

Critical Reviews and Perspectives

Current proteomics methods applicable to dissecting the DNA damage response

Monita Muralidharan^{1,2}, Nevan J. Krogan^{1,2}, Mehdi Bouhaddou^{1,2,3,4} and Minkyu Kim^{1,2,5,*}

¹Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA,

²Quantitative Biosciences Institute (QBI), University of California, San Francisco, CA 94158, USA, ³Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90024, USA,

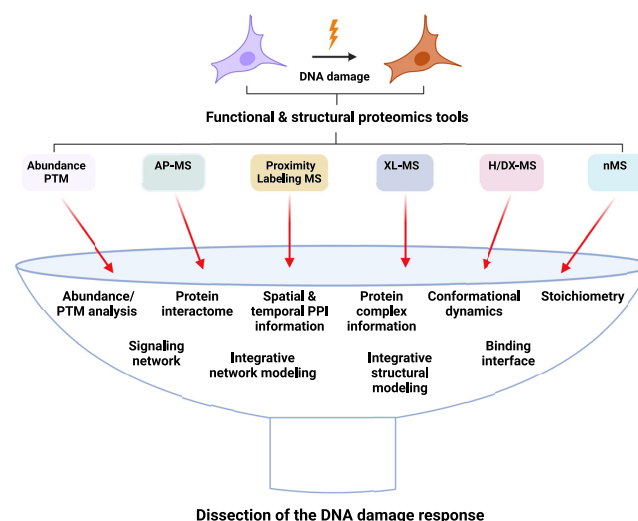
⁴Quantitative and Computational Biosciences Institute (QCBio), University of California, Los Angeles, CA 90024, USA and ⁵Department of Biochemistry and Structural Biology, University of Texas Health Science Center, San Antonio, TX 78229, USA

Received February 02, 2023; Revised April 25, 2023; Editorial Decision April 26, 2023; Accepted May 05, 2023

ABSTRACT

The DNA damage response (DDR) entails reorganization of proteins and protein complexes involved in DNA repair. The coordinated regulation of these proteomic changes maintains genome stability. Traditionally, regulators and mediators of DDR have been investigated individually. However, recent advances in mass spectrometry (MS)-based proteomics enable us to globally quantify changes in protein abundance, post-translational modifications (PTMs), protein localization, and protein-protein interactions (PPIs) in cells. Furthermore, structural proteomics approaches, such as crosslinking MS (XL-MS), hydrogen/deuterium exchange MS (H/DX-MS), Native MS (nMS), provide large structural information of proteins and protein complexes, complementary to the data collected from conventional methods, and promote integrated structural modeling. In this review, we will overview the current cutting-edge functional and structural proteomics techniques that are being actively utilized and developed to help interrogate proteomic changes that regulate the DDR.

GRAPHICAL ABSTRACT



INTRODUCTION

Liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) is being widely used to quantitatively measure both protein levels and post translational modifications. LC-MS/MS approaches can be performed in two different ways: 'top-down' approaches profile intact proteins and protein complexes, whereas 'bottom-up' or shotgun approaches profile proteins that have been digested to peptides. Peptides are then separated over a column (e.g. reversed phase C18) using increasing concentrations of organic solvent (e.g. acetonitrile), and as peptides elute from the column, they go through electrospray ionization (ESI) and the intact mass to charge ratio (m/z) of precursor or

*To whom correspondence should be addressed. Tel: +1 210 450 7346; Email: minkyu@uthscsa.edu

MS1 ions are measured. Precursor MS1 ions are then subjected to MS/MS or MS2 analysis, involving fragmentation of the peptide ion and subsequent measurement of fragment or product ion m/z (1,2). Through this tandem MS process, peptides in the samples are identified by comparing the collected data to a species-specific proteome database of *in silico*-digested proteins and scoring the resulting matches (3,4).

LC-MS/MS allows simultaneous identification and quantification of thousands of proteins or PTM sites on peptides in a single run. Abundance of each peptide across samples can be quantified and compared in several different ways: sample comparison can be assessed by using stable isotope labeling of amino acids in cells (SILAC) prior to tryptic digestion (5). Similarly, samples can be labeled post-digestion using isobaric tags (e.g. iTRAQ, TMT) (6–8) and multiple samples can be combined (up to 18 for TMT and 8 for iTRAQ) (9–11) for multiplexed quantification. Each tag possesses identical chemical structures and is differentiated via isotopes at different positions that generate peptides with the same m/z in MS1. However, upon MS2 fragmentation the tags produce different mass reporter ions, allowing relative quantification of peptide abundance. Alternatively, the relative abundance of each peptide in two or more biological samples can be directly measured without utilizing any labeling or special growth conditions (label-free).

Regarding MS data acquisition, there are two different acquisition modes for collecting MS data. In Data-Dependent Acquisition (DDA) mode, MS instrument selects the most abundant (typically top 10–20) peptide ions from MS1 scan in each column retention time window, and then these selected precursor peptide ions are fragmented and analyzed in MS2. However, the stochastic precursor selection of DDA could result in inconsistent detection and undersampling of less abundant peptides, which cause higher variation across replicates. Alternatively, in Data-Independent Acquisition (DIA) mode, MS instrument gathers MS2 data from all precursor ions in a narrow mass window (e.g. 400–425 m/z), and then repeats this process across the entire mass range (e.g. 400–1200 m/z) to systematically collect MS2 from all detected peptide ions in each retention time window (e.g. 2–4 s). MS instrument continues collecting MS2 data across the entire retention time range, thus generating highly multiplexed and comprehensive maps of fragment ions that elute on the chromatography. DIA uses a spectral library to extract peptide information from MS2 data, which leads to more accurate peptide quantification, lower variation across replicates, and fewer missing peptide identification, compared to DDA. But DDA is preferred for targeted analysis and library generation as it offers more accurate identification than DIA due to longer acquisition time per data point and more robust database-based search methods (12–14).

Interrogating the structure-function correlation of proteins and protein complexes has been crucial for understanding several cellular processes and disease mechanisms and developing various therapies. It is established that protein conformations are not just single specific three-dimensional quaternary structures but have multiple transient conformations with dynamic properties (15). Protein

dynamics facilitate its functions, such as macromolecular associations and other downstream processes in cellular pathways (16). Conventional structural biology tools such as X-ray crystallography (XRC), cryo-electron microscopy (cryoEM), and nuclear magnetic resonance (NMR) solve structures of important protein complexes but need to consistently deal with dynamics and heterogenous assemblies and require large quantities of protein, which limits their application. Protein chemistry combined with various mass spectrometry-based structural proteomics approaches, including XL-MS, H/DX-MS, and nMS, provide large structural information of proteins and protein complexes, in part due to their wide applicability to many different types of proteins, including structured, disordered, monomeric, and multimeric proteins (17). Furthermore, structural proteomics methods enable studying transient interactions of proteins and their complexes up to the megadalton range *in vitro* and *in vivo* under physiological conditions (18–22). Although structural proteomics tools alone do not solve structures, they provide invaluable information complementary to conventional methods, promoting integrated structural modeling (23). This review covers a repertoire of various proteomics approaches that can further our understanding of dynamic changes in proteome and mechanistic basis of PPIs that regulate DNA repair (Figure 1; Table 1).

FUNCTIONAL PROTEOMICS

Abundance proteomics

Profiling of total proteome in cells can be performed by digesting proteins to peptides with trypsin (and/or other peptidases) and applying them to LC-MS/MS. The relative abundance of each identified peptide is then measured and individual peptide data for each protein are combined to give protein-level quantification. This analysis allows us to quantitatively measure changes in abundance of thousands of proteins in cells and tissues simultaneously (Figure 1). To increase the proteome coverage (i.e. number of peptides and associated proteins identified in a given MS run), samples are often fractionated by ion exchange or reversed phase chromatography (24,25) and each fraction is individually analyzed by LC-MS/MS to maximize peptide identification. Furthermore, to globally map cellular macromolecular complexes, Complex-Centric Proteome Profiling method was recently developed (26). In this method, cell lysates are fractionated by size exclusion chromatography, and native protein complexes in each fraction are digested into peptides and analyzed by DIA MS. Quantification and composition of protein complexes are then determined based on protein patterns across the fractions by error-controlled, complex-centric analysis. This technique revealed the mitotic proteome reorganization, including disassembly of the nuclear pore complex, by quantitatively comparing human protein complexes during transition from interphase to mitosis (27).

