

Perspective

From fuzziness to precision medicine: on the rapidly evolving proteomics with implications in mitochondrial connectivity to rare human disease

Khaled A. Aly,¹ Mohamed Taha Moutaoufik,¹ Sadhna Phanse,¹ Qingzhou Zhang,¹ and Mohan Babu^{1,*}

SUMMARY

Mitochondrial (mt) dysfunction is linked to rare diseases (RDs) such as respiratory chain complex (RCC) deficiency, MELAS, and ARSACS. Yet, how altered mt protein networks contribute to these ailments remains understudied. In this perspective article, we identified 21 mt proteins from public repositories that associate with RCC deficiency, MELAS, or ARSACS, engaging in a relatively small number of protein-protein interactions (PPIs), underscoring the need for advanced proteomic and interactomic platforms to uncover the complete scope of mt connectivity to RDs. Accordingly, we discuss innovative untargeted label-free proteomics in identifying RD-specific mt or other macromolecular assemblies and mapping of protein networks in complex tissue, organoid, and stem cell-differentiated neurons. Furthermore, tag- and label-based proteomics, genealogical proteomics, and combinatorial affinity purification-mass spectrometry, along with advancements in detecting and integrating transient PPIs with single-cell proteomics and transcriptomics, collectively offer seminal follow-ups to enrich for RD-relevant networks, with implications in RD precision medicine.

INTRODUCTION

An estimated 300 million people suffer from rare diseases (RDs) worldwide, and although every RD is uncommon by definition, there is anywhere between 5,000 and 8,000 of them, rendering RDs a frequently seen category in the clinic (Haendel et al., 2020; Nguengang Wakap et al., 2020). Therapeutic options for patients with RD are in retreat owing to the scarcity of information pertaining to disease driver genes, mutation-to-phenotype linkage, and the biochemical consequences of genetic alterations, as exemplified by differential protein levels in the impacted tissue, or pathologically rewired protein-protein interactions (PPIs) that trigger disease manifestations. In general, human disease research has gained traction in recent years owing to the rapid development of affordable genomic sequencing platforms, which initially came to light about four decades ago upon sequencing the mitochondrial (mt) genome (Anderson et al., 1981). Despite finding that mtDNA comprises only three dozen genes, mt are still interestingly involved in a myriad of pivotal activities in the living cells such as ATP generation, urea cycle, heme and pyrimidine biosynthesis, as well as complex cell fate mechanisms (Rajput et al., 2015). This strengthened the notion that nuclear-encoded genes produce additional proteins that directly or indirectly interact with mt in a manner that explains their versatile cellular roles. The term “mt-associated proteins” or “mt proteins” was thus coined, which alludes to proteins associating with mt, whether or not they are of mt origin.

The pool of mt proteins is well over 1,000 (Calvo et al., 2016; Maltby et al., 2015), and some are linked to RDs or other human diseases when altered. Mt are essentially home to the amphibolic tricarboxylic acid cycle that generates ATP by means of five membrane-associated mt protein complexes, termed complex I–V. Complexes I–IV are assembled from 79 to 81 protein subunits, with alterations in these subunits often associated with RDs. For example, mutations in MT-ND1 associate with Leber hereditary optic neuropathy as well as mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) (Blakely et al., 2005), whereas mutations in the mtDNA encoding subunit COXI-III (Schon et al., 2012) and in some of the nuclear DNA encoded subunits such as COX7B (Indrieri et al., 2012), COX6B1 (Massa et al., 2008), NDUFA4 (Pitceathly et al., 2013), COX6A1 (Tamiya et al., 2014), COX8A (Hallmann et al., 2016), and NDUFAF8 (Floyd et al., 2016) are associated with respiratory chain complex (RCC) deficiency. Similarly, alterations in SACSIN, a 4,579-aa mt protein of unknown function, is linked to autosomal recessive spastic

¹Department of Biochemistry, University of Regina, Regina, SK, Canada

*Correspondence: mohan.babu@uregina.ca
<https://doi.org/10.1016/j.isci.2020.102030>



ataxia of Charlevoix-Saguenay (ARSACS) (Engert et al., 2000). These representative RD examples demonstrate that mt dysfunction is connected to RD onset and progression, which is a common theme found in many other RDs.

Owing to the scarcity of RD patient samples, it becomes of considerable value to develop cutting-edge proteomic and interactomic platforms to analyze these handful of rare samples to the fullest possible extent and identify RD biomarkers for subsequent diagnostic developments and therapeutic interventions. Our timely review discusses these points, by predominantly focusing on the strides accomplished in the areas of proteomics and interactomics and how they offer promise in RD research. Here, we will focus on systems perspective into altered mt proteome and connectome, their linkage to RDs, as well as explaining how untargeted and targeted proteomics can improve our understanding of mt connectivity to RDs. We also dedicate a section to discuss low abundance and transient PPI detection tools with promise in RD research and another section on genotype-phenotype matchmaking tools. We will conclude with how the aforesaid approaches are shaping a new horizon in the era of RD precision medicine.

A SYSTEMS PERSPECTIVE OF MT PROTEINS IN RDS

Controversies exist on the exact number of mt proteins owing to the vague consensus regarding the best mt enrichment methodologies, reproducibility concerns, and varied resolution ranges of protein detection methods used to generate data. This prompted the creation of open access inventories of mt proteins that garner information from published research, with notable discrepancies due to the aforementioned reasons. On the upside, commonalities between various databases are beneficial for follow-up analysis of mt proteins associated with a particular disease of interest, based on the confidence stemming from different database consensuses.

To shed light on database overlaps and discrepancies, we have recently surveyed (Zilocchi et al., 2020) five established databases (Figure 1A and Table S1) and extracted their mt protein inventory, including MitoCarta2.0 (Calvo et al., 2016), Integrated Mitochondrial Protein Index (IMPI) (Smith and Robinson, 2019), UniProt (UniProt, 2019), Gene Ontology (GO) (The Gene Ontology Consortium, 2019), and COMPARTMENTS (Binder et al., 2014). Although databases such as IMPI and COMPARTMENTS share 1,197 mt proteins in common, MitoCarta2.0 and UniProt only share 877 mt proteins, with 647 proteins shared between all five databases (Figure 1A). These discrepancies reveal the need for advanced proteomic and interactomic data refinement tools and algorithms to filter out fuzziness that hinders human disease research progress. Aside from that and despite being relatively simple to understand as a standalone organelle, mt impairment can lead to numerous RDs affecting many body organs, including the brain, muscles, liver, and others (Wallace and Fan, 2009). Examples include the Rett syndrome, Leigh syndrome, Leber hereditary optic neuropathy, Kearns-Sayre syndrome, RCC deficiency, MELAS, as well as ARSACS, a rare neurodegenerative disorder, in addition to thousands of Mendelian disorders without given names, yet identified based on preliminary driver mutations.

Mt also play a role in rare forms of otherwise common neuropsychiatric and neurodegenerative diseases, including early-onset bipolar disorder, juvenile Parkinson's disease (PD), and familial Alzheimer's disease (AD). Such overarching involvement of mt dysfunction in numerous diseases suggests the presence of overlapping symptoms that pose roadblocks in the identification of disease-specific pathologies. For example, and to shed light on how altered mt proteins contribute to the onset and/or progression of RDs and which drivers overlap with more prevalent human diseases such as bipolar disorder (BD), we searched the DisGeNET inventory (Pinerio et al., 2017), the largest collection of gene-disease association database, and extracted all genes that associate with RCC deficiency, MELAS, and ARSACS, as three representative RDs, as well as the more prevalent BD to highlight common and distinct players between rare and more common human diseases.

We next overlaid the harnessed outputs with mt proteins found in MitoCarta2.0, UniProt, IMPI GO, and COMPARTMENTS. This revealed 114 unique mt proteins associated with all four diseases (93 mt proteins associated with BD, 12 with MELAS, 8 with RCC deficiency, and 1 with ARSACS; Figure 1Bi and Table S2), 2 of which (SOD2, POLG) were found to be common players in more than one of these RDs (Figure 1Bii). The antioxidant enzyme superoxide dismutase 2, SOD2, was found to associate with BD, RCC deficiency, and MELAS and not with ARSACS, whereas the DNA polymerase γ catalytic subunit (POLG) associates with RCC deficiency and BD (Table S2). These common players explain the potential for shared symptoms between



combined with mass spectrometry (BF-MS), affinity capture-MS/Western/RNA, two-hybrid, proximity labeling-MS, co-localization, co-purification, and co-crystallization (Figure 1C) were used to determine 2,700 reported PPIs to date (Figures 1C and 1D, Table S3). For example, SOD2 interacts with 42 other mt proteins. SOD2 plays a critical role in detoxifying superoxide, a by-product of the electron transport chain, by converting it into peroxide, thereby scavenging reactive oxygen species (ROS) and preventing cell death. Interaction partners of SOD2 include the mt aldehyde dehydrogenase protein, ALDH2, an oxidative stress protector, and superoxide dismutase 1, SOD1, which maintains fitness and downregulates apoptosis, among others (Figure 1D). This suggests that impaired housekeeping of ROS may be linked to RCC and MELAS. In accord with these data, it was found that antioxidant treatments alleviate MELAS manifestations (Pek et al., 2019). Furthermore, SOD2 interaction with the amyloid precursor protein (APP; Figure 1D) was found to be connected to mt-controlled ROS and APP processing, an otherwise altered phenomenon in brain diseases mediated by RCC deficiency (Area-Gomez et al., 2018; Wilkins and Swerdlow, 2017). Thus, analyzing publicly available datasets can offer insights into common mt protein drivers in different RDs. Also, POLG is linked to RCC deficiency and BD and interacts with 17 mt proteins (Figure 1D), suggesting that POLG variants may impact mt fitness with likely role in other rare diseases, making it a less attractive target for RD-specific biomarker discovery.

Similar analysis can be performed on the compiled dataset to analyze disease-specific RD PPIs, exploring their linkage to mt dysfunction and designing downstream experimental validations for targeted therapeutic development. Besides mt protein public inventories, RD-specific databases and domains are emerging, such as Genetics and Rare Disorder Information Center (GARD), National Organization for Rare Disorders (NORD), Orphanet, PatientsLikeMe, and Therapeutics for Rare and Neglected Disease (TRND) that serve as data repositories, including clinical trial data, and also offer other advocacy and support services. Thus, the ongoing update of these public domains is critical in filtering out false-positive disease PPIs of individual portals, by querying database commonalities to minimize false positives. This is in addition to developing advanced tools and disease models that reveal strong, weak, and transient PPIs, thereby expanding the current datasets and offering better resolution of experimental data as discussed below.

Information on PPIs stored in public databases largely depend on the reliability of methods used to generate or predict these data. Thus, RD PPI refinements hinge on using disease models that mirror the polygenic risk factors and pathophysiological features of the disease under study. Classical disease models include cell lines such as SH-SY5Y used to study ARSACS in which the causative gene, SACS, is knocked down, or knockout mouse models where SACS exons are replaced with inactivating cassettes (Girard et al., 2012). Alternately, mice can be subjected to genotoxic treatments to induce the disease state. Another *in vitro* front termed organ-on-chip is showing success in modeling metabolic coupling of endothelial and neuronal cells (Maoz et al., 2018). These models may not fully recapitulate RD pathophysiology and potentially yield fuzzy proteomics and/or interactomics data that hamper the efforts in pinpointing disease-specific PPI networks. Skin fibroblasts represent a promising alternative that offers the advantage of being isolated from the patient and can be propagated for longer times without exhibiting genetic alterations. One pitfall of using fibroblasts as disease models, especially in rare neurological diseases, stems from their specialized roles in the cell that are not related to brain signaling. Thus, it is more vital to develop suitable models of neurological diseases, such as differentiated neurons derived from induced pluripotent stem cells (iPSCs) (Cavalli et al., 2020).

iPSCs offer the resilience of differentiation into numerous neural cell types, leading to better purity, tissue specificity, and closer imitation of the brain biology. Analysis of mtPPI defects and interactome alterations derived from iPSC neurons enables the study of synaptic cleft maturation, mt energetics and dynamics, as well as bidirectional trafficking in a brain-like disease model. One drawback of linear iPSCs is their lack of three-dimensional (3D) neuronal complexity that requires the presence of other cell types, such as glia, to create the complex and mosaic nature of the human brain tissue. This led to the development of 3D brain organoids as better disease models that offer more inclusive features of the human brain (Huch and Koo, 2015). Nevertheless, proper selection of the disease model is key to fine-tune precision proteomic and interactomic outputs, for subsequent translational research.

