for computational groups developing methods to understand sequence-function relationships and mutational impact.

2. Methods

2.1. Selection of proteins involved in LLPS and MLO associated

Our starting point was an integrated set of MLO associated human proteins that were collected in a previous group effort [35]. It consists of the entries of four databases of LLPS and MLOs associated proteins that were compiled, merged, completed and stored in a local database: **PhaSePro** [31], **PhaSepDB** [32] **DrLLPS** [33] and **LLPSDB** [34]. This set is periodically updated with the databases' new releases. The consolidated dataset is available at https://mlos.leloir.org.ar/ [35].

The role of the proteins in the LLPS process and their association with the MLOs, is taken from the annotation of the source database. There are four types of Protein roles: Drivers, Regulators, Clients and Unassigned when no database describes its role. In addition, we grouped their experimental evidence supporting the roles as low throughput and high throughput for user evaluation of their confidence.

2.2. Mutation collection

We obtained human coding variants from **ClinVar** release 20200402 [36], a public archive of human genetic variants and their interpretation with respect to a clinical condition or phenotypes, along with supporting evidence for such association. **DisGeNET** [37] offers several datasets based on gene-disease associations (GDAs) and variant-disease associations (VDAs). For our database we took mutations from the curated VDA dataset (October 2020), which at the same time integrates variants from UniProt, ClinVar, GWASdb [38] and GWAS catalog [39].

From **UniProt** [40] we used the dataset of human variants that are manually annotated in UniProtKB/Swiss-Prot (release-2021_02). Lastly, **COSMIC** release v94 was used to obtain the coding point mutations in human cancers [41].

In all cases, we mapped variants with genomic coordinates from the human genome assembly GRCh38 onto the canonical protein sequence. Disease and other altered phenotypic effects annotations in ClinVar, COSMIC, DisGeNET and UniProt are not consistent between databases nor within the same database. They are frequently cross referenced to one or many ontologies that collect medical terms, and/or diseases, such as Disease Ontology (DO) [42], the Human Phenotype Ontology (HPO) [43], Medical Subject Headings (MeSH) [44], Medical Genetics (MedGen, https://www.ncbi.nlm.nih.gov/medgen/), The Monarch Merged Disease Ontology (MONDO) [45], National Cancer Institute Thesaurus (NCI, https://ncim.nci.nih.gov/), Online Mendelian Inheritance in Man (OMIM) [46], among others. In some cases there is no reference to any ontology. Furthermore, a mutation can be associated with several diseases and vice versa. Thus, in this context studying a variant, a protein or a disease is challenging. As an example, mutation R521C in FUS protein is associated with different diseases in different ontologies: Melanoma of skin (SNOMEDCT_US: 93655004), amyotrophic lateral sclerosis ALS6 (MEDCIN: 315716 and MedGen: C1842675) and Gastric Carcinoma (NCI: C4911). In addition, there are many synonymous annotated for the same disease in one ontology, as an example "Cancer of Stomach", "Cancer of the Stomach", "Carcinoma of Stomach", "Gastric Cancer", etc, are references to the same disease in NCI. Another case are synonymous in different ontologies, as example: Cutaneous Melanoma (MedGen: C0151779), Melanoma of skin (SNOMEDCT_US: 93655004) and "Melanoma, Cutaneous Malignant" (OMIM: 155600).

Finally, there are different grades of specificity of a disease that are referred to as different terms, as an example, "Acanthoma" is a type of "Skin Neoplasms". Therefore, mapping all disease terminology into a single ontology is not feasible. So, to facilitate the user navigating through this tangle of terms in dozens of ontologies to study a variation or a protein, DisPhaseDB includes all available disease annotations and, when there are no references to an ontology, reference to the source mutation database.

2.3. Additional information

We also included molecular features such as structural domains from Pfam database (Mistry et al., 2020), Intrinsically disordered Regions (IDRs) and Low-Complexity Regions (LCRs) from MobiDB [47], post-translational modifications (PTMs) retrieved from PhosphoSitePlus [48] and Prion-like domains (PLDs) predicted by PLAAC [48–49]. These features are displayed on the protein sequence using the "Feature-Viewer" tool to visualize positional data [50].