Comparative proteomic analysis between BRCA1-deficient and BRCA1-proficient mouse tumors upon cisplatin treatment revealed that proteins involved in centrosome organization, chromosome condensation, HR repair and nucleotide metabolism were up-regulated in cisplatin-sensitive BRCA1-deficient tumors (28). However,

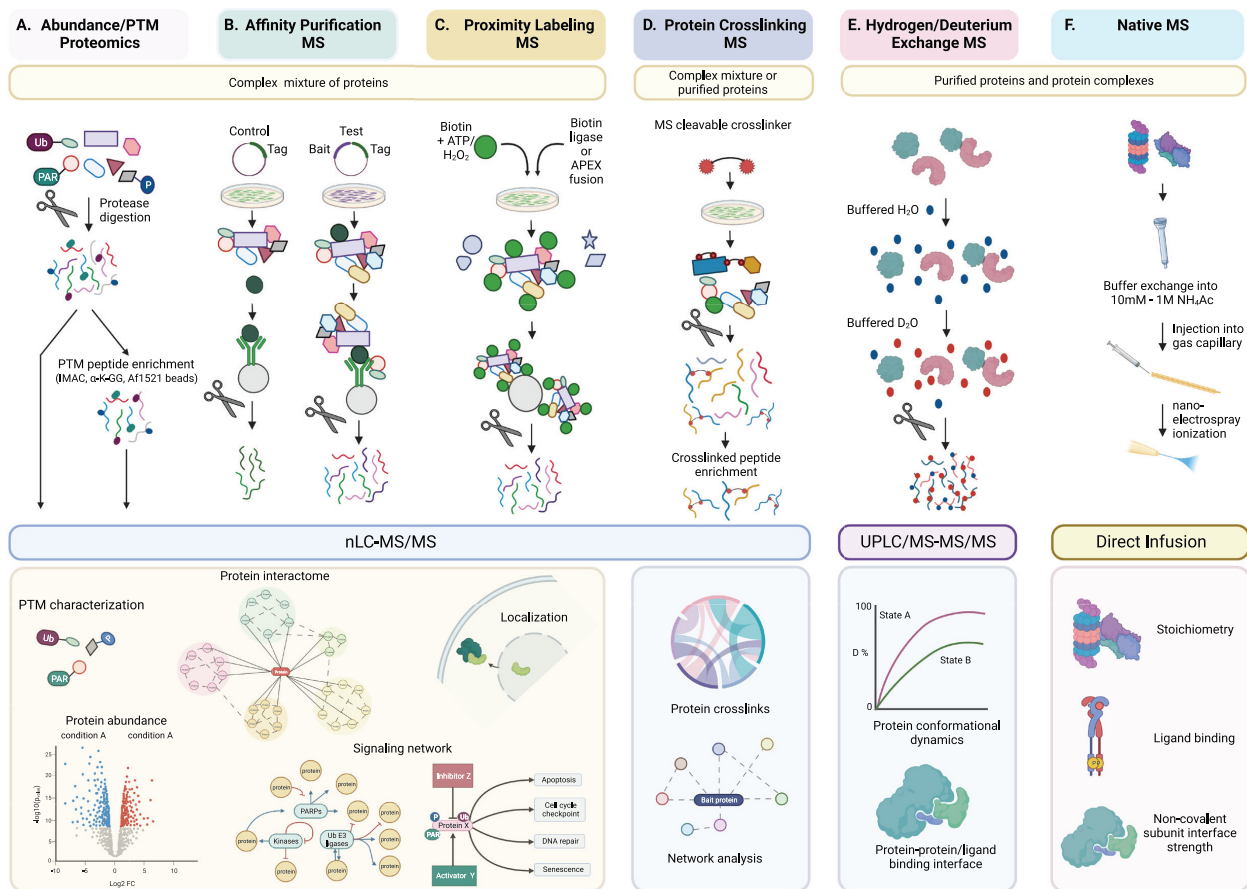


Figure 1. Schematic representation of functional and structural proteomics methods. (A) Proteins in cell lysates are digested and peptides are subsequently ionized and fragmented through nLC-MS/MS either directly (abundance proteomics) or after enrichment of PTM peptides (PTM proteomics). (B) Affinity Purification-MS (AP-MS) is to systematically identify PPIs by expressing and purifying affinity-tagged ‘bait’ proteins in cells and detecting stably bound protein interactors by MS. (C) Proximity labeling MS relies on a promiscuous biotin protein ligase (BirA/BioID/TurboID) or an engineered ascorbic acid peroxidase (APEX) fused to a protein of interest (bait) to label proteins in close proximity to the bait protein through covalent transfer of biotin or biotin derivatives in a distance-dependent manner. Upon addition of biotin, the proximal proteins are biotinylated, enriched with streptavidin beads, and identified by quantitative MS. These MS datasets from functional proteomics are integrated by network modeling and analysis to extract biological insight. (D) Protein crosslinking MS (XL-MS) involves crosslinking or surface modification of protein complexes followed by proteolysis and MS identifications of amino acid residue pairs that are within close spatial proximity, thus providing structural insights into proteins and protein assemblies. (E) Hydrogen/Deuterium exchange MS (H/DX-MS) measures dynamic changes in protein structures over a specified time scale by measuring the relative uptake of protein backbone amide hydrogen for deuterium (and *vice versa*). Residue-level information can be derived from gas-phase fragmentation of the deuterated peptides. (F) Native MS (nMS) involves analyzing and characterizing protein complexes, while retaining their inter and intramolecular interactions. The stoichiometry of a protein complex can be deduced from the calculated mass, and structural preference can be informed by the number of charges it carries. Data from the diverse structural proteomics methodologies can be combined to obtain maximum structural insight through integrative modeling. nLC, nano liquid chromatography; UPLC, ultra-performance liquid chromatography. Edges and arrows connecting proteins to other proteins and/or biological processes in the interactomes and networks are not only based on MS-driven proteomics data but also inferred from network modeling and analysis.

several studies show that overall changes in protein abundance during or after DNA damage exposure are much less conspicuous than PTM changes (29,30), necessitating the profiling of PTMs of proteins in addition to protein levels.

Post-translational modification (PTM) proteomics

PTMs (e.g. phosphorylation, ubiquitylation, and ADP-ribosylation among others) of many DDR proteins are known to induce changes in protein stability, cellular localization, enzymatic activity, and/or interaction with DNA and/or other proteins (31–38). To properly profile and quantify PTMs of proteins by MS, digested peptides from

lysates, bearing a given type of PTMs, need to be enriched prior to LC-MS/MS analysis (Figure 1). For phosphopeptide enrichment, Fe³⁺-based Immobilized Metal Affinity Chromatography (IMAC) beads are widely used and about 10 000–20 000 phosphorylated peptides (phosphopeptides) can be typically detected in a single LC-MS/MS run (39). Similarly, other metal-based (e.g. Ti⁴⁺, Zr⁴⁺) IMAC beads and titanium dioxide (TiO₂) can be also utilized to enrich phospho-peptides (39,40). To enrich ubiquitylated peptides, an antibody-bead conjugate that recognizes a di-glycine remnant of ubiquitin left on Lysine residues (K-GG) of protein substrates after trypsin digestion is used (41,42). This enrichment followed by

Table 1. Exemplary Proteomic Studies of the DNA Damage Response

Proteomic method	Pros	Cons	Study	Reference
Abundance proteomics	Quantitatively measure changes in abundance of thousands of proteins simultaneously	Proteome coverage could be limited for high complexity samples, in which fractionation before LC-MS/MS is recommended	Comparative proteomic analysis between BRCA1-deficient and BRCA1-proficient mouse tumors upon cisplatin treatment	(28)
			Comparative proteomic analysis between BRCA1-deficient and BRCA1-proficient mouse tumors upon cisplatin treatment	(28)
Phospho-proteomics	Determine the identity of phosphorylated proteins and the amino acid residues (and their level) which hold the phosphate group	Phospho-peptide enrichment step may not recover low abundant peptides and some enrichment beads (IMAC, TiO ₂) have higher efficiency for Ser/Thr than Tyr	Global changes in protein level and phosphorylation site profiles by SILAC with cisplatin treatment	(29,30)
			ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage	(43)
			Site-specific phosphorylation dynamics of the nuclear proteome during the DNA damage response	(44)
			Analyzing the phospho-proteome regulated by Spinophilin (PP1 regulatory subunit) that interacts with BRCA1	(46)
Ub-proteomics	Enables the large-scale analysis of ubiquitinated proteins and amino acid residues in cells	K-GG remnant profiling after trypsin digestion fails to recognize -GG- modified protein N-termini; loses Ub chain topology information; and may recognize ISG15- and NEDD8-modified proteins, which also yield K-GG remnant peptides upon cleavage by trypsin	Quantitative proteomic assessment of Ubiquitination and Acetylation in the DNA Damage Response	(47)
			Profiling of Proteome and Ubiquitome Changes in Human Lens Epithelial Cell Line after Ultraviolet-B Irradiation	(49)
ADPr-proteomics	Provides global information of ADPr proteins and amino acid residues	Numerous modifiable amino acid residues and the labile nature of ADPr bond pose a challenge for site-specific localization of ADPr Requires electron transfer dissociation (ETD)-based methods to preserve labile ADPr	Systems-wide analysis of Serine ADP-Ribosylation in response to oxidative stress	(62)
			Proteome-wide identification of poly(ADP-Ribosyl)ation targets in different genotoxic stress responses	(63,64)
AP-MS	Allows to identify in vivo PPIs by affinity purification using antibodies recognizing endogenous proteins of interest or affinity tags	Data could vary by antibodies employed Transient or weak PPIs may not be captured effectively Cannot distinguish between direct and indirect PPIs	Analyzing the effect of pathogenic mutations on the BRCA1 interactome	(46)
			TEADs associate with DNA repair proteins to facilitate recovery from DNA damage	(76)
			PPAR γ interacts with the MRN complex and the E3 ubiquitin ligase UBR5	(127)
iPOND, NCC	Enables to identify proteins specifically localized to chromosomal loci under active DNA replication	DNA labeling with EdU or biotin-dUTP and de/crosslinking steps are required NCC requires cell membrane permeabilization	Analysis of protein dynamics at active, stalled, and collapsed replication forks	(78,82–86)
ChIP-MS	Capture PPIs on crosslinked chromatin at specific gene loci and/or genome-wide	Relies on antibodies Requires de/crosslinking steps	Telomeres were immunoprecipitated and known and novel telomere-binding proteins were identified by MS	(79,80)