To gain more information on the functional categories in which mtPPIs belong, a protein classification and GO annotation analysis was performed (Figure 1E and Table S4). For example, proteins in the mtPPI network were enriched in many mt processes such as mt organization, including mt disassembly

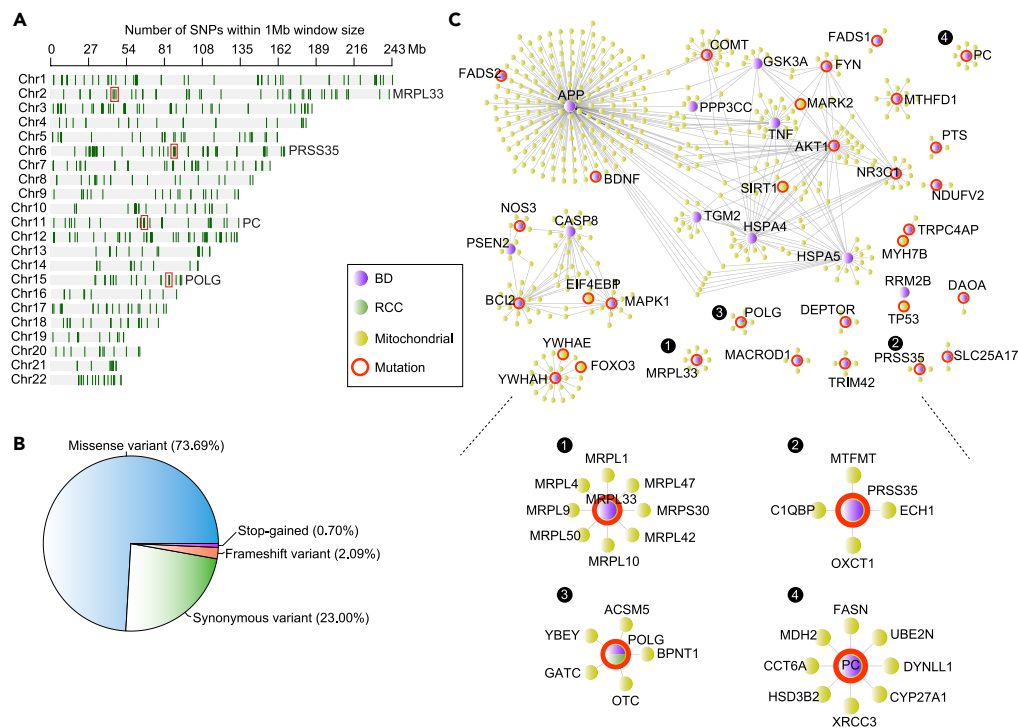


Figure 2. Bipolar disorder (BD) SNP variants

(A) BD SNP variants reported in the DisGeNET, within a window size of 1 Mb mapped to chromosomal localization in the genome. Four representative mt protein-encoding SNPs (MRPL33, PRSS35, PC, POLG) are highlighted in red boxes. (B) Prediction of BD SNPs on protein level. (C) Mt-associated BD SNPs mapped to the mtPPI network. MRPL33 (1), PRSS35 (2), POLG (3), and PC (4) interactions were clustered and illustrated.

(13 proteins), mt morphogenesis (4 proteins), mt membrane organization (46 proteins), mt RCC assembly (41 proteins), and mt genome maintenance (7 proteins). This is in addition to other processes such as lipid oxidation (14 proteins), redox process (159 proteins), and electron transport chain (52 proteins). Next, we focused on single nucleotide polymorphisms (SNPs) associated with various mt proteins and how they impact mtPPIs in connection with human disease. Owing to the scarcity of publicly deposited information pertaining to SNPs associated with RDs, and as proof of the usefulness of this approach, SNP variants belonging to BD were extracted from the DisGeNET database, along with their chromosomal location (Figure 2A). The consequences of those variants were summarized (Figure 2B), revealing 74% (423 of 574) are missense variants, 23% (132 of 574) are synonymous variants, 2.1% (12 of 574) are frameshift variants, and 0.7% of the SNPs led to a stop codon. Four representative mt SNPs with MRPL33 (mt ribosomal protein), PRSS35 (mt serine protease), POLG, and PC (mt pyruvate carboxylase) were mapped to our derived mtPPI network (Figure 2C), followed by clustering, suggesting the integration of these four mt proteins within small protein complex assemblies, such as the case with PC assembling a nine-membered protein complex upon clustering. This approach can prioritize the study of how these partner proteins are altered in a specific disease of interest, such as BD, for better understanding of the biochemical attributes underlying various disease phenotypes to catalyze faster therapeutic efforts.

UNTARGETED PROTEOMICS OF THE MT PROTEIN INTERACTOME IN RDS

These coarse PPI mappings through untargeted proteomics described below are suitable for early identification of rare and other disease biomarkers and protein complex discoveries. Of note, untargeted proteomics are tailored to analyze intricately mosaic samples such as miniaturized organoids and iPSC-differentiated neurons from healthy subjects and patients. These include BF-MS, cross-linking coupled with MS (XL-MS), MS-based aggregate contactome mapping, and label-free sequential window acquisition of all theoretical fragment-ion spectra MS (SWATH-MS). Untargeted proteomic approaches can thus be beneficial for rapid identification of potential RD biomarkers, with ensuing follow-ups by targeted methodologies.

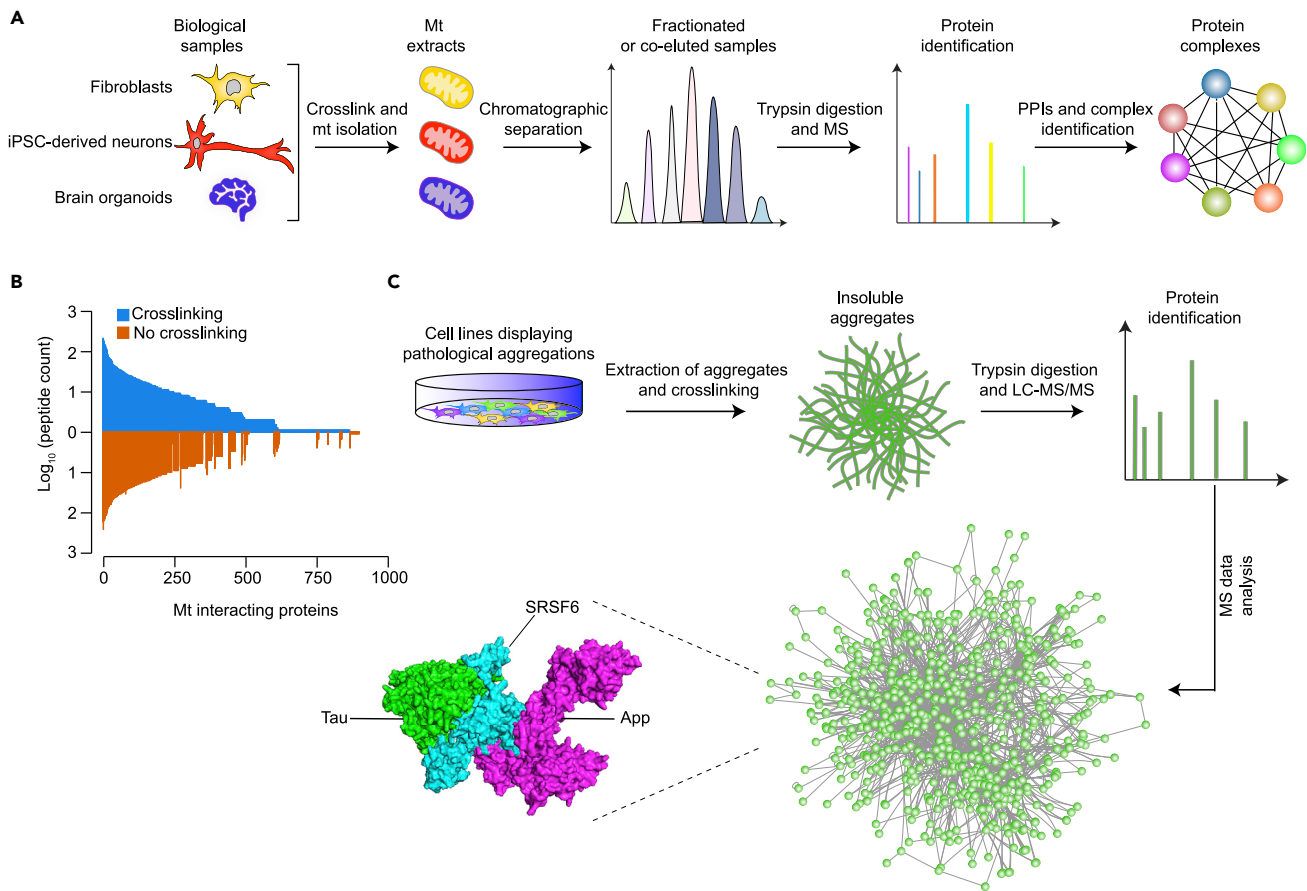


Figure 3. Disease-specific MtPPIs by BF-MS and aggregate contactome mapping

(A) BF-MS in the identification of disease-specific mtPPIs.

(B) Weak and transient mtPPIs captured in the presence of a cross-linker and detected by MS.

(C) Insoluble aggregates are extracted from organism models or cell lines before processing by MS to identify PPIs. Protein-protein docking between SRSF6 (cyan), Tau (green), and APP (magenta) is illustrated.

BF-MS to identify RD-specific Mt interactions

BF-MS has been instrumental in recent years to capture strong, weak, and transient PPIs from mt and other biological samples that have passed quality control assessment for purity validation (Havugimana et al., 2012; Moutaoufik et al., 2019; Pourhaghighi et al., 2020; Wan et al., 2015). Depending on the nature of the experiment, patient fibroblasts, iPSCs, organoids or disease-specific tissues can be analyzed by BF-MS. If mt dysfunction is the topic of concern such as in RCC deficiency, MELAS, ARSACS, and other RDs, iPSCs are subjected to chemical cross-linking early in the BF-MS approach, such as using dithiobis (succinimidyl propionate), a lipophilic chemical engineered to exhibit amine-reactive NHS ester at each end of the spacer, while lacking charged groups, thus able to permeate biological membranes and capture PPIs involving membrane as well as other soluble proteins (Figure 3A). This strategy yielded notable success when versatile-affinity (VA)-tagged bait membrane proteins from HEK293 cells containing 3x-FLAG, 6x-His, and StreptIII epitopes were subjected to pull-down attempts with or without cross-linking, followed by MS detection that revealed that the presence of a cross-linker has favored the capturing of numerous weak or transient mtPPIs (Malty et al., 2017), thereby validating the robustness and effectiveness of the BF-MS approach in the presence of dithiobis as mild cross-linker, to minimize possible artifacts (Figure 3B). A similar experimental setup can be performed on mt isolated from the neurons of patients with RD, followed by capturing RD-specific cross-linked neuronal mtPPIs when compared with healthy controls.