2.4. Server construction and access

The server backend consists of a http web-server developed in Python 3.8+ using the Flask framework and MySQL. The client web application was developed with the AngularJS framework.

3. Results

3.1. DisPhaseDB in numbers

We present DisPhaseDB, available at https://disphasedb.leloir.org.ar.

DisPhaseDB contains 5,741 LLPS proteins, all of them with experimental evidence that supports their association to the MLOs. For these proteins we collected human disease mutations from

up-to-date databases including **UniProt**, **ClinVar**, **DisGeNET** and **COSMIC**. After merging the four databases, the total number of unique coding variants (protein mutations) is 1,660,059. **COSMIC** contributes 1,464,124, **ClinVar** 221,097, **DisGeNET** 56,813 and **UniProt** 22,965. Supplementary Fig. 1 shows the overlap of the four protein variation resources, showing that all of them are needed to have a better landscape of mutation in LLPS proteins. The most common type are missense mutations, followed by synonymous mutations (66.57% and 23.41% respectively) (Supplementary Fig. 2).

It is evident that an amino acid change due to a missense mutation could influence protein structure, function and LLPS behavior. However, synonymous SNPs can have a substantial contribution to disease risk and other complex traits. There are various molecular mechanisms that underlie these effects such as: altering splicing efficiency and/or accuracy, losing information of exon–intron boundaries [51], affecting post-transcriptional processing and regulation of RNA, influencing the kinetics of mRNA translation [52] and affecting the timing of cotranslational folding due to rare codons [53], among others.

On average, proteins in DisPhaseDB have around 200 mutations, although few proteins are exceptionally highly mutated (Fig. 1). As an example, TITIN (20,552 mutations) is a key component of striated muscles and mutations in this protein are related to different types of cardiomyopathies and muscular dystrophies [54–56]. BRCA1, BRCA2 and APC (9,172, 12,063 and 9,237 mutations respec-

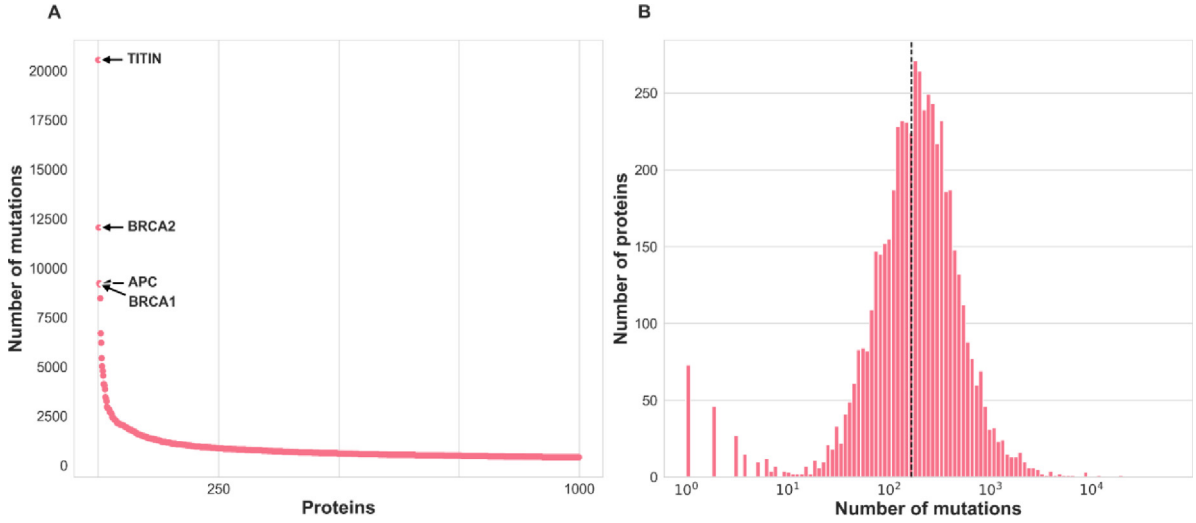

Fig. 1. A) Number of mutations by protein (only the first 1000 most mutated proteins are shown). B) Distribution of proteins by the number of mutations.