Table 1. Continued

Proteomic method	Pros	Cons	Study	Reference
Proximity proteomics	Detects transient and/or weak interactions	Proteins within proximity of a bait are identified, some of which are not necessarily interacting proteins	APEX reveals Shieldin as a 53BP1 effector complex in DSB repair	(91)
	Biotin-streptavidin binding is compatible with purification under harsh denaturing conditions	May have high backgrounds from abundant cellular proteins and/or endogenous biotinylation	BioID identifies Ku70-associated proteins	(96)
	Offers spatial and temporal resolution of PPIs	Biotinylation labeling time varies (1 min - 24 hr) depending on the enzyme used	BioID analysis of ATRX reveals its association with SLF2 that helps inhibit telomere exchanges	(97)
XL-MS	Elucidates PPIs and protein complex information by providing distance constraints within a protein or between proteins	Crosslinking efficiency is often low (1–5%), leading to marginal crosslinks	PPAR γ interact with the MRN complex via NBS1	(127)
	Depending on application needs (target functional group, solubility, membrane permeability, MS-cleavability, etc), a correct crosslinker can be selected	Long crosslinking time may induce protein aggregates	Mapping the regions involved in the Timeless–Tipin interaction	(128)
		Hybrid dipeptides generated by crosslinkers could greatly expand the search space during spectra matching		
H/DX-MS	Offers to analyze protein structures, dynamics, folding, and interactions	Deuterium/hydrogen scrambling effect (redistribution of the exchanged protons within a peptide) could complicate the analysis	Dynamic changes in DNA-binding domains of RPA upon ssDNA binding is revealed	(149)
	Not limited by the size of proteins or protein complexes	Back exchange of deuterium label during chromatographic separation due to intrinsic amino acid hydrogen exchange rate differences could lead to loss and alteration of signal and biased measurement	Helical subdomain (HD) of the PARP1 catalytic domain undergoes rapid unfolding when PARP1 encounters a DNA break, relieving the autoinhibition	(150)
	Requires few microliters of low micromolar concentration of proteins			
	Detects co-existing protein conformations			
nMS	Widely applicable to samples varying in mass, degree of flexibility, symmetry, and polydispersity	The relative abundance of detected complexes may deviate from that in solution because of different complexes' distinct ionization, transmission, and detection probabilities	A MutS intermediate state that is simultaneously bound to a DNA mismatch and a nucleoside triphosphate is detected	(160)
	Multiple oligomeric states of proteins and complexes can be analyzed simultaneously	In the gas phase, hydrophobic interactions weaken, and electrostatic interactions become stronger than in-solution making detecting specific assemblies impossible without prior crosslinking	The effect of mutations in <i>Staphylococcus aureus</i> uracil-DNA glycosylase inhibitor on its association with uracil-DNA glycosylase is analyzed	(161)
	Requires few microliters of low micromolar concentration of proteins			
	Does not require samples to be chemically labeled or crosslinked.			
Integrative structural modeling	Computationally integrates structural data from multiple sources to enable structure determination of macromolecules and their complexes	Accuracy of output models relies on quantity and resolution of the input information	Structures of protein complexes are derived by integrative modeling using structural proteomics data as well as data collected from conventional methods	(167,170,174,176)

LC-MS/MS allows quantitative analysis of thousands of non-redundant phosphorylated or ubiquitinated peptides and exact sites of modification after searching MS spectra against putative protein sequences in a species-specific proteome database.

Phospho-proteomic studies reveal that there is a considerable variation in the kinetics and directionality of phospho-proteome changes following DNA damage (43,44), and that over one-third of the captured phosphopeptides are dephosphorylated within minutes of DNA damage (43), indicating that phosphatases are not only involved in counter-balancing DNA double-strand break (DSB)-induced phosphorylation events by resetting activated DDR factors to the initial homeostatic state following the repair of damaged DNA, but also play a primary role in initiating the repair process by removing constitutive phosphorylation that inhibits the function of DNA repair factors to activate these proteins (45,46). Ubiquitome analysis has been also successfully used to profile changes in the ubiquitome in response to DNA damage (47). Because ubiquitination often regulates protein stability, pairing ubiquitin proteomics with abundance proteomics and proteasome inhibitor treatments (e.g. MG132) is often used to identify proteins whose levels are regulated by ubiquitination (48,49).

ADP-ribose (ADPr) unit is transferred by poly(ADP-ribose) polymerases (PARPs) from NAD⁺ to a carboxylic acid residues (Glu, Asp) (50–52), but other residues containing a thiol group (Cys), an alcohol group (Ser, Thr, Tyr), a guanidinium group (Arg) or Lys are also found to serve as acceptors (53–63). Despite the role of PARylation in DDR, high-throughput profiling of ADPr-modified proteins has been hindered due to the difficulty of PAR enrichment. However, the Afl521 macrodomain, which has a strong ADPr-binding affinity, was shown to successfully enrich and identify ADP-ribosylated proteins from cell lysates induced under different genotoxic stress conditions (64). Interestingly, another study reveals that Ser residues are the major ADPr site in response to oxidative stress, and that significant portion of these Ser residues are co-targeted by mutually exclusive ADP-ribosylation and phosphorylation, indicating a crosstalk between the two PTMs (62) that remains to be further validated for their roles in cellular stress response.

For the MS analysis of ADPr proteins, numerous modifiable amino acid residues and the labile nature of the bond between ADPr and the amino acid residues pose a challenge for site-specific localization using the conventional collision-induced dissociation (CID) MS/MS method. Instead, the radical-driven electron transfer dissociation (ETD) method has proven to preserve labile ADPr and allow confident determination of the exact ADPr-modified amino acid residues (62). Recently, the activated ion ETD (AI-ETD) method that uses infrared photoactivation during the ETD reaction was developed, which allows to overcome the high degree of non-dissociative electron transfer (ETnoD) during regular ETD, especially for ADPr-peptide precursors with low m/z ratio (63). AI-ETD was shown to map significantly more ADPr sites than ETD and identify more physiological ADPr sites with less input materials, many of which were not previously reported.

Protein-protein interaction mapping by affinity purification MS

Coordinated regulation of protein complex formation is central to the DDR response. For example, BRCA1, a hereditary breast and ovarian cancer susceptibility gene (65–67) that plays a crucial role in the repair of DNA double-strand breaks (DSB) (68), carries out its repair functions in concert with a large number of proteins involved at distinct steps of homologous recombination (69). Affinity Purification-MS (AP-MS) is a proteomics method to systematically identify PPIs by expressing and purifying affinity-tagged ‘bait’ proteins in cells and detecting stably bound protein interactors (prey) by MS (46,70–72). Samples that express the affinity tag alone or repress the expression of tagged bait protein are often used in parallel as negative control for statistical PPI scoring (Figure 1). Alternatively, direct immunoprecipitation of the endogenous protein of interest may also be performed if a sensitive and specific antibody exists (73).

Recently, Kim *et al.* applied AP-MS to comprehensively analyze the BRCA1 interactome and how these interaction profiles change with pathogenic BRCA1 mutations (46). This AP-MS study revealed a number of previously unidentified BRCA1-interacting proteins in addition to known interactors and showed that these interactions are altered by mutations in different domains of BRCA1: C-terminal BRCT domain mutants (S1655F, 5382insC, M1775R) completely or significantly lost the interaction with multiple homologous recombination repair proteins (including FAM175A, BRIP1, RBBP8, UIMC, BRE), whereas N-terminal RING domain mutants (I26A, C61G, R71G) maintain these interactions although they are not as strong as with wild-type BRCA1. These results concur with previous suggestions that RING domain mutants may retain residual activity; for example, the C61G variant is only moderately sensitive to cisplatin and PARP inhibitors and becomes readily resistant to these drugs (74,75). AP-MS of transcriptional enhanced associate domain family members 1–4 (TEADs) also revealed their association with DNA repair proteins (XRCC5, XRCC6, PARP1, RIF1) to facilitate cellular recovery from DNA damage (76).