Mt extracts are subjected to biochemical fractionation via two or more chromatographic techniques (e.g., multibed ion exchange, size-exclusion, isoelectric focusing, sucrose density gradient centrifugation),

followed by multiplexing and MS analysis. MS-spectra of co-eluting proteins can next be searched against human reference sequences such as X!Tandem, Comet, and MS-GF⁺ engines to increase the chance of identifying the co-eluting proteins and subsequently plotting PPI maps for patient samples versus healthy control counterparts for comparisons. BF-MS has been recently used to study how PPIs are rewired during neuronal differentiation from stem cells (Moutaoufik et al., 2019). This powerful platform can spot gain-of-function RD-specific PPIs as potential disease biomarkers, which can be targeted by small molecular inhibitors or peptide therapeutics that disrupt pathological PPI interfaces. This is particularly relevant when gain-of-function PPI interfaces can be tracked using 3D reference coordinates for docking attempts to interrupt dimer interfaces, especially upon the discovery that disease alleles with mutations can be reflected in PPI dimer interfaces (Sahni et al., 2015). Of note, some PPIs remain existent in both healthy and patient disease maps, yet the composition and stoichiometry of these PPIs may deviate from the norm. These unusual patterns mediated by mutation-driven variants may also be of therapeutic value but require supplementary methodologies such as fluorescence fluctuation spectroscopy, absolute quantification, and multiple reaction monitoring to validate their disease specificity for robust therapeutic interventions. Thus, BF-MS is an attractive interactome approach that scans system-wide PPIs to identify alterations from the norm that can be shortlisted for targeted validations, with implications in mt connectivity to RDs.

XL-MS to detect RD-related Mt protein with inter/intra-XL peptides

Despite the initial intent in using chemical cross-linking combined with MS (XL-MS) for the painstaking detection and quantification of PPIs in various biological specimens, XL-MS has rapidly expanded into other exciting territories, such as deducing medium-resolution protein structures, macromolecular complex topologies, and conformational state quantifications (Liu et al., 2015; O'Reilly and Rappsilber, 2018). Interacting proteins in close proximity are covalently bonded by a chemical cross-linker that links surface residues of the interacting protein partners, thereby probing for surface-exposed domains, which indirectly identifies hydrophobic protein patches within a macromolecular complex. By examining distance restraints upon considering the most extended conformation of the cross-linker in comparison with various cross-linked residues, protein conformational states and complex topologies can be robustly resolved (O'Reilly and Rappsilber, 2018). This area is also rapidly building upon using classical lysine-lysine into the development of aromatic glyoxal cross-linkers to covalently link arginine, thereby capturing interactions that are lysine deficient (Jones et al., 2019). Also, cross-linking-integrated workflows that couple MS-cleavable cross-linkers and dual fragmentation strategy (i.e., sequential collision-induced dissociation and electron transfer dissociation) to identify cross-links against full proteome databases are being consistently developed and refined (Liu et al., 2017).

The role of XL-MS in unveiling the scope of mt linkage to RDs is gaining attention owing to the involvement of mt dysfunction in brain and heart and aging complications. XL-MS was applied to identify 2,427 cross-linked peptide pairs linking 459 mt proteins, and RCC proteins were found to engage in a network of inter-complex interactions that mediate ATP synthesis, thus validating the higher-order mt respirasome assembly hypothesis, which results from the interaction between individually intact RCCs (Schweppe et al., 2017). Qualitative and quantitative identification of inter and intra cross-linked lysine residues, along with C α -C α distance considerations, has enabled XL-MS to deduce RCC complex topologies and offer structural insights that are validated by existing cryo-EM and crystallographic data (Schweppe et al., 2017). For instance, several homo-multimeric cross-links in the respiratory chain complex III were detected and align with established structural data pertaining to their role in ATP synthesis (Schweppe et al., 2017). Furthermore, lysine cross-link mapping of complex V associations has robustly identified interaction sites and the stoichiometry of complex V rotor, stator, and ATPA/B constituents, along with explaining their structural orientation within a functional context.

XL-MS thus complements structural approaches to understand mt protein complex interactions and topologies with implications in RDs, which can benefit the area of mt drug discovery. Restoring mt fitness and consequently reversing disease manifestations is becoming a viable route owing to recent advancements in which XL-MS has been seminally deployed. The synthetic tetrapeptide, elamipretide (SS-31), was found to improve mt fitness, and hence dubbed the term "mt therapy" (Chavez et al., 2020). XL-MS has unlocked the biochemical attributes behind phenotypic improvement in mt function mediated by SS-31, upon mapping the network of SS-31 mt interacting partners, which were found to fall into two categories: oxidative phosphorylation proteins that mediate ATP production and a second group involved in 2-oxoglutarate metabolism. Protein interactors with SS-31 from both groups are common binders of the mt cardiolipin, an inner mt membrane (IMM) lipid that constitutes ~20% of the total IMM lipid content. This confirms

SS-31 specificity for IMM and maps the interacting partners coordinating mt fitness improvement by SS-31 therapy. The working model of how SS-31 improves mt function suggests that it engages in mtPPIs that induce tighter cristae curvatures in a manner that stabilizes the IMM to optimize respirasome assembly and function. This feat in advancing mt therapy can be applied to mt dysfunction studies in RD research, in which XL-MS in the presence of SS-31 from patient samples can reveal disease-specific SS-31-linked PPIs that are gained or lost, thereby linking disease manifestations to underlying biochemical defects, which can refine the design of more targeted mt peptide therapeutics for RD treatments.

MS-based aggregate contactome mapping of Mt proteins in RDs

Contactome mapping is an emerging untargeted proteomics approach in neurodegenerative disease (ND) research, with clear potential in probing mt protein interactions in RDs for biomarker discovery and subsequent therapeutic intervention. This concept relies on the observation that “seed” proteins are often observed in ND protein aggregates, and although aggregate compositions in any ND may vary from one patient to another, it appears that seed constituents of these aggregates are non-random. For example, A β ₄₂ and tau are persistently found in AD aggregates as an AD pathological hallmark. This suggests that seed constituents likely coordinate common and may also mediate preferential PPIs within aggregate interfaces in different patients, which can be traced to discover individual-specific biomarkers for blockage using aggregate dispersal therapies (Balasubramaniam et al., 2019). Anti-aggregate therapeutics aim to diffuse these pathological “protein lumps” in AD and other neurological diseases.

Aggregate contactome mapping can be achieved through the so-called “click-chemistry” concept that involves a chemical cross-linker to stabilize physiological interactions within the aggregate vicinity (Figure 2), and subsequent MS-detection to reveal aggregate hubs, while additionally identifying the scope of hub-hub PPIs within the aggregate contactome. Briefly, the aggregate contactome starts with collecting protein aggregates by low-speed centrifugation from cell line disease models, such as the human neuroblastoma SH-SY5Y cell line expressing the Swedish familial-AD mutant of APP, and aggregate incubation in the presence of a non-denaturing detergent to solubilize aggregate protein interacting partners. This is followed by chemical cross-linking, tryptic digest, and LC-MS/MS analysis to detect cross-linked peptide spectra. This method has successfully revealed 506 contactome protein hubs that are involved in >7,000 interactions, with ~6,000 interactions deemed specific to the disease cell line model as opposed to the non-mutant counterpart (Balasubramaniam et al., 2019), which highlights the complexity in protein aggregate compositions in AD. For illustration, we used molecular docking to predict the interaction interface between the reported interacting proteins SRSF6, Tau, and APP for the subsequent development of aggregate dispersal drugs (Figure 3C). Experimentally, gene knockdowns in select PPIs that are abundant within or critical for contactome maturation can be constructed, followed by measuring aggregate dispersal patterns for therapeutic intervention using aggregate dispersal drugs that target select PPIs mediating aggregate formation.

Aggregate contactome mapping for subsequent dispersal using anti-aggregate therapeutics can be applied for RDs known to form protein aggregates as pathological disease hallmarks, such as the case with lysosomal storage disease (LSD). LSD is caused by an inborn error of metabolism that leads to abnormal material storage in body cells, and this inherited disorder remains without definitive cure. Thus, analysis of contactome interfaces within aggregates that may trigger LSD can significantly improve the understanding of this RD and enhance the identification of novel LSD biomarkers and anti-dispersal drug designs (Monaco and Fraldi, 2020). Similarly, amyotrophic lateral sclerosis (ALS), a mt dysfunction-driven RD that impacts the life of one in every 100,000 people in the United States, is accompanied by the aggregation of ubiquitinated proteins in motor neurons (Gill et al., 2019). ALS remains among the poorly managed RDs, and contactome mapping of mt protein aggregates isolated from ALS patient samples or disease models mimicking ALS can reveal seed hub interactions that can be targeted by aggregate dispersal drugs. This approach can also be applied to an arsenal of other RDs such as cerebral cavernous malformation, leptomeningeal amyloidosis, inflammatory myopathies, and myotonic dystrophy syndromes, which commonly manifest in protein aggregate formation as pathological hallmarks, revealing the potential for aggregate dispersal studies and interface contactome mapping in opening new avenues of biomarker discovery and anti-aggregate therapeutic discovery for select RDs.

Label-free SWATH-MS to uncover Mt biomarkers for RDs

SWATH-MS has achieved important strides in RD proteomics. Data independent acquisition (DIA) via SWATH introduces an interesting concept that considers all possible theoretical fragment ion spectra as

a starting point and reversibly concludes parent ions. In SWATH-MS, label-free protein samples are digested with trypsin, and the collection of digested peptides is subsequently analyzed in a DIA mode, where all ionized peptides within a certain mass range are fragmented using overlapping windows of ~ 25 m/z each (Gillet et al., 2012; Ludwig et al., 2018), thus enabling re-interrogation of data upon improving detection capabilities that expands the list of identified proteins in various clinical samples. This is unlike the case with information-dependent acquisition, where low protein abundance data are lost permanently.

SWATH-MS style of re-examining co-fragmentation patterns of the co-eluting peptides creates a digital biobank for each sample and offers orders of magnitude increase in large-scale quantitative proteomics with great precision, while addressing data reproducibility concerns. This qualifies SWATH-MS as a high-throughput quantitative proteomics tool that can be of particular value in biomarker discovery for RDs (Luo et al., 2017). In fact, almost half of SWATH-MS applications were applied to the area of clinical proteomics and biomarker discovery research (Narasimhan et al., 2019). Unlike classical selected or multiple reaction monitoring approaches, SWATH-MS is more useful in RD biomarker discovery attempts such as the case with nasopharyngeal carcinoma, where carbonic anhydrase 2 was identified as a reliable biomarker for this rare cancer type (Luo et al., 2017). Similarly and owing to the advantageous edge of SWATH-MS over classical tools in its ability to analyze differential proteomes among diverse cohorts, it was successfully applied for analyzing system proteomes of liver mt (Williams et al., 2016), with potential implications in RDs in which mt dysfunction connectivity to rare liver diseases is relevant, such as the case with α -1 antitrypsin deficiency (Teckman et al., 2004). On the brain disease front, SWATH-MS offers promise in studying mt energetics and their linkage to RDs. For instance, and using SWATH-MS, differential expression of the presynaptic mt proteome at various developmental stages of neuronal synapses has been achieved as a proof of concept (Stauch et al., 2019), tendering great value in RDs in which neuromuscular synapses are altered in connection to mt dysfunction, such as the case in ALS, ARSACS, and other RDs where altered neuronal mt accelerate disease manifestations.

In addition, secretome profiling in rare cancer types, such as pancreatic cancer, can be achieved by SWATH-MS, which was applied for differential secretome mapping between two variants of the TP53 transcriptional regulator to fine-tune variant-specific biomarker discovery outputs (Butera et al., 2020). Another important yet significantly understudied area of clinical research is metabolomics and their linkage to human disease. A comparative profiling study of both human serum and plasma using SWATH-MS showed subtle variations in proteomic profiles between these samples but revealed significant metabolomic differences, which sheds light on how sample handling and storage can impact blood metabolome profiles, thereby questioning the reliability of putative disease biomarkers (Nambu et al., 2020). More recently, patient saliva is becoming one of the most attractive body fluids to collect and analyze as patient-convenient and a non-invasive approach in disease diagnostics. However, the quality of sample preparation can drastically impact positive testing outcomes, and SWATH-MS is being gradually applied as the label-free method of choice to assess saliva sample preparation quality (Zhang et al., 2020).