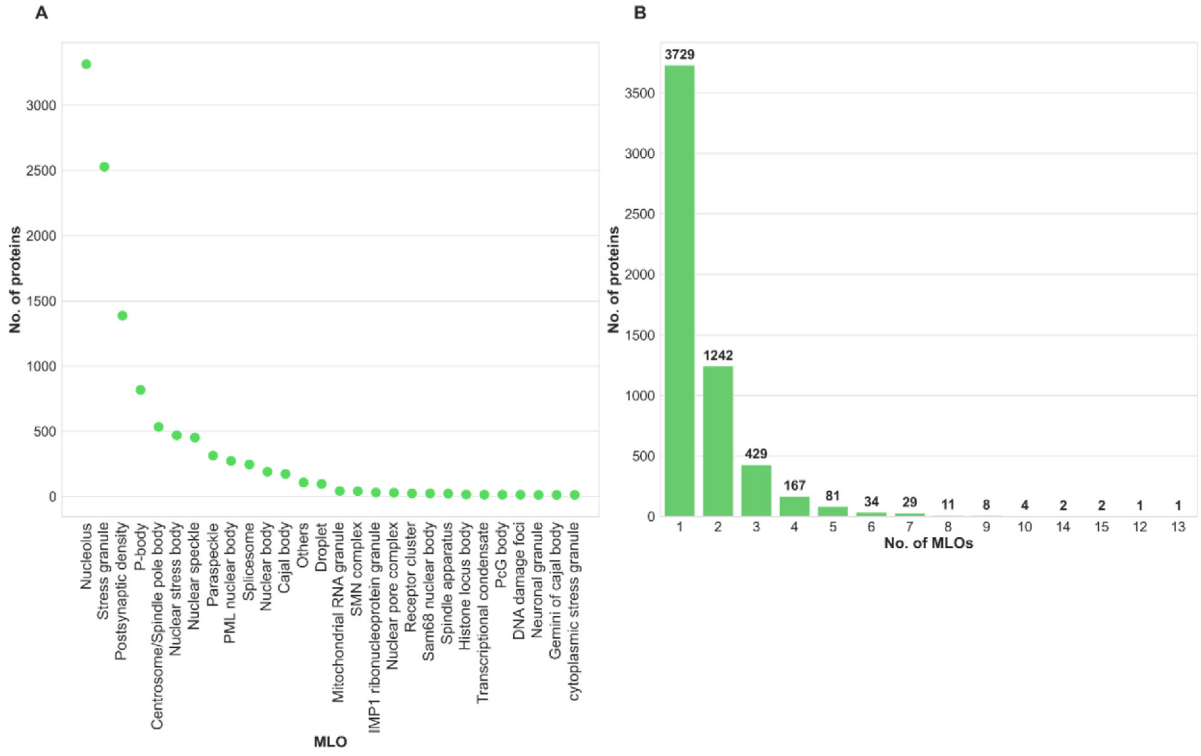

Fig. 2. A) Protein distribution by MLO in DisPhaseDB, showing only MLOs with more than 10 associated proteins. B) Number of MLOs in which a protein can be present.

tively) are proteins involved in DNA repair and tumor suppressor [57–59]. It is well known mutations in these proteins produce an increased risk for different types of cancer, especially breast, ovarian and colorectal cancer [60–62]. Mutations do not appear equally in different protein regions, IDR and LCR have more mutations than the ordered portion of the protein (Supplementary Fig. 3).

Each protein is associated with one or more LLPS source databases and, when possible, with their role in the LLPS process. Protein roles can vary depending on the MLO and the source database leading to diverse situations. A protein can be annotated as Driver in a particular MLO and as Client in another, also a protein can have a role in one database and be unassigned or have a different one in another for the same MLO. There are 285 proteins classified as Drivers, 357 regulators, 3,157 potential clients, and 4,105 have no role assigned in their source databases or MLO (Supplementary Fig. 4 shows the distribution of proteins by their role and, disaggregated by MLO).

Mutated proteins of DisPhaseDB are associated with 103 MLOs, varying in number across them. As an example, the nucleolus has 3,315 associated proteins while the synaptosome has only 1. Most proteins are associated with a single MLO (3,729), being the maximum 13 MLOs (1 protein) (Fig. 2).