Protein-DNA interaction mapping

The DDR also involves choreographed interactions between DNAs and proteins and cataloging the dynamic spatial and temporal changes in DNA/chromatin-associated proteome is essential to thoroughly understand the DNA repair and replication processes. To identify proteins enriched at DNA replication forks, isolation of Proteins on Nascent DNA (iPOND) and Nascent Chromatin Capture (NCC) techniques were developed (77,78). Additionally, chromatin immunoprecipitation (ChIP) followed by MS (ChIP-MS) was devised to probe chromatin-bound proteins at specific gene loci as well as genome-wide (79,80).

In iPOND, newly synthesized DNAs are labeled with 5-ethynyl-2'-deoxyuridine (EdU) incorporation by the replisome DNA polymerases. After crosslinking proteins to DNAs with formaldehyde, biotin is covalently conjugated to an alkyne group of EdU using click chemistry, and 100–200 bp protein-bound chromatin fragments are generated

by sonication. Finally, proteins bound on the EdU labeled chromatin fragments are captured by streptavidin beads and analyzed by quantitative MS after protein-DNA de-crosslinking. Duration of EdU labeling (pulse) and subsequent chase after removal of EdU determines the spatial information of proteins relative to the replication fork. Furthermore, the pulse and chase samples are critical to distinguish proteins engaged in DNA replication from those that are simply associated with bulk chromatin (81). NCC is similar to iPOND except it utilizes biotin-dUTP instead of EdU, such that NCC does not require a biotin conjugation step. However, due to impermeability of biotin-dUTP, it is necessary to permeabilize cell membranes in a hypotonic buffer during biotin-dUTP labeling. Quantitative MS analysis of enriched proteins in iPOND and NCC has been performed by either label-free or SILAC/iTRAQ/TMT-based protocols (78,82–86) to identify new DNA replication proteins. In these studies, labeling protocols tend to provide more quantitatively precise and reproducible datasets than label-free methods (81). ChIP-MS is similar to AP-MS, in that both methods rely on antibody-based immunoprecipitation/purification, but ChIP-MS is used to specifically identify PPIs on crosslinked chromatin. Proteins in the male-specific lethal (MSL) complex on the chromatin template as well as proteins enriched in telomeres were monitored by ChIP-MS and successfully identified novel proteins relevant to the function of MSL complex and telomere maintenance, respectively (79,80).

Proximity-dependent labeling MS

Proximity-dependent labeling has emerged as a complementary method to AP-MS to study PPIs and relies on a promiscuous biotin protein ligase (BirA/BioID/TurboID) (87,88) or an engineered ascorbic acid peroxidase (APEX) (89) fused to a protein of interest (bait) to label proteins in close proximity to the bait protein through covalent transfer of biotin or biotin derivatives in a distance-dependent manner. Upon addition of biotin, the proximal proteins are biotinylated, enriched with streptavidin beads, and identified by quantitative MS (Figure 1). The covalent biotin labeling of proximal proteins is particularly useful to capture transient or weaker PPIs, compatible with purification under harsh cell lysis conditions (e.g. higher detergent and urea concentrations in lysis buffer), and offers spatial and temporal resolution of PPIs within various sub-cellular locations (87,89,90). APEX-based proximity proteomics characterized the endogenous network of BRCA1, 53BP1, and MDC1 and revealed Shieldin is a 53BP1 effector complex in DSB repair (91). Recently, biotin ligase enzymes have been engineered to split into two parts which enable contact-specific proximity-dependent labeling. While the two split enzyme parts remain inactive apart, they become active when bound together by PPIs. (92–94). These split enzyme-based approaches offer greater targeting specificity of biotinylation than full-length enzymes alone and provide powerful tools for validating and functionally characterizing high-confidence PPIs identified by AP-MS and/or other methods (95). Abbasi and Schild-Poulter applied BioID to the Ku interactome analysis and identified ~250 Ku70-associated proteins in HEK293 cells, many of

which are involved in RNA metabolism, chromatin remodeling, and microtubule dynamics as well as DNA repair and replication, revealing additional cellular roles of the Ku complex (96). Another BioID analysis of ATRX revealed its association with SLF2 that helps inhibit telomere exchanges (97).

MS data quantification, interpretation and integration using bioinformatics and network modeling

Once data is collected, they first pass through quality control. Specifically, one must ensure reproducibility in peptide detections and intensity quantifications between biological replicates. Replicates that fail quality control are discarded. Global proteomics data of whole cell lysates (e.g. abundance proteomics, phosphoproteomics, ubiquitin proteomics, or similar) can be subjected to relative quantification to compare each experimental condition to a control. One software, called MSstats, offers quantitative and comparative measurement of protein levels, PTMs, or PPIs between two biological states, such as diseased versus healthy or mutant versus wild-type (98). This data preprocessing step generates \log_2 fold changes and *P*-values, and thus defines differentially expressed or regulated proteins (DEP). For AP-MS data, additional algorithms are used to remove background signal; even after performing an affinity purification, some proteins detected may be non-specifically binding the tube or other reagents used during sample preparation. The analysis workflow typically utilizes software packages (e.g. compPASS, SAINT, MIST) (99–103) to assign quantitative scoring metrics (i.e. abundance, specificity, and reproducibility) to all pairwise PPIs and filter high-confidence interacting proteins (e.g. SAINTscore ≥ 0.9 , Bayesian false discovery rate ≤ 0.05). These quantitative, differential, pre-processed data are subsequently interpreted using more sophisticated computational algorithms, which enable the integration of new data with prior knowledge.

A highly-used, and often first-line, approach to data interpretation is gene enrichment analysis, which enables the assignment of biological terms to differentially abundant proteins. Gene sets have been defined by several groups, many of which are distributed by the Molecular Signatures Database (MSigDB, www.gsea-msigdb.org). Commonly used databases include Reactome (104), KEGG (105), Wikipathways (106) and Gene Ontology (107). One enrichment analysis algorithm, called Gene Set Enrichment Analysis (GSEA), creates a ranked list of proteins derived from experiment (often ranked by \log_2 fold change) and compares it to annotated gene sets (e.g. cellular pathways) to determine whether genes in each gene set occur toward the top or bottom (large or small fold changes) of the list. If so, this indicates the gene set is 'correlated' with proteins under regulation in the experiment (108). A similar approach uses a Fisher's exact or hypergeometric test to assess the overlap of DEP to each gene set (109). The end result is the same: significantly enriched gene sets represent cellular pathways that are regulated in the experiment. These results provide general biological understanding and often guide subsequent validation experiments. Importantly, these gene set databases possess many terms related to the DNA damage response. In Gene Ontology, there are over 40 terms

that either possess ‘DNA repair’ or ‘DNA damage,’ spanning pathways such as ‘double strand break repair via non-homologous end joining’ and ‘single strand break repair via break induced replication’. Gene enrichment analysis is a good way to gain a high-level summary of the data but is often insufficient to more deeply interrogate the data to reveal mechanistic hypotheses for further experimental testing.

Network modeling, a branch of computational modeling that uses pre-defined networks of gene-gene interactions, can be used to integrate MS datasets and extract biological insight at the interface between distinct proteomics datasets. Several freely-available, large (often genome-wide scale) biological networks exist [e.g. PathwayCommons (110), STRING (111), ReactomeFI (112)], which capture decades of experimental knowledge detailing how genes and proteins physically and/or functionally interact in protein complexes or functional pathways. Data from an experiment can be overlaid onto these networks and analyzed using network modeling algorithms to understand biological interconnectivity across multiple datasets (Figure 1).

One approach is called network propagation (113). Briefly, proteins of interest from the experiment (‘nodes’) are labeled in the network. This signal is then propagated through each proteins’ molecular interactions (‘edges’) to nearby nodes in an iterative manner, often using a random walk or heat kernel mathematical process. A permutation test is subsequently performed by shuffling node labels or randomizing network structure in order to assign a p-value to each node, which represents the network proximity, and functional connectivity, of each gene to the initial labeled nodes. Network propagation can be used to identify the extent of intersection between two or more gene lists by evaluating their shared neighborhoods within the network. This approach is particularly useful when two gene lists have poor overlap at the gene level but high potential overlap at the pathway level. We and others have previously used this approach to discover links between the human papillomavirus (HPV) virus-host physical interactome and the mutational landscape of HPV-negative tumors, revealing converging routes to oncogenesis (114), and to integrate genome-wide CRISPR hits from human cells infected with distinct coronaviruses, revealing host pathway targets for pan-coronavirus therapeutics (115).