Although SWATH-MS offers promise in numerous clinical research and drug discovery applications, intra-laboratory data precision, sensitivity, selectivity, and reproducibility can also be achieved by SWATH-MS, which was recently performed on 11 different laboratory sites worldwide and consistently quantified >4,000 high-quality proteins in HEK293 cells, revealing the potential of SWATH-MS as a method of choice for reproducible quantitative proteomics across various research groups (Collins et al., 2017). Owing to its considerable selectivity and sensitivity, SWATH-MS stands as a powerful frontline tool in studying mt energetics and their linkage to RDs such as ARSACS and ALS as explained in the previous sections by having a precision edge when compared with shotgun proteomics.

TARGETED PROTEOMICS IN MT PROTEINS OR INTERACTOME IN RDS

Targeted proteomics are suitable to zoom in on a subcellular or suborganellar scale. This is exemplified by peroxidase-mediated covalent tagging of proximal endogenous proteins in the neuronal synaptic cleft or proximity labeling to spatially resolve neuronal mt protein interaction maps at pre- and post-synaptic termini in patients with RD versus healthy counterparts. In fact, it is beneficial to map mt interactome alterations in RDs in which neuronal activity is perturbed at terminal synapses, such as in Huntington's disease. A collection of targeted proteomics approaches and their benefit in mt research in connection with RDs are discussed below.

Triple-SILAC genealogical proteomics to unveil RD mechanisms

One obstacle that impedes the understanding of rare neurological diseases is the lack of knowledge on how genetic defects are linked to observable disease phenotypes. Classically, MS-based methods are pursued to map proteome changes in patient cell types, but without knowing the scope of how the underlying genetic alterations are responsible for patient-specific pathological protein levels. This can be overcome by genome sequencing of patient versus healthy samples to reveal SNP existence in different loci. However, such an approach does not fully resolve the puzzle, since it is difficult to sort out which SNP(s) drive(s) patient proteome alterations when compared with healthy subject maps, since many SNPs may exist in patient genomes and any or a combination of them may be responsible for the phenotype observed. Genealogical proteomics contribute to resolving this roadblock by comparing patient proteome profiles with consanguineous healthy subjects instead of non-relative healthy individuals as a reference point for data filtration (Zlatic et al., 2018). This enhances refining disease-specific protein levels, since healthy and patient subjects belonging to the same family line may share common disease-irrelevant SNPs and also exhibit protein accumulation or deprivation commonalities that are not disease specific. This can minimize false positives compared with non-relative healthy controls.

Genealogical proteomics were recently applied on Menkes disease, an X-linked rare monogenic childhood ND caused by alterations in the copper transporter ATP7A, leading to an improper distribution of copper exemplified by its accumulation in the kidneys and the intestine, while being deficiently supplied to the brain cells. Within a family, patient fibroblast proteome profiles were compared with their healthy kin using triple-SILAC (Figure 4A). Wild-type cells were supplemented with “light” unlabeled ^{12}C - and ^{14}N -arginine and lysine amino acids, while ATP7A null cells were incubated with either “medium” ^{13}C -arginine and ^2H -lysine or “heavy” ^{13}C - and ^{15}N -tagged arginine and lysine-containing media for isotope amino acid incorporations (Zlatic et al., 2018). A total of 214 non-redundant proteins exhibited increased or decreased production in patients versus healthy kin, including 15 mt proteins, along with enzymes that require ATP7A for copper loading into the Golgi complex, and a collection of other copper metabolism enzymes. These triple-SILAC derived 214 differentially abundant protein datasets were queried against the mouse phenotype KEGG pathway database to assess their linkage to phenotypic traits. Of interest, the proteome query captured Menkes phenotypes such as abnormal skin tensile and altered liver morphology, which KEGG associates with *Atp7a* mouse mutations, with no phenotypic overlaps when unrelated datasets were queried against the database.

Triple-SILAC genealogical proteomics can be of particular value in RDs in which mt altered protein profiles drive disease onset and/or progression. Since an altered mt proteome in patient samples may or may not be directly linked to the RD under study, mt proteome signature heterogeneities between different patient samples confine narrowing down which of the mt proteins (Table S1) are linked to the RD disease of interest. Triple-SILAC genealogical proteomics can shortlist mt disease drivers by comparing mt proteome signatures of RD patient versus healthy kin, who may exhibit proteomic alterations that are unrelated to the RD under study. Although genealogical proteomics uncover distinct proteome profiling in patient samples, applying the same concept for interactome mapping within a pedigree can robustly filter out spurious PPIs associated with family=unrelated healthy subjects from the comparative datasets.

Biotin proximity labeling to map mtPPIs or complexes in RDs

Originally implemented as a complementary approach to affinity purification for PPI detections, earlier developments of the BioID (proximity-dependent biotin identification) technology have conceptually resembled yeast two-hybrid pairwise PPI analyses, by fusing the *Escherichia coli* biotin ligase (BirA) to the bait of interest, while fusing the suspected partner to the biotin acceptor peptide (BAP). If the bait and prey interact, BAP is brought in proximity with BirA, an enzyme that catalyzes the conversion of inactive biotin into an activated molecule termed biotinoyl-5-AMP (Li et al., 2019). Upon biotin activation, BirA retains the active product, which can still label the BAP in close proximity (Figure 4A). Passing the biotin-labeled acceptor over a streptavidin affinity matrix can capture fused prey proteins for MS analysis. Subsequent discovery of the BirA variant (R118G; BirA*), which can no longer retain activated biotin, has advanced BioID through the development of the so-called promiscuous proximity biotinylation.

In the updated version (Kim et al., 2014), BirA* remains fused to the bait of interest and inactive biotin can be catalytically converted into a non-retained active form that in this case biotinylates interacting protein partners within 10 nm of the bait vicinity. Properly conformed interacting partners with the bait can be

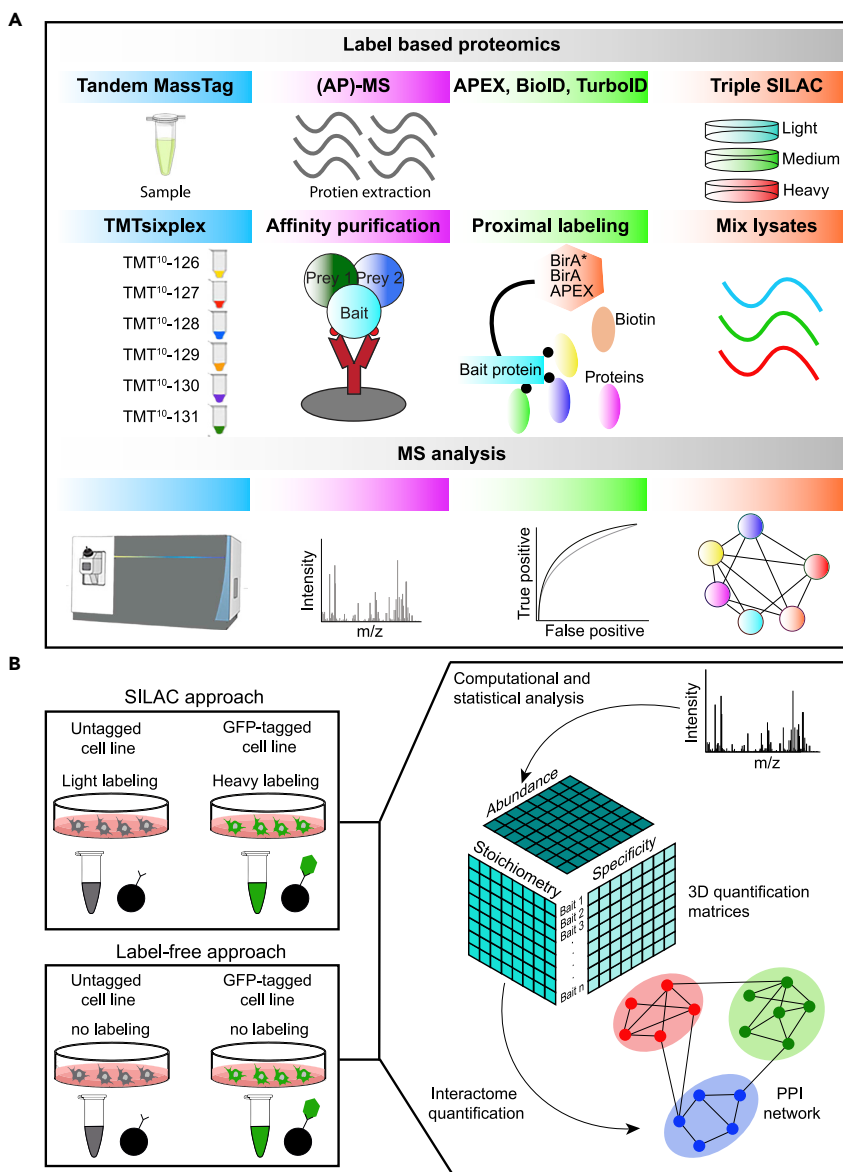


Figure 4. Targeted proteomic and triple quantification approaches

(A) Illustration of label-based tandem mass tag (TMT) peptide labeling, combined with affinity purification (AP)-MS/BioID (MAC-tag), biotin proximity labeling approaches (APEX, BioID, TurboID) and triple-SILAC genealogical proteomics for interactome mapping.

(B) QUBIC-based triple quantification (abundance, stoichiometry, and specificity) interactome using SILAC and label-free approaches. Proteins were affinity purified prior to MS analysis, followed by three-dimensional quantification. Proteome quantification of both stoichiometry and specificity of bait proteins as well as abundance of their interactors is outlined.

labeled without peptide fusions, and this development also no longer investigates the interaction between baits and one fusion protein at a time. The collection of biotinylated interaction partners is labeled, passed over streptavidin, and then identified by MS. BioID has a higher tendency to capture transient PPIs and thus is suitable for targeted proteomics. Furthermore, disease-linked proteins contain intrinsically disordered regions (IDRs) that may undergo posttranslational modifications (PTMs), rendering unfolding and refolding domain properties critical for understating human disease attributes related to IDR-containing proteins.

Targeted *in vivo* BioID was recently shown to effectively label >20,000 biotin sites in predicted IDRs of proteins localized to several subcellular compartments, making BioID an important approach in studying RDs

(Minde et al., 2020). X-linked severe combined immunodeficiency (X-SCID) is a rare immune disease, affecting 1 in 100,000 males at birth. Mutations in the interleukin-2 receptor gamma chain (IL2RG) are the major cause of the disease, but the scope of biochemical alterations coordinating disease pathophysiology remain poorly studied. Targeted BioID showed for the first time that IL2RG is involved in aberrant interactions with the endoplasmic reticulum and Golgi proteins that result in IL2RG mislocalization, which contributes to progressive disease manifestations (Tuovinen et al., 2020). The BioID technology has also been implemented in mt research, with notable potential in mt connectivity to RDs. A high-density BioID-based human mt proximity interaction network has been elucidated upon interrogating 100 different mt baits representing various mt sub-compartments, which led to the identification of 528 mt proteins engaged in thousands of high-confidence proximity interactions (Antonicka et al., 2020). This approach is beneficial on two fronts: first, the characterization of mt proteins of unknown function, and second, mapping these mt interactions in RD patient samples to highlight loss- or gain-of-function interactions of therapeutic value.