Also mutated proteins are associated with one or more diseases, Fig. 3 (upper panel) shows the number of DisPhaseDB mutated proteins associated with all the Mesh ontology subheadings in the disease category. These headings are nodes near the root of the ontology, but the annotations allow going forward to more specificity, for example Supplementary Fig. 5 shows the terms under "neoplasms" subheading disaggregated by site. Since 80% of the mutations in DisPhaseDB are contributed by COSMIC (somatic mutations in cancer). Fig. 3 (lower panel) shows the distribution of mutated proteins by disease removing those mutations contributed by COSMIC. Even though removing COSMIC mutations, proteins associated with neoplasms are still predominant.

3.2. Server usage

DisPhaseDB offers either a quick search by protein, MLO or disease or an advanced search applying one or several filters. Possible filters are by protein, role, MLO, disease name or keyword,

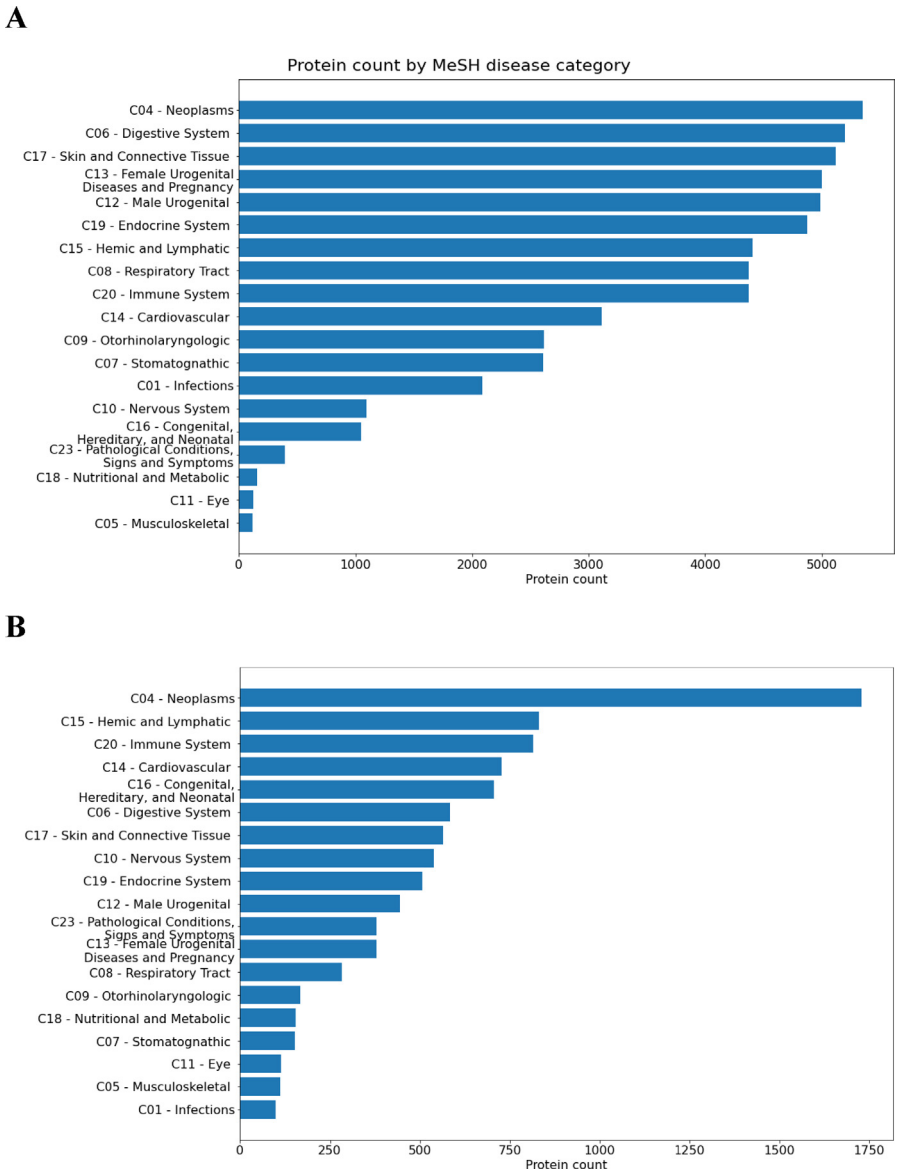

Fig. 3. A) Distribution of the total mutated proteins among all the subheading in MeSh ontology. B) Same as A, but excluding COSMIC contributed mutations to see the tendencie (COSMIC contributes with 1,464,124 mutations out of 1,660,059).