Data integration using network propagation can be especially useful when a single dataset is insufficient to extract the level of mechanistic insight desired. It can also be used as an exploratory tool to screen for interesting connections between distinct data types. Furthermore, network propagation is useful when data is sparse in nature, which can be the case for proteomics datasets; in these cases, network propagation can be used to explore how proteins coalesce into the same complex or pathway. However, because network propagation provides a wealth of information, the results can be difficult to interpret. Additional approaches are needed to make it easier for users to pinpoint the most relevant mechanisms from the resulting network.

Network modeling tools could be transformative for discovering new mechanisms underlying DDR signaling. For example, one might be interested in understanding the mechanisms underlying how a specific cancer driver point mutation (e.g. BRCA1 mutation) drives cancer pro-

gression. To study this, one could perform global proteomics (abundance proteomics and PTM analysis) comparing wild-type cells to cells that possess the mutation of interest. Additionally, one could perform differential AP-MS or proximity-dependent labeling to quantitatively compare how a protein-coding point mutant alters PPIs. The protein-level overlap between these datasets might be small; indeed, proteins detected to be differentially regulated in the global proteomics data may reflect gene expression responses several steps downstream the mutant, and thus may not be contained within the physical interactome. Network propagation could be used here to reveal the molecular networks that ‘connect’ the differential PPIs to the downstream signaling events. Specifically, network propagation could be performed in two parts. First, label and propagate nodes corresponding to differentially interacting PPIs. Second, label and propagate nodes corresponding to differentially regulated proteins from the global proteomics studies. Genes that are significant ($P < 0.05$) in both propagation analyses represent molecular networks that lie at the interface between BRCA1 mutant protein interactions and the downstream signaling effects of the BRCA1 mutation. Such an analysis may delineate a novel cellular pathway linking a BRCA1 mutation to its signaling ramifications, and thereby suggest subsequent validation experiments to further understand BRCA1 function. This approach is specifically applicable to DDR signaling because one important question in the field is how specific mutants (especially variants of uncertain significance of key DDR proteins) impact protein complex formation and DNA repair outcomes, and may provide a more comprehensive view of proteomic changes and signaling pathways induced by a given mutation, which is often difficult to capture by any single proteomics attempt.

STRUCTURAL PROTEOMICS

Crosslinking MS (XL-MS)

XL-MS is a powerful technique that informs protein-protein interactions and provides structural insights into proteins and protein assemblies by identifying amino acid residue pairs that are within close spatial proximity (116–121). Typically, this approach involves crosslinking or surface modification of protein complexes followed by proteolysis and MS identifications (Figure 1). Applications of XL-MS vary from studying macromolecular complexes (122,123) to their structural dynamics (124,125).

Qiu et al. conducted a comprehensive investigation of DNA-binding proteins *in vivo* using XL-MS and identified >100 proteins, including those involved in DNA replication and repair (126). Another study found that peroxisome proliferator activated receptor γ (PPAR γ) interacts with the MRE11-RAD50-NBS1 (MRN) complex and the E3 ubiquitin ligase UBR5 by AP-MS and that PPAR γ promotes UBR5-mediated degradation of ATM interactor (ATMIN), leading to ATM activation and the initiation of DNA repair upon DNA damage (127). Subsequent XL-MS experiment identified PPAR γ peptides crosslinked to NBS1, which are located in the DNA-binding domain and the ligand-binding domain of PPAR γ , indicating that PPAR γ interact with the MRN complex via NBS1, and

that NBS1 binding may interfere with PPAR γ transcription activator function. Holzer et al. elucidated the structural basis for the human Timeless–Tipin complex using various complementary tools including XL-MS as well as XRC and nMS. They used lysine-specific crosslinker and demonstrated that the Timeless–Tipin interaction is extensive, such that three distinct regions of Timeless between amino acids 400 and 1100 are crosslinked to three different sites in Tipin. Using this integrated approach they were able to show that Timeless forms a binding surface capable of surrounding the smaller protein Tipin (128).

There are a few inherent challenges of XL-MS: (i) the combinatorial nature of the crosslinks and mixture of unmodified peptides increases the complexity of the spectra; (ii) sequencing of crosslinked peptides often results in low-quality or uninterpretable MS/MS spectra because of insufficient or complex fragmentation; (iii) identification and filtering of nonspecific interactions are challenging especially in validating novel observed PPIs. Thus, high-throughput bioinformatic approaches are required to decipher complete structural information. Some of these challenges can be overcome by increasing data acquisition rates, higher sequencing capacity, and novel hybrid fragmentation techniques like electron transfer/higher-energy collision dissociation (EThcD) and ultraviolet photodissociation (UV-PD) (129). Regardless, XL-MS is a valuable tool for structural characterization of protein complexes by providing real-time insight into the PPIs and protein-DNA/RNA interactions that may undergo structural changes in response to DNA damage.

Hydrogen/deuterium exchange MS (H/DX-MS)

H/DX-MS measures dynamic changes in protein structures over a specified time scale by measuring the relative uptake of protein backbone amide hydrogen for deuterium (and *vice versa*) (Figure 1). The rate of this reaction is dependent on various factors like solvent accessibility, chemical properties of the underlying amino acid sequence, stability of hydrogen bonding networks, and inductive effect of the neighboring groups (130–133). It has been applied to explore protein/protein complex conformations (134), protein-ligand binding sites (135), allosteric effects of proteins (136), protein folding dynamics (137), intrinsically disordered proteins (138), and protein – membrane interactions (139). The resolution of data depends on the fragmentation technique adopted during the MS data acquisition. Residue-level information can be derived from gas-phase fragmentation of the deuterated peptides using ETD or electron-capture dissociation (ECD) (140,141). The vibrationally cold energy utilized in these techniques results in minimized hydrogen scrambling (i.e. redistribution of the exchanged protons within a peptide) as opposed to other methods like collision-induced dissociation (CID) (142–144). In 2013, Resetca et al. developed a method called time-resolved electrospray ionization mass spectrometry (TRESI-MS), which uses a microfluidic chip in-line with all the steps involved in a ‘bottom-up’ H/DX workflow (145). This development provided faster sample preparation times and improved reproducibility, making it feasible to characterize rapid structural transitions that occur

during protein folding (146), ligand binding (147), or post-translational modification (148).

Ahmad et al. applied H/DX-MS to explore ssDNA-driven dynamics of the individual domains of single-strand DNA (ssDNA)-binding protein, replication protein A (RPA) (149). They observed changes in DNA binding domains (DBD) A-E of RPA upon ssDNA binding (149): DBD-A and B were shown to be dynamic and no protection was observed upon ssDNA binding, while DBD-C displayed extensive changes signifying a major role in stabilizing RPA on ssDNA binding. These H/DX-MS data proposed that DBD-A and -B of RPA serve as the dynamic half and DBD-C, -D and -E function as the less-dynamic half, such that ssDNA buried in DBD-A and B could be more accessible to RPA-interacting proteins (149). Dawicki-McKenna et al. monitored conformational dynamics of PARP1 using H/DX-MS and observed a >10 000-fold faster exchange within specific portions of the helical subdomain (HD) of the catalytic domain, which undergoes rapid unfolding when PARP1 encounters a DNA break (150). This result indicates that DNA damage detection by PARP1 relieves the autoinhibition posed by the HD.

Native MS (nMS) and intact protein analysis

nMS and intact protein MS involve analyzing and characterizing macromolecules, predominantly protein complexes, while retaining their inter and intramolecular interactions (Figure 1). To ionize protein complexes directly from aqueous solutions, buffers in which protein complexes are suspended during purification steps are exchanged into volatile buffers, such as ammonium acetate which has a pH range of 6–8 and evaporates readily during ionization. nMS resorts to a soft transfer of the analyte from the solution to the gas phase using nano-electrospray ionization (nESI). During ionization the macromolecules are charged, and the solvent molecules are stripped off before the ions reach the mass analyzer under (ultra)high-vacuum conditions. One may argue that proteins may not always retain a fully native state during nMS. However, under careful and optimized conditions, nMS provides gas-phase ions of proteins that retain many of their native features (151,152). The stoichiometry of a protein complex can be deduced from the calculated mass, and structural preference can be informed by the number of charges it carries (153). Structured protein conformations typically carry limited charges per unit mass due to the limited solvent accessibility owing to their compact nature. In contrast, unfolded/denatured proteins have a larger solvent-accessible surface area which enables them to accumulate higher charges (154–156). nMS combined with CID or surface-induced dissociation (SID) disrupts noncovalent interactions between protein subunits based on their strength and thus can help decipher the stoichiometry and topology of protein complexes (157,158). SID is a rapid dissociation that is almost a single-step energy deposition process when ions collide with a surface. No extensive unfolding of proteins and subunits is observed in this process (159).

nMS analysis of *Escherichia coli* MutS detected a MutS dimer state that is simultaneously bound to a DNA mismatch and a nucleoside triphosphate, which has been

thought as an intermediate for faithful initiation of mismatch repair (160). Furthermore, this analysis could distinguish between binding of nucleoside diphosphate and nucleoside triphosphate in the two composite ATPase sites of MutS dimer, providing mechanistic and structural insight into how mismatched DNA binding and release are controlled by binding and hydrolysis of ATP. This study demonstrates that nMS is a powerful tool to detect mechanistically relevant reaction intermediates of allosteric enzyme complexes that cannot be easily addressed by other techniques. Besides, nMS was utilized to quantitatively measure the macromolecular association between uracil–DNA glycosylase and its inhibitor in *Staphylococcus aureus* and revealed that naturally occurring mutations in the inhibitor lead to appreciable changes in the dissociation constants for the complex (161).