Conceptually similar to the BioID, APEX (ascorbic acid peroxidase) catalyzes the conversion of biotin-phenol to the biotin-phenoxyl radical when hydrogen peroxide is supplied to mediate peroxidase-based proximity labeling. BioID and APEX have both been more recently subjected to protein fragment complementation analysis, termed split-BioID and split-APEX. It was discovered that BirA or APEX can be split into N- and C-terminal fragments that are re-constituted by fusing each fragment to a protein of interest (De Munter et al., 2017). If the two fusion proteins interact, they bring BirA or APEX fragments to close proximity (Figure 4A), thereby resuming active biotinylation or peroxidation. This leads to labeling other partners interacting with the two fusion proteins within a proximal space, which has profoundly advanced interactomics research for protein complexes that require conditional assembly, such as partner-dependent macromolecular complex formation (Trinkle-Mulcahy, 2019). Recently, the mt nucleoid has been attracting some focus, and a derivative of the APEX method, termed Twinkle-APEX2, was applied to fuse target peroxidase to mt nucleoids in human fibroblasts. This led to enriching for mt nucleoid proteins, nearly half of which have not been captured using classical APEX and/or immunoprecipitation studies (Han et al., 2017). TurboID is a more recent variant of BirA that was engineered to create 14 *birA* mutations to increase biotin proximity labeling in 10 min as opposed to 24 h when compared with BioID (Branon et al., 2018). The advantage of APEX over BioID or its TurboID derivative is in the labeling time it requires. APEX labeling is complete in 1 or 2 min, as opposed to 24 h or 10 min in BioID or TurboID, respectively. This improves the temporal resolution and sheds light on compartmental PPI dynamics.

In addition, there is growing conviction that RDs are linked to altered microRNA (miR) levels in various body fluids and tissues. For instance, Duchenne muscular dystrophy (DMD) is a lethal X-linked rare disorder caused by mutations in dystrophin that encodes a cytoskeletal protein (Salvatore et al., 2011). Serum levels of muscle-specific miRs are altered in DMD, and similar findings have been reported in ALS. When normal mice become deficient in miR-206, they develop normal neuromuscular synapses, whereas ALS mice deficient in the same miR showed significant disease acceleration (Salvatore et al., 2011). However, the scope of protein complexes regulating miR expression in RDs remains elusive. Split-BioID was recently used to probe protein complexes that regulate miR-mediated translation repression, which mapped protein complexes spatiotemporally mediating this process in native cellular environments (Schopp et al., 2017), offering a promising tool in biomarker discovery for RDs.

Combined affinity purification-MS and BioID for interactome mapping in RDs

Although affinity purification (AP)-MS and BioID coupled with MS (Figure 4A) can both capture physiologically relevant protein-specific interactomes, BioID or APEX may be better suited to capture weak or transient PPIs owing to rapid untagged labeling of prey protein partners. Recently, combining both AP-MS and BioID has been attempted with notable modifications (Liu et al., 2018). AP-MS relies on fusing the protein of interest to an epitope tag followed by capturing the endogenous bait with an affinity matrix specific to the fusion tag (e.g., GFP, HA, 6x-His, Strep, MYC, or FLAG). Affinity-purified baits are processed by MS to identify interacting partners, which has been successful in high-throughput and targeted studies. Although BioID is becoming more competitive in capturing weak interactions, a recent Gateway-compatible MAC (Multiple Approaches Combined) tag has enabled the usage of both techniques via a single construct. The ORF of interest is cloned on Strep- and HA-encoded vector as a double fusion tag, where the vector also encodes the R118G BirA variant (BirA*).

Isogenic cell lines propagating the vector are divided into three groups; the first is supplied with biotin for catalytic conversion into active biotin, thereby labeling PPI partners in close proximity for subsequent MS analysis. The second batch is subjected to classical AP-MS upon direct purification of the Strep-tagged protein, and the third batch is visualized by fluorescence microscopy using Alexa Fluor 488-labeled anti-HA immunostaining for protein localization determination. Of interest, biotinylated partners can also be visualized using Alexa Fluor 594 streptavidin, and when combined with Alexa Fluor 488-labeled anti-HA, co-localization experiments can be performed. The HA epitope can also be used for chromatin immunoprecipitation (ChIP)-seq applications and validation of the cross-linked interacting partners by XL-MS, making the MAC-tag strategy as versatile as the Swiss Army knife (Liu et al., 2018). To benchmark the utility of the MAC-tag concept, it has been validated for 18 *bona fide* subcellular and 3 sub-organellar markers including mt, endoplasmic reticulum, and peroxisome with proper localization outcomes, making it a powerful tool for consideration when studying compartment-specific RD interactome profiling.

Tandem Mass Tag peptide labeling of spatial proteome in RDs

Dynamic proteome analysis of localization and translocation events in tissue-specific cell types can be a daunting process that requires comparative quantification, while minimizing artifacts. Label-free quantification (LFQ) and tandem mass tag (TMT) multiplexing workflows have been attempted to uncover the spatial proteome of mice primary neurons. Spatial proteome is an emerging proteomics area that attempts to map the localization of all cellular proteins in a systems-wide panoramic context (Christoforou et al., 2016; Rhee et al., 2013). This principle starts with the biochemical separation of the organelle of interest, followed by quantifying protein distributions in various sub-fractions. When those quantifications are compared with known marker proteins, subcellular localization assignment of unknown others becomes feasible.

Earlier attempts relied on the differential centrifugation of cell lysates into six different fractions, while mixing each fraction with a SILAC heavy “reference” membrane fraction, followed by MS analysis of heavy to light ratios to yield abundance profiles across different gradients (Itzhak et al., 2016). In the LFQ updated workflow version, the heavy labeled reference step was removed and abundance profiles were compared with a reference computational algorithm (Itzhak et al., 2017). Acute neurons were isolated in replicates from mice and subjected to the LFQ workflow upon differential centrifugation, followed by comparing protein abundances to the reference machine learning-based algorithm to localize various proteins into 12 clusters pertaining to different organelles and neuronal compartments such as the cytosol, mt, membrane, and nuclear proteins. Quantification can also be performed by chemical labeling upon digesting fractionated proteins with trypsin, followed by conjugation with a TMT reagent (Figure 3). Although each tag has an identical mass, proteolytic fragmentation essentially yields reporter ions with different masses that can be used to quantify parent peptides.

Recently, 6- and 10-plex TMT (Figure 4A) have enabled MS analysis of several fractions in a single MS run (Nusinow et al., 2020; Pourhaghighi et al., 2020). Exploring both label-free and label-based TMT approaches shows that both are reliable and consistent in annotating cell proteomes to relevant compartments, with TMT labeling providing comparable outputs in capturing select protein translocations. The advantage of TMT is multiplexing, which enables comparative analyses even of triplicate samples in less than 9 h of MS time. TMT isobaric labeling was used to identify key proteome commonalities that drive AD and PD in postmortem brain tissues (Ping et al., 2018), with similar experimental setups being feasible in RD research.

Binary protein interactome mapping with implications in Mt connectivity to RDs

Binary protein interaction experimental platforms have grown considerably in recent years, with exceptionally well-designed methodological approaches (Stelzl et al., 2005; Venkatesan et al., 2009; Woodsmith et al., 2017). Owing to space constraints, we will discuss representative examples, along with their potential implications in RD research. Among these methods, the Mammalian Membrane Two-hybrid (MaMTH) offers interesting applicability in understanding the scope of RD connectivity to mt alterations. The MaMTH is based on fusing a membrane protein of interest to the C-terminus of ubiquitin (Cub) along with a spacer and an artificial transcription factor (TF), while the prey of interest is fused to the N terminus of ubiquitin (Nub). Interaction between bait and prey excises the TF, which in turn activates a reporter system that can be recorded. Pairwise interaction analyses between mt and RD proteins of interest can annotate mt proteins associated with RDs and not with healthy subjects (Petschnigg et al., 2014).

Another modification to this two-hybrid approach is termed Barcode Fusion Genetics-Yeast Two-Hybrid (BFG-Y2H), in which protein matrix pairs are screened in a single multiplexed strain. In this method, DNA barcodes are fused using Cre recombination, yielding chimeric protein-pair barcodes for next generation sequencing, which has demonstrated significant improvement in capturing protein interactions (Yachie et al., 2016). This two hybrid-based approach is easily scalable in a high-throughput format, which is also a common theme in Y2H-based methodologies (Yachie et al., 2016). For example, Y2H was successfully implemented in a high-throughput manner, along with orthogonal validations, to identify ~14,000 PPIs involving 4,000 human proteins. Interactome coverage was further increased by expanding the ORFeome collection to encompass ~90% of the protein-coding genome, thereby quadrupling the PPI maps to uncover interactions associated with many cellular processes (Luck et al., 2020). Depending on the choice of baits, the same process can be applied to study mtPPIs in RDs that are either gained or lost, when compared with healthy counterparts.

LOW ABUNDANCE AND TRANSIENT MT PPI DETECTION METHODS FOR RDS

This two-edged obstacle is subjected to ongoing investigations, not only to refine methodologies for capturing low-abundance proteins and transient mtPPIs but also to link abundance thresholds to pathway functions in healthy and/or diseased cell populations, with potential implications in altered mt proteomes in connection to RDs.

QUBIC-based triple interactome in three quantitative dimensions

Quantitative bacterial artificial chromosome (BAC)-green fluorescent protein (GFP) interactomics (QUBIC) has been attempted for human interactome mapping that captures low-abundance and transient PPIs with great efficiency (Hein et al., 2015). Two factors play a critical role in interactome mapping: strength of interactions and the abundance of interacting proteins, both of which can impact the output based on the mapping method used. Thus, interaction networks are the collective result of binary affinities and cellular distributions of all proteins. QUBIC has been implemented to offer three quantitative interactome dimensions: detecting specific interactions via local co-enrichment profiles, interaction stoichiometries, and cellular abundance of the interacting proteins (Figure 4B), which can uncover weak interactions in orders of magnitude when compared with established methods. The QUBIC approach led to the discovery that the human interactome is dominated by weak and transient PPIs (Hein et al., 2015), which can be suitable to study dynamic changes in organelles such as the mt.

BACs are large bacterial chromosomes ranging from 150 to 300 kb that can accommodate large mammalian genes, along with their regulatory elements (Poser et al., 2008). BACs are engineered to clone mammalian genes of interest, along with their cognate promoter, intron-exon make up and regulatory elements, while tagging the protein product with GFP. BACs propagate independently of the *E. coli* chromosome; thus they can be purified and transfected into HeLa cells and the GFP tag is used for dual purpose: imaging and AP/immunoprecipitation (IP) followed by MS. A total of 1,125 mice orthologs of the human counterparts were cloned on BACs as N- or C-terminal GFP-tagged baits and expressed as transgenic fusions in HeLa cells to map human interactomes spanning all protein classes. Mice orthologs were chosen as alternative surrogates owing to their ability to resist RNA interference. Specific interactors with baits were identified via LFQ using computational algorithms, as previously explained, that use co-enrichment profiles as guide for suggestive interactions. This local co-enrichment as a first dimension led to identifying >5,000 proteins involved an upward of 30,000 statistically significant interactions.

In the second and third dimensions, IP baits along with their interaction partners were quantified to determine equimolar similarities or discrepancies that can shed light on protein complex stoichiometries (Hein et al., 2015). Although tagging the protein of interest at certain terminus may reveal stoichiometry disparities when compared with fusion and purification from the other terminus, the LFQ approach applied for stoichiometry estimations has analyzed IP profiles when baits are N- or C-terminally tagged, as well as overlaying stoichiometry outputs with those obtained from cloning the human counterpart of the mouse ortholog. Another gap covered by a third dimension was linking complex stoichiometries to cellular protein abundances via whole proteome quantification using the proteome ruler concept (Wisniewski et al., 2014). This novel approach has reliably quantified the human interactome in three quantitative dimensions that include label-free quantification of both IPs and complete proteomes.