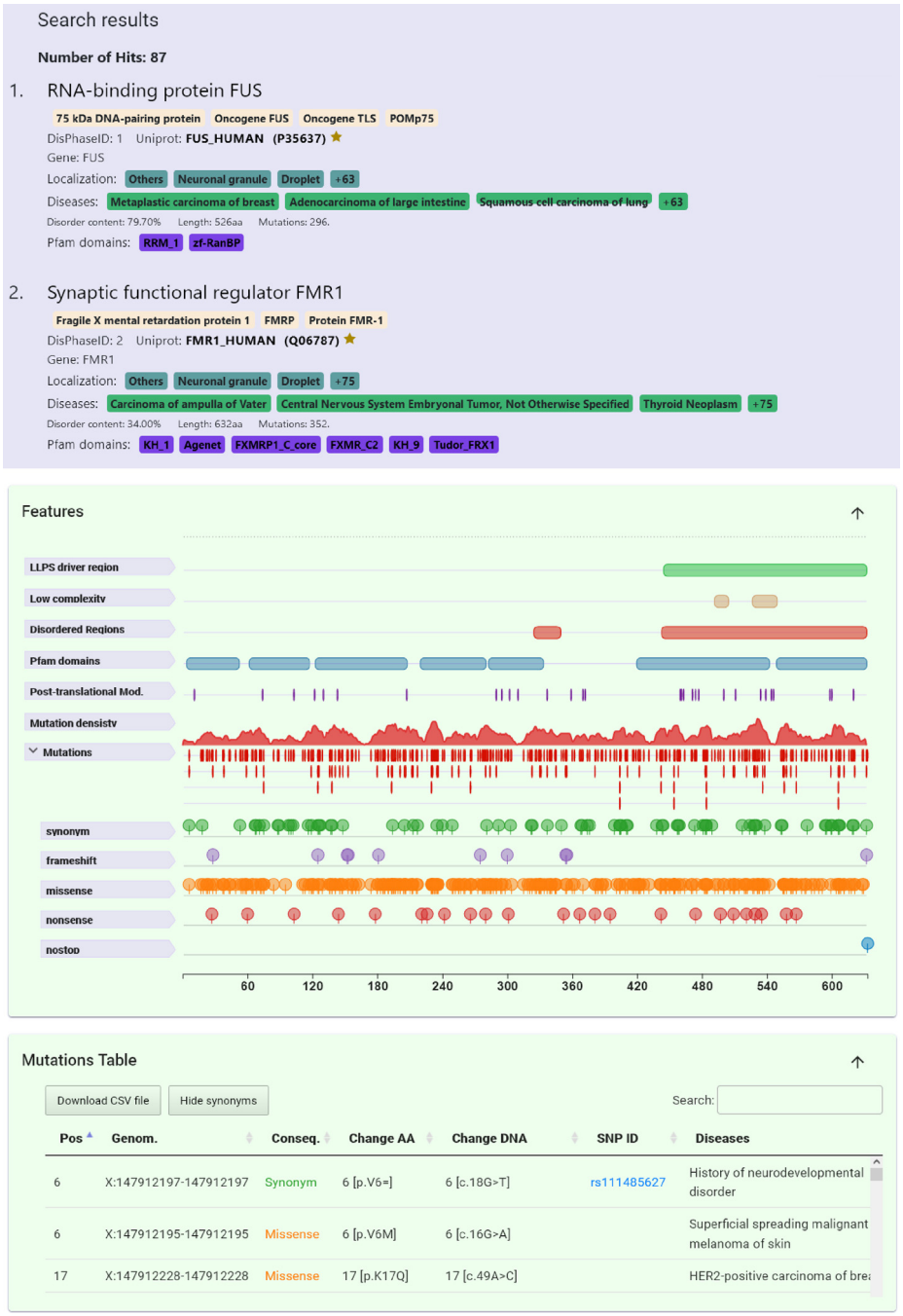

Fig. 4. Example of search by Hepatobiliary Neoplasm disease. The top panel shows the first two proteins of a list of 87 related to the query disease. Middle panel shows the protein features mapped onto the sequence and the bottom panel shows a portion of the list of disease related mutations in which the protein is involved (three out of 352 mutations).

by evidence (low or high throughput experiments), by protein disorder content and mutation type (missense, frameshift, nonsense, etc). In addition, filters can be combined in such a way that users can customize the set of proteins according to their need or interest.