Despite significant advances in nMS methods, there are some analytical challenges that remain to be addressed. For example, complexes expressed and purified from mammalian cells often carry PTMs and/or small ligands, which cause small mass differences and result in a heterogeneous mixture. It is challenging to resolve heterogeneous protein complexes, although charge changing experiments can sometimes help solve this problem (162,163). Additionally, biomolecules need to be prepared to be amenable to native electrospray ionization. But they often require nMS-incompatible components, including metal ions, small molecules, or lipids to maintain the integrity. Improving gas-phase desalting and adduct removal techniques would not only maintain the biomolecule in a ‘native-like’ environment but also reduce adduct formation and aid solvent removal at low activation energies, resulting in clean native spectra (164). Alternative ionization methods, such as desorption electrospray ionization (DESI) that enable biomolecules to directly ionize from biological tissues (165) may also help mitigate this problem. DESI involves directing a fast-moving charged solvent stream at an angle relative to the sample surface under atmospheric conditions to extract analytes from the surfaces resulting in the secondary ions being propelled toward the mass analyzer (166).

Integrative structural modeling

Data from the diverse structural proteomics methodologies discussed above can be combined to obtain maximum structural insight through integrative modeling (IM) (23,167,168). Structure determination of protein assemblies is vital for a mechanistic understanding of their function. IM takes experimental input data from structural, biochemical, proteomic, and genetic studies to optimize and compute a comprehensive model of protein complexes with properties that satisfy information within the uncertainty of the data (23,167,169,170). The primary sources of input information for IM are XRC, NMR spectroscopy or comparative modeling, XL-MS, H/DX-MS, stoichiometric data from nMS, solution scattering data from small-angle X-ray scattering (SAXS), and protein–protein interactions from AP-MS. Structures of several protein complexes have been derived using integrative methods to explain their architecture and evolutionary principles of large assemblies (167,171–173), rationalize the effect of disease

mutations (167,174), and derive structural heterogeneity of flexible protein complexes (170,175,176).

Gutierrez *et al.* applied the IM approach to investigate in-solution architecture of CSN9-dependent structural changes and conformational dynamics of the COP9 signalosome (CSN) complex (170). Based on XL-MS data, XRC structures, and comparative models, they proposed that CSN9 binding triggers the CSN complex to adopt a configuration that facilitates CSN–CRL interactions, thereby enhancing CSN deneddylase activity. Most importantly, they observed additional conformations and configurations of CSN in solution that were absent in the static structure.

TFIIH is a conserved 10-subunit complex that plays a critical role in both RNA polymerase II transcription (174,177) and DNA repair (178). Mutations in TFIIH subunits lead to different cancers (179,180) and autosomal recessive disorders, such as xeroderma pigmentosum (XP), and trichothiodystrophy (TTD) (178). EM studies derived structural information of yeast and human TFIIH (181–184). By adopting an integrative approach using data from XL-MS, biochemical analyses, and previously reported EM maps, Luo *et al.* determined the molecular architecture of both human and yeast TFIIH (174). They identified four new conserved regions that function as hubs for TFIIH assembly and >35 conserved topological features within TFIIH, illuminating a network of interactions involved in TFIIH assembly and regulation of its activities.

FUTURE PERSPECTIVES

Development and application of myriads of tools over the past few years have helped in shaping our understanding of how protein complexes, PTM signaling, PPIs and protein structure/function regulate the DDR. This opens up an arena of DDR research in the future that can be decoded with integrated multidisciplinary approaches to gain a deeper mechanistic understanding of these genome maintenance processes both *in vitro* and *in vivo*.

There are several areas of mass spectrometry that are currently transforming the field. One is single-cell proteomics, which is a relatively new field focusing on analyzing the proteome of individual cells. This technique can provide important insights into cellular heterogeneity in signaling responses (185,186), which could reveal cell-to-cell variability in DDR.

Ion mobility mass spectrometry (IM-MS) is another powerful technique that is used to further separate ions based on their collisional cross section, affected by their size, shape, and charge (187). IM-MS is particularly useful for deconvolving complex mixtures where traditional mass spectrometry may struggle to differentiate between closely related species (188). Furthermore, IM-MS can be used in combination with other mass spectrometry-based techniques, such as native MS and top-down proteomics, to obtain structural information about proteins and protein complexes (189,190).

Additionally, data independent acquisition (DIA) methods such as SWATH and diaPASEF allow for the simultaneous quantification of thousands of proteins in complex mixtures (191–194). DIA methods are particularly useful

for large-scale proteomics studies, where traditional DDA methods may be limited by their dynamic range and sensitivity. SWATH and diaPASEF are both highly reproducible and can provide accurate and precise quantitative measurements, making them valuable tools for proteome profiling.

Although further innovation and addressing limitations in existing technologies remain crucial, we are about to unravel the complexity of regulatory mechanisms in DNA repair at an extraordinary level of detail and at a rapid pace.

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

ACKNOWLEDGEMENTS

We would like to thank members of the Krogan lab for critical reading of the manuscript. Figure and graphical abstract were created with Biorender.com.

FUNDING

National Institutes of Health [U54CA274502, P50AI150476, U19AI135990 to N.J.K.]; Martha and Bruce Atwater Breast Cancer Research Fund; Benioff Initiative for Prostate Cancer Research; UTHSCSA Startup Research Fund (to M.K.); F. Hoffmann-La Roche and Vir Biotechnology and gifts from The Ron Conway Family (to N.J.K.).

Conflict of interest statement. None declared.

REFERENCES

- Peng, J. and Gygi, S.P. (2001) Proteomics: the move to mixtures. *J. Mass Spectrom.*, **36**, 1083–1091.
- Wysocki, V.H., Resing, K.A., Zhang, Q. and Cheng, G. (2005) Mass spectrometry of peptides and proteins. *Methods*, **35**, 211–222.
- Eng, J.K., McCormack, A.L. and Yates, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, **5**, 976–989.
- Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.
- Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics*, **1**, 376–386.
- Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K.A. and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.*, **75**, 1895–1904.
- Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S. *et al.* (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics*, **3**, 1154–1169.
- Wiese, S., Reidegeld, K.A., Meyer, H.E. and Warscheid, B. (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, **7**, 340–350.
- McAlister, G.C., Huttlin, E.L., Haas, W., Ting, L., Jedrychowski, M.P., Rogers, J.C., Kuhn, K., Pike, I., Grothe, R.A., Blethrow, J.D. *et al.* (2012) Increasing the multiplexing capacity of tmts using reporter ion isotopologues with isobaric masses. *Anal. Chem.*, **84**, 7469–7478.
- Rauniyar, N. and Yates, J.R. 3rd (2014) Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.*, **13**, 5293–5309.
- Braun, C.R., Bird, G.H., Wühr, M., Erickson, B.K., Rad, R., Walensky, L.D., Gygi, S.P. and Haas, W. (2015) Generation of multiple reporter ions from a single isobaric reagent increases multiplexing capacity for quantitative proteomics. *Anal. Chem.*, **87**, 9855–9863.
- Gillet, L.C., Navarro, P., Tate, S., Röst, 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.
- Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanutelli, V., Butscheid, Y., Escher, C. *et al.* (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics*, **14**, 1400–1410.
- Hu, A., Noble, W.S. and Wolf-Yadlin, A. (2016) Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res*, **5**, 419.
- Wright, P.E. and Dyson, H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
- Boehr, D.D., Nussinov, R. and Wright, P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, **5**, 789–796.
- Politis, A. and Schmidt, C. (2018) Structural characterisation of medically relevant protein assemblies by integrating mass spectrometry with computational modelling. *J. Proteomics*, **175**, 34–41.
- Liu, F., Lössl, 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.
- Adhikary, S., Deredge, D.J., Nagarajan, A., Forrest, L.R., Wintrose, P.L. and Singh, S.K. (2017) Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E1786–E1795.
- Martens, C., Shekhar, M., Lau, A.M., Tajkhorshid, E. and Politis, A. (2019) Integrating hydrogen-deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins. *Nat. Protoc.*, **14**, 3183–3204.
- Arlt, C., Nutschan, K., Haase, A., Ihling, C., Tänzler, D., Sinz, A. and Sawers, R.G. (2021) Native mass spectrometry identifies the HybG chaperone as carrier of the Fe(CN)2CO group during maturation of *E. coli* [NiFe]-hydrogenase 2. *Sci. Rep.*, **11**, 24362.
- Bandekar, S.J., Chen, C.-L., Ravala, S.K., Cash, J.N., Avramova, L.V., Zhalnina, M.V., Gutkind, J.S., Li, S. and Tesmer, J.J.G. (2022) Structural/functional studies of Trio provide insights into its configuration and show that conserved linker elements enhance its activity for Rac1. *J. Biol. Chem.*, **298**, 102209.
- Rout, M.P. and Sali, A. (2019) Principles for integrative structural Biology studies. *Cell*, **177**, 1384–1403.
- Villén, J., Beausoleil, S.A., Gerber, S.A. and Gygi, S.P. (2007) Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 1488–1493.
- Wang, Y., Yang, F., Gritsenko, M.A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M.E., Lopez-Ferrer, D., Reno, T. *et al.* (2011) Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics*, **11**, 2019–2026.
- Heusel, M., Bludau, I., Rosenberger, G., Hafen, R., Frank, M., Banaei-Esfahani, A., van Drogen, A., Collins, B.C., Gstaiger, M. and Aebersold, R. (2019) Complex-centric proteome profiling by SEC-SWATH-MS. *Mol. Syst. Biol.*, **15**, e8438.
- Heusel, M., Frank, M., Köhler, M., Amon, S., Frommelt, F., Rosenberger, G., Bludau, I., Aulakh, S., Linder, M.I., Liu, Y. *et al.* (2020) A global screen for assembly state changes of the mitotic proteome by SEC-SWATH-MS. *Cell Syst.*, **10**, 133–155.
- Warmoes, M., Jaspers, J.E., Xu, G., Sampadi, B.K., Pham, T.V., Knol, J.C., Piersma, S.R., Boven, E., Jonkers, J., Rottenberg, S. *et al.* (2013) Proteomics of genetically engineered mouse mammary