Of importance, stable or core complexes were found to exhibit consistent stoichiometries among experimental replicates, whether baits are N- or C-terminally tagged. This is unlike transient complexes or satellite components such as regulatory proteins associated with core complexes that do not represent the machinery core, which tend to exhibit stoichiometry discrepancies. Thus, reciprocal verifications should reflect core interactomes, whereas the lack of suggests non-core complex proteins associations. This tool offers a new dimension by analyzing less-stable RD mt protein complexes of otherwise stable complexes in healthy counterparts, which could be of particular relevance in biomarker discovery as defined in this case by stoichiometric changes in mt protein complexes. Although this approach sheds light on the fact that the vast majority of human protein interactions are weak associations, it still does not offer sufficient solutions on developing workflows that more robustly capture transient PPIs beyond the trial-and-error endeavor, as evidenced by the lack of reciprocity when different ends are tagged or orthologs are tried, posing an ongoing bottleneck in interactomics research.

Imaging-MS and proximity ligation imaging cytometry

Monitoring PPIs and PTMs in rare cell types such as stem and rare immune cells remains largely unaddressed to date. A newly developed approach termed proximity ligation imaging cytometry (PLIC) offers promise to advance rare cell proteomics. This method has been demonstrated through a proof of concept in medullary thymic epithelial cells (meTECs), which are rare immune cell types amounting to less than 0.1% of thymus cells. Single cell suspensions of thymi were prepared from fat-free surgically removed thymi from 5-week-old mice, followed by enzymatic and mechanical treatment to create a suspension of single cells for PLIC analysis. PLIC essentially combines nucleotide-based proximity ligation assay (PLA) with imaging flow cytometry (IFC). It conceptually relies on targeting the protein of interest with an oligonucleotide-conjugated antibody. If another protein is targeted with a second oligonucleotide-conjugated antibody and both proteins interact, the two oligonucleotides get into proximity, prompting the ligation of additional connector DNA strands and yielding a circular DNA, the product of which can be detected by fluorescence microscopy (Avin et al., 2017; Soderberg et al., 2006). An updated version of this methodology has recently enabled multiparametric fluorescent proteomic analysis of single cells in suspension by coupling PLA with IFC, which led to the morphological examination and proteomic quantitation of thousands of events in rare cell types with great precision (Avin et al., 2017).

As proof of assay sensitivity, fluorescent signals were detected upon an interaction between autoimmune regulator (AIRE) and sirtuin-1 (SIRT1) proteins in meTECs, and those signals were not recorded for other negative control cell types such as CD45+ populations, demonstrating sensitivity and specificity. When fluorescent signals are combined with single protein staining, results often show the lack of complete overlap between interacting proteins in co-localization attempts, suggesting that relying on fluorescent signaling may over quantify PPIs. Another contributor to false-positive signals is sample section thickness, which can be heterogeneous in a manner that over or underrepresents the population of cell type under study. To overcome these pitfalls, an integrated IFC within PLIC includes data processing algorithms that filter out false-positive signals for thousands of single cells. PLIC can also be used to quantify dimerization and oligomerization states of individual proteins by targeting the same protein in any cell type with two different oligonucleotide-labeled antibodies, and if the protein dimerizes or oligomerizes, the outcome will yield fluorescent signals that can offer insights into interaction stoichiometries.

PLIC offers additional importance in PTM analysis of RDs by probing for the protein of interest with two different unlabeled primary antibodies that target select protein and its modified molecule, respectively, such as lysine-acetylation specific antibody. If the PTM includes substrate acetylation, then both primary antibodies can come to proximity, followed by oligonucleotide-labeled secondary antibody that reconstitutes a fluorescent signal due to PTM of the target protein. PLIC thus offers a significant push to PPI and PTM research, especially in rare cell types, which can be widely implicated in mt connectivity to RD onset and/or progression. For example, numerous mt respiratory chain proteins that are involved in the assembly of complexes I–IV are phosphorylated under healthy conditions (Stram and Payne, 2016), resulting in increased or decreased activity. In RDs such as RCC deficiency, PLIC can be the method of choice in monitoring the fluorescence signal patterns in response to fluctuations in PTMs of special complex subunits of interest, when compared with healthy subject signals, thus offering a new dimension into the impact of PTM alterations in RDs.

SINGLE-CELL PROTEOMICS IN RARE CELL POPULATIONS

Single-cell proteomics is the new hype, taking humanity several steps closer to precision medicine. The approach is gaining interest because no two cells are alike. Such heterogeneity needs to be investigated on an individual cell level, especially for rare cell types, to better understand human disease specificity (Doerr, 2019). This requires the development of reliable methods to isolate and/or sort single cells for downstream proteomics such as fluorescent probing, flow cytometry, mass cytometry (CyTOF), droplet microfluidics, microengraving, single-cell barcode chips (SCBCs), magnetic ranking cytometry (MagCR), and other MS-based approaches. Fluorescent probing enables the dynamic analysis of temporal and spatial distributions (Purvis et al., 2012), and GFP has been the key contributor in these attempts, with GFP variants such as the superfolder GFP improving both folding and functional retention of fusion proteins on select target cells (Kamiyama et al., 2016).

GFP alternatives include mEos2 and Dendra2 that photoconvert in the presence of UV light and other fluorescent reporters that are cell permeable, which are used for cytosolic studies via capillary electrophoresis, in addition to far-field imaging that integrates photobleaching for long-term tracking in single cells (Chudakov et al., 2007; Kelley, 2020; Wiedenmann et al., 2004). Flow cytometry has undergone remarkable advancement with single-cell implications owing to its capacity to monitor more than 15 parameters per cell, with its recently updated fluorescence-activated cell sorting being widely used to isolate single cells based on prior garnered information for subsequent downstream proteomic applications.

CyTOF capitalizes on antibody isotope labeling, with immunostained cells subsequently droplet nebulized for argon plasma exposure to evaporate the cell and ionize its constituents, followed by time-of-flight MS analysis (Bandura et al., 2009). Droplet microfluidics in turn encapsulate single cells in oil-water emulsion droplets in picoliter volumes, with the capacity to analyze single-cell secretomes beyond the flow cytometry grasp. This in addition to the dense microwell-generated microengraving that involves well coating with antibodies specific to a particular cell type that can robustly capture specific cell types for downstream proteomics (Love et al., 2006).

SCBCs in contrast are microchamber based with arrayed barcodes carrying spatial addresses specific to select proteins and conjugated to DNA fragments that convert the barcode into antibody arrays, which is being gradually used in single-cell secretome mappings (Bailey et al., 2007). MagCR is among the latest technologies that conjugates cell surface-specific antibodies to magnetic nanoparticles that upon capturing the cell type of interest can be sorted via microfluidic devices for downstream proteomics (Poudineh et al., 2017). The aforesaid collection of single-cell capturing methods showcase the newest developments for the specific objective of isolating single cells for proteomic applications, which can open new avenues in RD research.

PROTEOMICS AND OTHER DATASETS FOR IDENTIFYING DISEASE DRIVER GENES

The complexity of rare and other human disease renders monothematic approaches unsustainable. Research groups focusing on proteomics and omics approaches, bioinformatics, and artificial intelligence must explore interdisciplinary collaborations to untangle disease facets that require multiple sources of data. This can eventually establish a new *modus operandi* of multiple data source integrations and refined outputs that systematically unify concepts that transform independent bulk data into streamlined quality-controlled findings. This is particularly important owing to the growing acceptance that not a single one of the aforesaid methods is sufficient to resolve complex RD puzzles.

The Human Genome Project (International Human Genome Sequencing Consortium, 2004) has made the human genome sequence available, which revealed more than 3 million SNPs. This represented an earlier foundation for subsequent genome-wide association studies (GWASs) that have been of value in identifying genetic variations deviating from the norm, whether on an individual or population-specific scale, in an attempt to uncover alterations linked to human disease. The downside of GWASs stems from the lack of certainty on whether these SNPs will upregulate, downregulate, or obliterate gene expression and the scope of proteomic or transcriptomic consequences. This is because most SNPs are found in non-coding regions that often contain promoters and other regulatory elements of gene expression. Thus, non-coding region SNPs likely impact gene expression in any direction. Complementary approaches become necessary to better understand the biological significance of coding and non-coding region SNPs, with implications in biomarker discovery and precision medicine. Linking the genetic makeup to

transcriptome profiling offers insights into cellular dynamics and reveals the impact of DNA alterations on gene expression levels. Transcriptomics can also be pivotally coupled with proteomics as a form of quality control, since proteomics, especially in RDs, still lack sufficiently robust methods to ensure proteome and interactome profiles mapped via targeted and untargeted proteomics are tissue and/or cell type specific (Moutaoufik et al., 2019). Rapid developments in sorting techniques and single-cell RNA sequencing have made it possible to capture cell types of interest on a single cell level. Thus, combining transcriptomic data with proteomic maps can significantly facilitate filtering out false-positive PPIs that are not backed by cognate transcriptome signatures due to tissue contamination.

Multidisciplinary research is also extending into the area of drugomics, where drug-protein interaction studies are being deposited in databases such as Drug2Gene and Drug Bank (Roider et al., 2014; Wishart et al., 2006). The connectivity map project is making it possible to identify the impact of drugs on genome-wide transcript levels of the entire human genome (Lamb et al., 2006). This reveals the scope of genes whose transcription levels are impacted by the same drug, which can accelerate functional annotations of unknown orphans by linking them to pathways exhibiting similar response to drug treatments. Furthermore, since biochemical pathways result in the production of signature metabolites, the area of metabolomics can link metabolic variations across various samples to altered biochemical pathways, which can be used to validate proteomics and genetic alteration backgrounds. Although proteomics can help guide targeted diagnostics, integrating other biological datasets, such as single-cell transcriptomics, genomics, and drugomics, can refine driver gene lists in rare and other human diseases.

GENOTYPIC AND PHENOTYPIC MATCHMAKING TOOLS

Another issue of considerable difficulty in RD research is variant validation. In other words, when disease-causing mutations are identified by tools such as whole-exome sequencing, it becomes important to validate the authenticity of variant linkage to the RD under study, which is a difficult task owing to the scarcity of other patients with the same rare disease type. To overcome this bottleneck, new tools are emerging, such as the PhenomeCentral Portal that is fed with exome or genetic variant sequences gathered from RD patient samples, along with phenotypic descriptions associated with these variants (Buske et al., 2015). Clinicians can next access the portal and feed their patient clinical features to the database, which matches the new query with stored patient genotypic/phenotypic records and instantly suggests causative genes based on semantic similarities that enable hypothesis-free and hypothesis-driven prioritization of candidate genes that clinicians can focus on for exome sequence validations.

PhenomeCentral accepts exome and genetic variant records even if these entries lack sufficient molecular diagnostic information. This is because the portal considers that in the future when similar variants are fed into the system along with complete molecular or phenotypic explanations, it can match records and refine its stored information in what appears to be an ongoing effort that can significantly advance RD research matchmaking between previously stored genotypes along with observable phenotypes with the new queries. Matchmaking tools are gaining similar momentum in the genomic and proteomics interface, as exemplified by the development of Orphanet (Pavan et al., 2017), a promising RD data reference source for sharing orphan drug discovery. More than 320,000 possible links between known drugs obtained from DrugBank and proteins of unknown function involved in rare and orphan diseases suggested that an upward of 18,000 drugs can be repurposed as rare or orphan disease therapeutic options. In addition, artificial intelligence is also being rapidly integrated to model diseased human cells, which can overcome the scarcity of RD samples. Examining these models can vary based on nutrient, lipid, and metabolite to suggest potential molecules of therapeutic value solely based on machine learning algorithms that stratify and match patients and prioritize *in silico*-examined therapeutic options (Zilocchi et al., 2020). This can help clinicians involved in RDs to find their needle in the haystack, and with more contributions from the scientific and clinician community, more tools, methods, and databases can be developed to accelerate RD therapeutic developments.