As an example, Fig. 4 shows a search by a particular disease: Hepatobiliary Neoplasm. The output is a list of proteins involved in this disease with relevant annotations. By clicking a protein, further characteristics are expanded. As an example, synaptic functional regulator FMR1 (UP: Q06787) is selected. The information displayed is divided in the following sections: I) a protein summary with general information and the fasta sequence; II) protein MLO location; III) protein features mapped onto the sequence such as regions, domains, disorder content and mutations (disaggregated by type), among others IV) a mutation summary and V) a mutation table to download. Fig. 4 is a composite of different parts of the search results and output for illustrative purposes.

4. Discussion

To the best of our knowledge, there is not an integrated and comprehensive resource for mutations in MLOs associated proteins. For this reason, we integrated all state-of-the-art resources of proteins involved in LLPS and MLOs with four relevant disease

databases that annotate medical terms and phenotypic effects. The selected variant databases with clinical relevance are not redundant showing very little overlap among them. In such a way to cover the range of diseases and variant effects.

Variant databases are often not user friendly and they cross-reference to different disease ontologies and many other databases. This highlights the need for a unification of these resources.

DisPhaseDB also provides mutation mapping onto the protein sequence and associated metadata, such as disordered, low complexity and ordered regions, post translational modifications, among other features.

Therefore this resource will be helpful to investigate sequence-function relationships and mutational impact on LLPS proteins, to assist researchers to better understand complex human diseases under the lens of phase separation.

Funding

AMN, FO, MA are PhD fellows, EMP is Postdoctoral fellow and FS, JI and CMB are researchers of National Scientific and Technical Research Council (CONICET) - Argentina. This work was partially funded by PICT-2018-01015.

CRediT authorship contribution statement

Alvaro M. Navarro: Formal analysis, Investigation, Methodology, Software, Visualization, Validation. **Fernando Orti:** Data curation, Methodology, Software. **Elizabeth Martínez-Pérez:** Data curation, Investigation, Methodology, Software. **Macarena Alonso:** Data curation, Methodology. **Franco Simonetti:** Conceptualization, Methodology, Supervision, Writing review. **Javier Iserte:** Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing-original draft. **Cristina Marino-Buslje:** Conceptualization, Formal analysis, Funding acquisition, Resources, Investigation, Supervision, Writing original draft, review & editing.

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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.csbj.2022.05.004.

References

[1] Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol 2017;18(5):285–98.
[2] Sanders DW, Kedersha N, Lee DSW, Strom AR, Drake V, Riback JA, et al. Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell* 2020; 181 (2): 306–24.e28.
[3] Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. Trends Genet TIG 2011;27(8):295–306.
[4] Gomes E, Shorter J. The molecular language of membraneless organelles. J Biol Chem 2019;294(18):7115–27.
[5] Su Qi, Mehta S, Zhang J. Liquid-liquid phase separation: orchestrating cell signaling through time and space. Mol Cell 2021;81(20):4137–46.
[6] Lenard AJ, Hutten S, Zhou Q, Usluer S, Zhang F, Bourgeois BMR, et al. Phosphorylation regulates CIRBP arginine methylation, transportin-1 binding and liquid-liquid phase separation. Front Mol Biosci 2021;8(October):689687.
[7] Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science 2017;357(6357). https://doi.org/10.1126/science.aaf4382.
[8] Ryan VH, Fawzi NL. Physiological, pathological, and targetable membraneless organelles in neurons. Trends Neurosci 2019;42(10):693–708.
[9] Prasad A, Bharathi V, Sivalingam V, Girdhar A, Patel BK. Molecular mechanisms of TDP-43 misfolding and pathology in amyotrophic lateral sclerosis. Front Mol Neurosci 2019;12(February):25.
[10] Alberti S, Dormann D. Liquid-liquid phase separation in disease. Annu Rev Genet 2019;53(December):171–94.
[11] Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. Biochem J 2017;474(8):1417–38.
[12] Tsang B, Pritišanac I, Scherer SW, Moses AM, Forman-Kay JD. Phase separation as a missing mechanism for interpretation of disease mutations. Cell 2020;183(7):1742–56.
[13] Luo R, Fan Yu, Yang J, Ye M, Zhang D-F, Guo K, et al. A novel missense variant in ACAA1 contributes to early-onset Alzheimer's disease, impairs lysosomal function, and facilitates amyloid-β pathology and cognitive decline. Signal Transd Target Ther 2021;6(1):325.
[14] Wu X, Cai Q, Feng Z, Zhang M. Liquid-liquid phase separation in neuronal development and synaptic signaling. Dev Cell 2020;55(1):18–29.
[15] Murakami T, Qamar S, Lin JQ, Kaminski GS, Schierle ER, Miyashita A, et al. ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. Neuron 2015;88(4):678–90.
[16] Hofweber M, Dormann D. Friend or foe—post-translational modifications as regulators of phase separation and RNP granule dynamics. J Biol Chemi 2019;294(18):7137–50.
[17] Gopal PP, Nirschl JJ, Klinman E, Holzbaur ELF. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. Proc Natl Acad Sci U S A 2017;114(12):E2466–75.
[18] Schisa JA, Elaswad MT. An emerging role for post-translational modifications in regulating RNP condensates in the germ line. Front Mol Biosci 2021;8(April):658020.
[19] Tang S-J. Potential role of phase separation of repetitive DNA in chromosomal organization. Genes 2017;8(10). https://doi.org/10.3390/genes8100279.
[20] Specht CG. A quantitative perspective of alpha-synuclein dynamics - why numbers matter. Front Synap Neurosci 2021;13(October):753462.
[21] Luo Y-Y, Jun-Jun Wu, Li Y-M. Regulation of liquid-liquid phase separation with focus on post-translational modifications. Chem Commun 2021;57(98):13275–87.
[22] Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, et al. Protein phase separation: a new phase in cell biology. Trends Cell Biol 2018;28(6):420–35.
[23] van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Keith Dunker A, et al. Classification of intrinsically disordered regions and proteins. Chem Rev 2014;114(13):6589–631.
[24] Dignon GL, Best RB, Mittal J. Biomolecular phase separation: from molecular driving forces to macroscopic properties. Annu Rev Phys Chem 2020;71(April):53–75.
[25] Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, et al. Pi-pi contacts are an overlooked protein feature relevant to phase separation. *eLife*. 2018; 7 (February). https://doi.org/10.7554/eLife.31486.
[26] Mittag T, Parker R. Multiple modes of protein-protein interactions promote RNP granule assembly. J Mol Biol 2018;430(23):4636–49.
[27] Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 2009;324(5935):1729–32.
[28] Schmidt, Broder H, Görlich D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* 4 (January). 2015. https://doi.org/10.7554/eLife.04251.
[29] Gadd JC, Kuyper CL, Fujimoto BS, Allen RW, Chiu DT. Sizing subcellular organelles and nanoparticles confined within aqueous droplets. Anal Chem 2008;80(9):3450–7.
[30] Alderson TR, Kay LE. NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. Cell 2021;184(3):577–95.
[31] Mészáros B, Erdős G, Szabó B, Schád É, Tantos Á, Abukhairan R, et al. PhaSePro: the database of proteins driving liquid-liquid phase separation. Nucleic Acids Res 2020;48(D1):D360–7.
[32] You K, Huang Qi, Chunyu Yu, Shen B, Sevilla C, Shi M, et al. PhaSepDB: A database of liquid-liquid phase separation related proteins. Nucleic Acids Res 2020;48(D1):D354–9.
[33] Ning W, Guo Y, Lin S, Bin Mei YW, Jiang P, Tan X, et al. DrLLPS: A data resource of liquid-liquid phase separation in eukaryotes. Nucleic Acids Res 2020;48(D1):D288–95.
[34] Li Q, Peng X, Li Y, Tang W, Zhu J, Huang J, et al. LLPSDB: A database of proteins undergoing liquid–liquid phase separation in vitro. Nucleic Acids Res 2019;48(D1):D320–7.
[35] Orti F, Navarro AM, Rabinovich A, Wodak SJ, Marino-Buslje C. Insight into membraneless organelles and their associated proteins: drivers, clients and regulators. Comput Struct Biotechnol J 2021;19(June):3964–77.
[36] Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;46(D1):D1062–7.
[37] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res 2020;48(D1):D845–55.
[38] Li MJn, Wang P, Liu X, Lim EL, Wang Z, Yeager M, et al. GWASdb: A database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res* 2012; 40 (Database issue): D1047–54.
[39] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47(D1):D1005–12.