FUTURE DIRECTIONS OR PERSPECTIVES

The “omics” age requires morphing from mere depositing of research outputs in public domains into actionable therapeutics. This calls for concerted collaborations to refine dataset creation and integration methods, filtering out fuzziness, and giving birth to rapid, yet effective, rare and other disease biomarker discovery platforms to materialize the precision medicine promise into a strong reality. Drug discovery and biomarker identification is a daunting area of research, let alone RDs, owing to the scarcity of available samples to analyze. Despite the advancement of next-generation sequencing that can robustly screen whole

genomes, whole exomes, and targeted sequences for genetic variations in a high-throughput scale, limited success can be achieved for RDs. This reveals the power of systems biology and the marginal advantage of interactomics over proteomics, since network-based approaches can link PPIs involving rare variants in one protein to others within a network map for the potential usage of currently available medications through repurposing, until more precise medications are developed. Alternately, PPIs involving pathological variants can be targeted by peptide therapeutics that are less toxic than synthetic chemicals, but this requires layers of PPI map overlays to painstakingly identify targetable interactomes, which can be a hard reach in RDs. Furthermore, 3D structures of protein candidates pertaining to RDs can be subjected to molecular docking, whereby available small molecules as drug leads can be *in silico* docked into these structures for downstream experimental validation. It is expected that this decade will reveal more advances that will take rare human disease research several steps closer to refining disease-associated data, thus moving away from fuzziness and consistently toward individual-specific RD therapeutics.

METHODS

All methods can be found in the accompanying [Transparent methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2020.102030>.

ACKNOWLEDGMENTS

M.T.M. was supported by a Saskatchewan Health Research Foundation Postdoctoral Fellowship. M.B. is supported by CIHR New Investigator award (MSH-130178), ARSACS Foundation, ALS Society of Canada, Parkinson Canada (2014-673), National Institutes of Health (R01GM106019), and CIHR (MOP-125952; RSN-124512, 132191; FDN-154318) grants.

AUTHOR CONTRIBUTIONS

K.A.A. and M.B. outlined the manuscript. M.T.M., S.P., and Q.Z. performed the analysis. M.T.M., S.P., and M.B. prepared the figures. K.A.A. and M.B. wrote the manuscript, with input from other authors. All read and approved the manuscript.

REFERENCES

- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457–465.
- Antonick, H., Lin, Z.Y., Janer, A., Aaltonen, M.J., Weraarpachai, W., Gingras, A.C., and Shoubridge, E.A. (2020). A high-density human mitochondrial proximity interaction network. *Cell Metab.* 32, 479–497.e9.
- Area-Gomez, E., de Groof, A., Bonilla, E., Montesinos, J., Tanji, K., Boldogh, I., Pon, L., and Schon, E.A. (2018). A key role for MAM in mediating mitochondrial dysfunction in Alzheimer disease. *Cell Death Dis.* 9, 335.
- Avin, A., Levy, M., Porat, Z., and Abramson, J. (2017). Quantitative analysis of protein-protein interactions and post-translational modifications in rare immune populations. *Nat. Commun.* 8, 1524.
- Bailey, R.C., Kwong, G.A., Radu, C.G., Witte, O.N., and Heath, J.R. (2007). DNA-encoded antibody libraries: a unified platform for multiplexed cell sorting and detection of genes and proteins. *J. Am. Chem. Soc.* 129, 1959–1967.
- Balasubramaniam, M., Ayyadevara, S., Ganne, A., Kakraba, S., Penthalha, N.R., Du, X., Crooks, P.A., Griffin, S.T., and Shmookler Reis, R.J. (2019). Aggregate interactome based on protein cross-linking interfaces predicts drug targets to limit aggregation in neurodegenerative diseases. *iScience* 20, 248–264.
- Bandura, D.R., Baranov, V.I., Ornatsky, O.I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J.E., and Tanner, S.D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* 81, 6813–6822.
- Binder, J.X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S.I., Schneider, R., and Jensen, L.J. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)* 2014, bau012.
- Blakely, E.L., de Silva, R., King, A., Schwarzer, V., Harrower, T., Dawidek, G., Turnbull, D.M., and Taylor, R.W. (2005). LHON/MELAS overlap syndrome associated with a mitochondrial MTND1 gene mutation. *Eur. J. Hum. Genet.* 13, 623–627.
- Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* 36, 880–887.
- Buske, O.J., Girdea, M., Dumitriu, S., Gallinger, B., Hartley, T., Trang, H., Misyura, A., Friedman, T., Beaulieu, C., Bone, W.P., et al. (2015). PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum. Mutat.* 36, 931–940.
- Butera, G., Brandi, J., Cavallini, C., Scarpa, A., Lawlor, R.T., Scupoli, M.T., Marengo, E., Cecconi, D., Manfredi, M., and Donadelli, M. (2020). The mutant p53-driven secretome has oncogenic functions in pancreatic ductal adenocarcinoma cells. *Biomolecules* 10, 884.
- Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* 44, D1251–D1257.
- Cavalli, E., Battaglia, G., Basile, M.S., Bruno, V., Petralia, M.C., Lombardo, S.D., Pennisi, M., Kalfin, R., Tancheva, L., Fagone, P., et al. (2020). Exploratory analysis of iPSCs-derived neuronal cells as predictors of diagnosis and treatment of Alzheimer disease. *Brain Sci.* 10, 166.
- Chavez, J.D., Tang, X., Campbell, M.D., Reyes, G., Kramer, P.A., Stuppard, R., Keller, A., Zhang, H., Rabinovitch, P.S., Marcinek, D.J., et al. (2020). Mitochondrial protein interaction landscape of SS-31. *Proc. Natl. Acad. Sci. U S A* 117, 15363–15373.

- Christoforou, A., Mulvey, C.M., Breckels, L.M., Geladaki, A., Hurrell, T., Hayward, P.C., Naake, T., Gatto, L., Viner, R., Martinez Arias, A., et al. (2016). A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* 7, 8992.
- Chudakov, D.M., Lukyanov, S., and Lukyanov, K.A. (2007). Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2. *Nat. Protoc.* 2, 2024–2032.
- Collins, B.C., Hunter, C.L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S.L., Chan, D.W., Gibson, B.W., Gingras, A.C., Held, J.M., et al. (2017). Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* 8, 291.
- De Munter, S., Gornemann, J., Derua, R., Lesage, B., Qian, J., Heroes, E., Waelkens, E., Van Eynde, A., Beullens, M., and Bollen, M. (2017). Split-BiolD: a proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS Lett.* 591, 415–424.
- Doerr, A. (2019). Single-cell proteomics. *Nat. Methods* 16, 20.
- Engert, J.C., Berube, P., Mercier, J., Dore, C., Lepage, P., Ge, B., Bouchard, J.P., Mathieu, J., Melancon, S.B., Schalling, M., et al. (2000). ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat. Genet.* 24, 120–125.
- Floyd, B.J., Wilkerson, E.M., Veling, M.T., Minoque, C.E., Xia, C., Beebe, E.T., Wrobel, R.L., Cho, H., Kremer, L.S., Alston, C.L., et al. (2016). Mitochondrial protein interaction mapping identifies regulators of respiratory chain function. *Mol. Cell* 63, 621–632.
- Gill, C., Phelan, J.P., Hatzipetros, T., Kidd, J.D., Tassinari, V.R., Levine, B., Wang, M.Z., Moreno, A., Thompson, K., Maier, M., et al. (2019). SOD1-positive aggregate accumulation in the CNS predicts slower disease progression and increased longevity in a mutant SOD1 mouse model of ALS. *Sci. Rep.* 9, 6724.
- Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11, O111.016717.
- Girard, M., Lariviere, R., Parfitt, D.A., Deane, E.C., Gaudet, R., Nossova, N., Blondeau, F., Prenosil, G., Vermeulen, E.G., Duchon, M.R., et al. (2012). Mitochondrial dysfunction and Purkinje cell loss in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *Proc. Natl. Acad. Sci. U S A* 109, 1661–1666.
- Haendel, M., Vasilevsky, N., Unni, D., Bologa, C., Harris, N., Rehm, H., Hamosh, A., Baynam, G., Groza, T., McMurry, J., et al. (2020). How many rare diseases are there? *Nat. Rev. Drug Discov.* 19, 77–78.
- Hallmann, K., Kudin, A.P., Zsurka, G., Kornblum, C., Reimann, J., Stuve, B., Waltz, S., Hattingen, E., Thiele, H., Nurnberg, P., et al. (2016). Loss of the smallest subunit of cytochrome c oxidase, COX8A, causes Leigh-like syndrome and epilepsy. *Brain* 139, 338–345.
- Han, S., Udeshi, N.D., Deerinck, T.J., Svinkina, T., Ellisman, M.H., Carr, S.A., and Ting, A.Y. (2017). Proximity biotinylation as a method for mapping proteins associated with mtDNA in living cells. *Cell Chem. Biol.* 24, 404–414.
- Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. *Cell* 150, 1068–1081.
- 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.
- Huch, M., and Koo, B.K. (2015). Modeling mouse and human development using organoid cultures. *Development* 142, 3113–3125.
- Indrieri, A., van Rahden, V.A., Tiranti, V., Morleo, M., Iaconis, D., Tammara, R., D'Amato, I., Conte, I., Maystadt, I., Demuth, S., et al. (2012). Mutations in COX7B cause microphthalmia with linear skin lesions, an unconventional mitochondrial disease. *Am. J. Hum. Genet.* 91, 942–949.
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.
- Itzhak, D.N., Davies, C., Tyanova, S., Mishra, A., Williamson, J., Antrobus, R., Cox, J., Weekes, M.P., and Borner, G.H.H. (2017). A mass spectrometry-based approach for mapping protein subcellular localization reveals the spatial proteome of mouse primary neurons. *Cell Rep.* 20, 2706–2718.
- Itzhak, D.N., Tyanova, S., Cox, J., and Borner, G.H. (2016). Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* 5, e16950.
- Jones, A.X., Cao, Y., Tang, Y.L., Wang, J.H., Ding, Y.H., Tan, H., Chen, Z.L., Fang, R.Q., Yin, J., Chen, R.C., et al. (2019). Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers. *Nat. Commun.* 10, 3911.
- Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L.A., Ishikawa, H., Leonetti, M.D., Marshall, W.F., Weissman, J.S., et al. (2016). Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* 7, 11046.
- Kelley, M.L.a.S.O. (2020). Single-cell analysis targeting the proteome. *Nat. Rev. Chem.* 4, 143–158.
- Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U S A* 111, E2453–E2461.
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935.
- Li, P., Meng, Y., Wang, L., and Di, L.J. (2019). BiolD: a proximity-dependent labeling approach in proteomics study. *Methods Mol. Biol.* 1871, 143–151.
- Liu, F., Lossel, P., Scheltema, R., Viner, R., and Heck, A.J.R. (2017). Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* 8, 15473.
- Liu, F., Rijkers, D.T., Post, H., and Heck, A.J. (2015). Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* 12, 1179–1184.
- Liu, X., Salokas, K., Tamene, F., Jiu, Y., Weldatsadik, R.G., Ohman, T., and Varjosalo, M. (2018). An AP-MS- and BiolD-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.* 9, 1188.
- Love, J.C., Ronan, J.L., Grotenbreg, G.M., van der Veen, A.G., and Ploegh, H.L. (2006). A microengraving method for rapid selection of single cells producing antigen-specific antibodies. *Nat. Biotechnol.* 24, 703–707.
- Luck, K., Kim, D.K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charlotiaux, B., et al. (2020). A reference map of the human binary protein interactome. *Nature* 580, 402–408.
- Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 14, e8126.
- Luo, Y., Mok, T.S., Lin, X., Zhang, W., Cui, Y., Guo, J., Chen, X., Zhang, T., and Wang, T. (2017). SWATH-based proteomics identified carbonic anhydrase 2 as a potential diagnosis biomarker for nasopharyngeal carcinoma. *Sci. Rep.* 7, 41191.
- Malty, R.H., Aoki, H., Kumar, A., Phanse, S., Amin, S., Zhang, Q., Minic, Z., Goebels, F., Musso, G., Wu, Z., et al. (2017). A map of human mitochondrial protein interactions linked to neurodegeneration reveals new mechanisms of redox Homeostasis and NF-kappaB signaling. *Cell Syst.* 5, 564–577.e12.
- Malty, R.H., Jessulat, M., Jin, K., Musso, G., Vlasblom, J., Phanse, S., Zhang, Z., and Babu, M. (2015). Mitochondrial targets for pharmacological intervention in human disease. *J. Proteome Res.* 14, 5–21.
- Maoz, B.M., Herland, A., FitzGerald, E.A., Grevesse, T., Vidoudez, C., Pacheco, A.R., Sheehy, S.P., Park, T.E., Dauth, S., Mannix, R., et al. (2018). A linked organ-on-chip model of the human neurovascular unit reveals the metabolic coupling of endothelial and neuronal cells. *Nat. Biotechnol.* 36, 865–874.
- Massa, V., Fernandez-Vizcarra, E., Alshahwan, S., Bakhsh, E., Goffrini, P., Ferrero, I., Mereghetti, P., D'Adamo, P., Gasparini, P., and Zeviani, M. (2008). Severe infantile encephalomyopathy caused by a mutation in COX6B1, a nucleus-encoded subunit of cytochrome c oxidase. *Am. J. Hum. Genet.* 82, 1281–1289.