[40] UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2021. Nucleic Acids Res 2021;49(D1):D480–9.
[41] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2019;47(D1): D941–7.
[42] Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human disease ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res 2019;47(D1):D955–62.
[43] Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The human phenotype ontology in 2021. Nucleic Acids Res 2021;49 (D1):D1207–17.
[44] Nelson SJ. Medical terminologies that work: the example of MeSH. In: 2009 10th International Symposium on Pervasive Systems, Algorithms, and NetworksIEEE. https://doi.org/10.1109/i-span.2009.84.
[45] Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, et al. The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res 2017;45(D1): D712–22.
[46] Amberger JS, Bocchini CA, Scott AF, Hamosh Aa. OMIM.org: leveraging knowledge across phenotype–gene relationships. Nucleic Acids Res 2018;47 (D1):D1038–43.
[47] Piovesan D, Necci M, Escobedo N, Monzon AM, Hatos A, Mičetić I, et al. MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res 2021;49 (D1):D361–7.
[48] Hornbeck PV, Kornhauser JM, Latham V, Murray B, Nandhikonda V, Nord A, et al. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. Nucleic Acids Res 2019;47(D1). https://doi.org/10.1093/nar/gky1159.
[49] Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell 2009;137(1):146–58.
[50] Paladin L, Schaeffer M, Gaudet P, Zahn-Zabal M, Michel P-A, Piovesan D, et al. The feature-viewer: A visualization tool for positional annotations on a sequence. Bioinformatics 2020;36(10):3244–5.
[51] Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3 (4):285–98.
[52] Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 2011;12(10):683–91.
[53] Kimchi-Sarfaty C, Jung Mi Oh, Kim I-W, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A 'silent' polymorphism in the MDR1 gene changes substrate specificity. Science 2007;315(5811):525–8.
[54] Matsumoto Y, Hayashi T, Inagaki N, Takahashi M, Hiroi S, Nakamura T, et al. Functional analysis of titin/connectin N2-B mutations found in cardiomyopathy. J Muscle Res Cell Motil 2005;26(6–8):367–74.
[55] Itoh-Satoh M, Hayashi T, Nishi H, Koga Y, Arimura T, Koyanagi T, et al. Titin mutations as the molecular basis for dilated cardiomyopathy. Biochem Biophys Res Commun 2002;291(2):385–93.
[56] Hackman P, Vihola A, Haravuori H, Marchand S, Sarparanta J, De Seze J, et al. Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. Am J Hum Genet 2002;71(3):492–500.
[57] Kawasaki Y, Sagara M, Shibata Y, Shirouzu M, Yokoyama S, Akiyama T. Identification and characterization of Asef2, a guanine–nucleotide exchange factor specific for Rac1 and Cdc42. Oncogene 2007. https://doi.org/10.1038/sj.onc.1210574.
[58] Shukla PC, Singh KK, Quan A, Al-Omran M, Teoh H, Lovren F, et al. BRCA1 is an essential regulator of heart function and survival following myocardial infarction. Nat Commun 2011;2(December):593.
[59] Liu J, Doty T, Gibson B, Heyer W-D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nat Struct Mol Biol 2010;17(10):1260–2.
[60] Mersch J, Jackson MA, Park M, Nebgen D, Peterson SK, Singletary C, et al. Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian. Cancer 2015;121(2):269–75.
[61] Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 2007;81(5):873–83.
[62] Yamaguchi K, Nagayama S, Shimizu E, Komura M, Yamaguchi R, Shibuya T, et al. Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis. Sci Rep 2016. https://doi.org/10.1038/srep26011.