- Minde, D.P., Ramakrishna, M., and Lilley, K.S. (2020). Biotin proximity tagging favours unfolded proteins and enables the study of intrinsically disordered regions. *Commun. Biol.* 3, 38.
- Monaco, A., and Faldut, A. (2020). Protein aggregation and dysfunction of autophagy-lysosomal pathway: a vicious cycle in lysosomal storage diseases. *Front. Mol. Neurosci.* 13, 37.
- Moutaoufik, M.T., Malt, R., Amin, S., Zhang, Q., Phanse, S., Gagarinova, A., Zilocchi, M., Hoell, L., Minic, Z., Gagarinova, M., et al. (2019). Rewiring of the human mitochondrial interactome during neuronal reprogramming reveals regulators of the respirasome and neurogenesis. *iScience* 19, 1114–1132.
- Nambu, M., Nishiumi, S., Kobayashi, T., Masuda, T., Ito, S., Yoshida, M., and Ohtsuki, S. (2020). Effects of differences in pre-analytical processing on blood protein profiles determined with SWATH-MS. *J. Proteomics* 223, 103824.
- Narasimhan, M., Kannan, S., Chawade, A., Bhattacharjee, A., and Govekar, R. (2019). Clinical biomarker discovery by SWATH-MS based label-free quantitative proteomics: impact of criteria for identification of differentiators and data normalization method. *J. Transl. Med.* 17, 184.
- Nguengang Wakap, S., Lambert, D.M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., and Rath, A. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur. J. Hum. Genet.* 28, 165–173.
- Nusinow, D.P., Szpyt, J., Ghandi, M., Rose, C.M., McDonald, E.R., 3rd, Kalocsay, M., Jane-Valbuena, J., Gelfand, E., Schweppe, D.K., Jedrychowski, M., et al. (2020). Quantitative proteomics of the cancer cell line encyclopedia. *Cell* 180, 387–402.e16.
- O'Reilly, F.J., and Rappsilber, J. (2018). Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* 25, 1000–1008.
- Pavan, S., Rommel, K., Mateo Marquina, M.E., Hohn, S., Lanneau, V., and Rath, A. (2017). Clinical practice guidelines for rare diseases: the Orphanet database. *PLoS One* 12, e0170365.
- Pek, N.M.Q., Phua, Q.H., Ho, B.X., Pang, J.K.S., Hor, J.H., An, O., Yang, H.H., Yu, Y., Fan, Y., Ng, S.Y., et al. (2019). Mitochondrial 3243A > G mutation confers pro-atherogenic and pro-inflammatory properties in MELAS iPS derived endothelial cells. *Cell Death Dis.* 10, 802.
- Petschnigg, J., Groisman, B., Kotlyar, M., Taipale, M., Zheng, Y., Kurat, C.F., Sayad, A., Sierra, J.R., Mattiazzi Usaj, M., Snider, J., et al. (2014). The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat. Methods* 11, 585–592.
- Pinero, J., Bravo, A., Queralt-Rosinach, N., Gutierrez-Sacristan, A., Deu-Pons, J., Centeno, E., Garcia-Garcia, J., Sanz, F., and Furlong, L.I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 45, D833–D839.
- Ping, L., Duong, D.M., Yin, L., Gearing, M., Lah, J.J., Levey, A.I., and Seyfried, N.T. (2018). Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. *Sci. Data* 5, 180036.
- Pitceathly, R.D., Rahman, S., Wedatilake, Y., Polke, J.M., Cirak, S., Foley, A.R., Sailer, A., Hurler, M.E., Stalker, J., Hargreaves, I., et al. (2013). NDUFA4 mutations underlie dysfunction of a cytochrome c oxidase subunit linked to human neurological disease. *Cell Rep.* 3, 1795–1805.
- Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovskiy, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* 5, 409–415.
- Poudineh, M., Aldridge, P.M., Ahmed, S., Green, B.J., Kermanshah, L., Nguyen, V., Tu, C., Mohamadi, R.M., Nam, R.K., Hansen, A., et al. (2017). Tracking the dynamics of circulating tumour cell phenotypes using nanoparticle-mediated magnetic ranking. *Nat. Nanotechnol.* 12, 274–281.
- Pourhaghighi, R., Ash, P.E.A., Phanse, S., Goebels, F., Hu, L.Z.M., Chen, S., Zhang, Y., Wierbowski, S.D., Boudeau, S., Moutaoufik, M.T., et al. (2020). BrainMap elucidates the macromolecular connectivity landscape of mammalian brain. *Cell Syst.* 10, 333–350.e14.
- Purvis, J.E., Karhohs, K.W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G. (2012). p53 dynamics control cell fate. *Science* 336, 1440–1444.
- Rajput, N.K., Singh, V., and Bhardwaj, A. (2015). Resources, challenges and way forward in rare mitochondrial diseases research. *F1000Res.* 4, 70.
- Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* 339, 1328–1331.
- Roider, H.G., Pavlova, N., Kirov, I., Slavov, S., Slavov, T., Uzunov, Z., and Weiss, B. (2014). Drug2Gene: an exhaustive resource to explore effectively the drug-target relation network. *BMC Bioinformatics* 15, 68.
- Sahni, N., Yi, S., Taipale, M., Fuxman Bass, J.I., Coulombe-Huntington, J., Yang, F., Peng, J., Weile, J., Karras, G.I., Wang, Y., et al. (2015). Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* 161, 647–660.
- Salvatore, M., Magrelli, A., and Taruscio, D. (2011). The role of microRNAs in the biology of rare diseases. *Int. J. Mol. Sci.* 12, 6733–6742.
- Schon, E.A., DiMauro, S., and Hirano, M. (2012). Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat. Rev. Genet.* 13, 878–890.
- Schopp, I.M., Amaya Ramirez, C.C., Debeljak, J., Kreibich, E., Skribbe, M., Wild, K., and Bethune, J. (2017). Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat. Commun.* 8, 15690.
- Schweppe, D.K., Chavez, J.D., Lee, C.F., Caudal, A., Kruse, S.E., Stuppard, R., Marcinek, D.J., Shadel, G.S., Tian, R., and Bruce, J.E. (2017). Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc. Natl. Acad. Sci. U S A* 114, 1732–1737.
- Smith, A.C., and Robinson, A.J. (2019). MitoMiner v4.0: an updated database of mitochondrial localization evidence, phenotypes and diseases. *Nucleic Acids Res.* 47, D1225–D1228.
- Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., et al. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* 3, 995–1000.
- Stauch, K.L., Villeneuve, L.M., Totusek, S., Lamberty, B., Ciborowski, P., and Fox, H.S. (2019). Quantitative proteomics of presynaptic mitochondria reveal an overexpression and biological relevance of neuronal MitoNEET in postnatal brain development. *Dev. Neurobiol.* 79, 370–386.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122, 957–968.
- Stram, A.R., and Payne, R.M. (2016). Post-translational modifications in mitochondria: protein signaling in the powerhouse. *Cell. Mol. Life Sci.* 73, 4063–4073.
- Tamiya, G., Makino, S., Hayashi, M., Abe, A., Numakura, C., Ueki, M., Tanaka, A., Ito, C., Toshimori, K., Ogawa, N., et al. (2014). A mutation of COX6A1 causes a recessive axonal or mixed form of Charcot-Marie-Tooth disease. *Am. J. Hum. Genet.* 95, 294–300.
- Teckman, J.H., An, J.K., Blumenkamp, K., Schmidt, B., and Perlmutter, D. (2004). Mitochondrial autophagy and injury in the liver in alpha 1-antitrypsin deficiency. *Am. J. Physiol. Gastrointest. Liver Physiol.* 286, G851–G862.
- The Gene Ontology Consortium. (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338.
- Trinkle-Mulcahy, L. (2019). Recent advances in proximity-based labeling methods for interactome mapping. *F1000Res* 8, F1000 Faculty Rev-135.
- Tuovinen, E.A., Gronholm, J., Ohman, T., Poysti, S., Toivonen, R., Kreutzman, A., Heiskanen, K., Trotta, L., Toivainen-Salo, S., Routes, J.M., et al. (2020). Novel hemizygous IL2RG p.(Pro58Ser) mutation impairs IL-2 receptor complex expression on lymphocytes causing X-linked combined immunodeficiency. *J. Clin. Immunol.* 40, 503–514.
- UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.
- Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* 6, 83–90.

Wallace, D.C., and Fan, W. (2009). The pathophysiology of mitochondrial disease as modeled in the mouse. *Genes Dev.* 23, 1714–1736.

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.

Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Rucker, C., Salih, A., Spindler, K.D., and Nienhaus, G.U. (2004). EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc. Natl. Acad. Sci. U S A* 101, 15905–15910.

Wilkins, H.M., and Swerdlow, R.H. (2017). Amyloid precursor protein processing and bioenergetics. *Brain Res. Bull.* 133, 71–79.

Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M.,

Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. *Science* 352, aad0189.

Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34, D668–D672.

Wisniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* 13, 3497–3506.

Woodsmith, J., Apelt, L., Casado-Medrano, V., Ozkan, Z., Timmermann, B., and Stelzl, U. (2017). Protein interaction perturbation profiling at amino-acid resolution. *Nat. Methods* 14, 1213–1221.

Yachie, N., Petsalaki, E., Mellor, J.C., Weile, J., Jacob, Y., Verby, M., Ozturk, S.B., Li, S., Cote,

A.G., Mosca, R., et al. (2016). Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* 12, 863.

Zhang, X., Sadowski, P., and Punyadeera, C. (2020). Evaluation of sample preparation methods for label-free quantitative profiling of salivary proteome. *J. Proteomics* 210, 103532.

Zilocchi, M., Broderick, K., Phanse, S., Aly, K.A., and Babu, M. (2020). Mitochondria under the spotlight: on the implications of mitochondrial dysfunction and its connectivity to neuropsychiatric disorders. *Comput. Struct. Biotechnol. J.* 18, 2535–2546.

Zlatic, S.A., Vrailas-Mortimer, A., Gokhale, A., Carey, L.J., Scott, E., Burch, R., McCall, M.M., Rudin-Rush, S., Davis, J.B., Hartwig, C., et al. (2018). Rare disease mechanisms identified by genealogical proteomics of copper homeostasis mutant pedigrees. *Cell Syst.* 6, 368–380.e6.