- tumors identifies fatty acid metabolism members as potential predictive markers for cisplatin resistance. *Mol. Cell. Proteomics*, **12**, 1319–1334.
29. Pines, A., Kelstrup, C.D., Vrouwe, M.G., Puigvert, J.C., Typas, D., Misovic, B., de Groot, A., von Stechow, L., van de Water, B., Danen, E.H.J. *et al.* (2011) Global phosphoproteome profiling reveals unanticipated networks responsive to cisplatin treatment of embryonic stem cells. *Mol. Cell. Biol.*, **31**, 4964–4977.
 30. Beli, P., Lukashchuk, N., Wagner, S.A., Weinert, B.T., Olsen, J.V., Baskcomb, L., Mann, M., Jackson, S.P. and Choudhary, C. (2012) Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol. Cell*, **46**, 212–225.
 31. Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R. 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y. *et al.* (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*, **316**, 1160–1166.
 32. Mu, J.-J., Wang, Y., Luo, H., Leng, M., Zhang, J., Yang, T., Besusso, D., Jung, S.Y. and Qin, J. (2007) A proteomic analysis of ataxia telangiectasia-mutated (ATM)/ATM-Rad3-related (ATR) substrates identifies the ubiquitin-proteasome system as a regulator for DNA damage checkpoints. *J. Biol. Chem.*, **282**, 17330–17334.
 33. Panier, S. and Boulton, S.J. (2014) Double-strand break repair: 53BP1 comes into focus. *Nat. Rev. Mol. Cell Biol.*, **15**, 7–18.
 34. Schwertman, P., Bekker-Jensen, S. and Mailand, N. (2016) Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat. Rev. Mol. Cell Biol.*, **17**, 379–394.
 35. Kannouche, P.L., Wing, J. and Lehmann, A.R. (2004) Interaction of human DNA polymerase η with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol. Cell*, **14**, 491–500.
 36. Andreassen, P.R., D'Andrea, A.D. and Taniguchi, T. (2004) ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev.*, **18**, 1958–1963.
 37. Caldecott, K.W. (2014) Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA Repair*, **19**, 108–113.
 38. Palazzo, L., Mikošević, P., Mikoš, A. and Ahel, I. (2019) ADP-ribosylation signalling and human disease. *Open Biol*, **9**, 190041.
 39. Leutert, M., Rodríguez-Mias, R.A., Fukuda, N.K. and Villén, J. (2019) R2-P2 rapid-robotic phosphoproteomics enables multidimensional cell signaling studies. *Mol. Syst. Biol.*, **15**, e9021.
 40. Yu, L.-R. and Veenstra, T. (2013) Phosphopeptide enrichment using offline titanium dioxide columns for phosphoproteomics. *Methods Mol. Biol.*, **1002**, 93–103.
 41. Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D. and Gygi, S.P. (2003) A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.*, **21**, 921–926.
 42. Udeshi, N.D., Mertins, P., Svinikina, T. and Carr, S.A. (2013) Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.*, **8**, 1950–1960.
 43. Besson, A., Schmidt, A., Ziv, Y., Elkon, R., Wang, S.-Y., Chen, D.J., Aebersold, R. and Shiloh, Y. (2010) ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. *Sci. Signal.*, **3**, rs3.
 44. Bennetzen, M.V., Larsen, D.H., Bunkenborg, J., Bartek, J., Lukas, J. and Andersen, J.S. (2010) Site-specific phosphorylation dynamics of the nuclear proteome during the DNA damage response. *Mol. Cell. Proteomics*, **9**, 1314–1323.
 45. Lee, D.-H. and Chowdhury, D. (2011) What goes on must come off: phosphatases gate-crash the DNA damage response. *Trends Biochem. Sci.*, **36**, 569–577.
 46. Kim, M., Park, J., Bouhaddou, M., Kim, K., Rojca, A., Modak, M., Soucheray, M., McGregor, M.J., O'Leary, P., Wolf, D. *et al.* (2021) A protein interaction landscape of breast cancer. *Science*, **374**, eabf3066.
 47. Elia, A.E.H., Boardman, A.P., Wang, D.C., Huttlin, E.L., Everley, R.A., Dephoure, N., Zhou, C., Koren, I., Gygi, S.P. and Elledge, S.J. (2015) Quantitative proteomic atlas of ubiquitination and acetylation in the DNA damage response. *Mol. Cell*, **59**, 867–881.
 48. Halim, V.A., García-Santisteban, I., Warmerdam, D.O., van den Broek, B., Heck, A.J.R., Mohammed, S. and Medema, R.H. (2018) Doxorubicin-induced DNA damage causes extensive ubiquitination of ribosomal proteins associated with a decrease in protein translation. *Mol. Cell. Proteomics*, **17**, 2297–2308.
 49. Chen, X., Li, P., Zhang, G., Kang, L., Qin, B., Mao, X., Qin, M., Cao, Y., Wang, Y. and Guan, H. (2020) Comprehensive profiling of proteome and ubiquitome changes in Human lens epithelial cell line after ultraviolet-B irradiation. *ACS Omega*, **5**, 32171–32182.
 50. D'Amours, D., Desnoyers, S., D'Silva, I. and Poirier, G.G. (1999) Poly(ADP-ribosylation) reactions in the regulation of nuclear functions. *Biochem. J.*, **342**, 249–268.
 51. Tallis, M., Morra, R., Barkauskaite, E. and Ahel, I. (2014) Poly(ADP-ribosylation) in regulation of chromatin structure and the DNA damage response. *Chromosoma*, **123**, 79–90.
 52. Li, P., Zhen, Y. and Yu, Y. (2019) Site-specific analysis of the asp- and Glu-ADP-ribosylated proteome by quantitative mass spectrometry. *Methods Enzymol.*, **626**, 301–321.
 53. McDonald, L.J. and Moss, J. (1994) Enzymatic and nonenzymatic ADP-ribosylation of cysteine. *Mol. Cell. Biochem.*, **138**, 221–226.
 54. Bonfiglio, J.J., Fontana, P., Zhang, Q., Colby, T., Gibbs-Seymour, I., Atanassov, I., Bartlett, E., Zaja, R., Ahel, I. and Matic, I. (2017) Serine ADP-ribosylation depends on HPF1. *Mol. Cell*, **65**, 932–940.
 55. Yan, F., Huang, C., Wang, X., Tan, J., Cheng, S., Wan, M., Wang, Z., Wang, S., Luo, S., Li, A. *et al.* (2020) Threonine ADP-ribosylation of ubiquitin by a bacterial effector Family blocks host ubiquitination. *Mol. Cell*, **78**, 641–652.
 56. Leslie Pedrioli, D.M., Leutert, M., Bilan, V., Nowak, K., Gunasekera, K., Ferrari, E., Imhof, R., Malmström, L. and Hottiger, M.O. (2018) Comprehensive ADP-ribosylome analysis identifies tyrosine as an ADP-ribose acceptor site. *EMBO Rep.*, **19**, e45310.
 57. Altmeyer, M. and Hottiger, M.O. (2009) Poly(ADP-ribose) polymerase 1 at the crossroad of metabolic stress and inflammation in aging. *Aging*, **1**, 458–469.
 58. Vyas, S. and Chang, P. (2014) New PARP targets for cancer therapy. *Nat. Rev. Cancer*, **14**, 502–509.
 59. Laing, S., Unger, M., Koch-Nolte, F. and Haag, F. (2011) ADP-ribosylation of arginine. *Amino Acids*, **41**, 257–269.
 60. Altmeyer, M., Messner, S., Hassa, P.O., Fey, M. and Hottiger, M.O. (2009) Molecular mechanism of poly(ADP-ribosylation) by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. *Nucleic Acids Res.*, **37**, 3723–3738.
 61. Martello, R., Leutert, M., Jungmichel, S., Bilan, V., Larsen, S.C., Young, C., Hottiger, M.O. and Nielsen, M.L. (2016) Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue. *Nat. Commun.*, **7**, 12917.
 62. Larsen, S.C., Hendriks, I.A., Lyon, D., Jensen, L.J. and Nielsen, M.L. (2018) Systems-wide analysis of serine ADP-ribosylation reveals widespread occurrence and site-specific overlap with phosphorylation. *Cell Rep.*, **24**, 2493–2505.
 63. Buch-Larsen, S.C., Hendriks, I.A., Lodge, J.M., Rykær, M., Furtwängler, B., Shishkova, E., Westphall, M.S., Coon, J.J. and Nielsen, M.L. (2020) Mapping physiological ADP-ribosylation using activated ion electron transfer dissociation. *Cell Rep.*, **32**, 108176.
 64. Jungmichel, S., Rosenthal, F., Altmeyer, M., Lukas, J., Hottiger, M.O. and Nielsen, M.L. (2013) Proteome-wide identification of poly(ADP-Ribosylation) targets in different genotoxic stress responses. *Mol. Cell*, **52**, 272–285.
 65. Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L.M., Haugen-Strano, A., Swensen, J. and Miki, Y. (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science*, **266**, 120–122.
 66. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M. and Ding, W. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, **266**, 66–71.
 67. King, M.-C., Marks, J.H., Mandell, J.B. and New York Breast Cancer Study Group New York Breast Cancer Study Group (2003) Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, **302**, 643–646.
 68. Venkitaraman, A.R. (2014) Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. *Science*, **343**, 1470–1475.
 69. Li, M.L. and Greenberg, R.A. (2012) Links between genome integrity and BRCA1 tumor suppression. *Trends Biochem. Sci.*, **37**, 418–424.

70. Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K. *et al.* (2012) Global landscape of HIV-human protein complexes. *Nature*, **481**, 365–370.
71. Shah, P.S., Link, N., Jang, G.M., Sharp, P.P., Zhu, T., Swaney, D.L., Johnson, J.R., Von Dollen, J., Ramage, H.R., Satkamp, L. *et al.* (2018) Comparative flavivirus-host protein interaction mapping reveals mechanisms of Dengue and Zika virus pathogenesis. *Cell*, **175**, 1931–1945.
72. Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O’Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L. *et al.* (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, **583**, 459–468.
73. Pimkina, J., Humbey, O., Zilfou, J.T., Jarnik, M. and Murphy, M.E. (2009) ARF induces autophagy by virtue of interaction with bcl-xl. *J. Biol. Chem.*, **284**, 2803–2810.
74. Drost, R., Dhillon, K.K., van der Gulden, H., van der Heijden, I., Brandsma, I., Cruz, C., Chondronasiou, D., Castroviejo-Bermejo, M., Boon, U., Schut, E. *et al.* (2016) BRCA1185delAG tumors may acquire therapy resistance through expression of RING-less BRCA1. *J. Clin. Invest.*, **126**, 2903–2918.
75. Wang, Y., Kraiss, J.J., Bernhardt, A.J., Nicolas, E., Cai, K.Q., Harrell, M.I., Kim, H.H., George, E., Swisher, E.M., Simpkins, F. *et al.* (2016) RING domain-deficient BRCA1 promotes PARP inhibitor and platinum resistance. *J. Clin. Invest.*, **126**, 3145–3157.
76. Calses, P.C., Pham, V.C., Guarnaccia, A.D., Choi, M., Verschueren, E., Bakker, S.T., Pham, T.H., Hinkle, T., Liu, C., Chang, M.T. *et al.* (2023) TEAD proteins associate with DNA repair proteins to facilitate cellular recovery from DNA damage. *Mol. Cell. Proteomics*, **22**, 100496.
77. Sirbu, B.M., Couch, F.B., Feigler, J.T., Bhaskara, S., Hiebert, S.W. and Cortez, D. (2011) Analysis of protein dynamics at active, stalled, and collapsed replication forks. *Genes Dev.*, **25**, 1320–1327.
78. Alabert, C., Bukowski-Wills, J.-C., Lee, S.-B., Kustatscher, G., Nakamura, K., de Lima Alves, F., Menard, P., Mejlvang, J., Rappalber, J. and Groth, A. (2014) Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat. Cell Biol.*, **16**, 281–293.
79. Wang, C.I., Alekseyenko, A.A., LeRoy, G., Elia, A.E.H., Gorchakov, A.A., Britton, L.-M.P., Elledge, S.J., Kharchenko, P.V., Garcia, B.A. and Kuroda, M.I. (2013) Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat. Struct. Mol. Biol.*, **20**, 202–209.
80. Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R. and Fujii, H. (2013) Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci. Rep.*, **3**, 3171.
81. Cortez, D. (2017) Proteomic analyses of the eukaryotic replication machinery. *Methods Enzymol.*, **591**, 33–53.
82. Sirbu, B.M., McDonald, W.H., Dungrawala, H., Badu-Nkansah, A., Kavanaugh, G.M., Chen, Y., Tabb, D.L. and Cortez, D. (2013) Identification of proteins at active, stalled, and collapsed replication forks using isolation of proteins on nascent DNA (iPOND) coupled with mass spectrometry. *J. Biol. Chem.*, **288**, 31458–31467.
83. Aranda, S., Rutishauser, D. and Ernfors, P. (2014) Identification of a large protein network involved in epigenetic transmission in replicating DNA of embryonic stem cells. *Nucleic Acids Res.*, **42**, 6972–6986.
84. Dungrawala, H., Rose, K.L., Bhat, K.P., Mohni, K.N., Glick, G.G., Couch, F.B. and Cortez, D. (2015) The replication checkpoint prevents two types of fork collapse without regulating replisome stability. *Mol. Cell*, **59**, 998–1010.
85. Lecona, E., Rodriguez-Acebes, S., Specks, J., Lopez-Contreras, A.J., Ruppen, I., Murga, M., Muñoz, J., Mendez, J. and Fernandez-Capetillo, O. (2016) USP7 is a SUMO deubiquitinase essential for DNA replication. *Nat. Struct. Mol. Biol.*, **23**, 270–277.
86. Munden, A., Wright, M.T., Han, D., Tirgar, R., Plate, L. and Nordman, J.T. (2022) Identification of replication fork-associated proteins in *Drosophila* embryos and cultured cells using iPOND coupled to quantitative mass spectrometry. *Sci. Rep.*, **12**, 6903.
87. Roux, K.J., Kim, D.I., Raida, M. and Burke, B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, **196**, 801–810.
88. 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.
89. Hung, V., Zou, P., Rhee, H.-W., Udeshi, N.D., Cracan, V., Svinkina, T., Carr, S.A., Mootha, V.K. and Ting, A.Y. (2014) Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol. Cell*, **55**, 332–341.
90. Lobingier, B.T., Hüttenhain, R., Eichel, K., Miller, K.B., Ting, A.Y., von Zastrow, M. and Krogan, N.J. (2017) An approach to spatiotemporally resolve protein interaction networks in living cells. *Cell*, **169**, 350–360.
91. Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A., Kremer, M., Typas, D., Lammers, M., Mailand, N., Nussenzweig, A. *et al.* (2018) DNA repair network analysis reveals shieldin as a key regulator of NHEJ and PARP inhibitor sensitivity. *Cell*, **173**, 972–988.
92. Schopp, I.M., Amaya Ramirez, C.C., Debeljak, J., Kreibich, E., Skribbe, M., Wild, K. and Béthune, J. (2017) Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat. Commun.*, **8**, 15690.
93. Han, Y., Branon, T.C., Martell, J.D., Boassa, D., Shechner, D., Ellisman, M.H. and Ting, A. (2019) Directed evolution of split APEX2 peroxidase. *ACS Chem. Biol.*, **14**, 619–635.
94. Cho, K.F., Branon, T.C., Rajeev, S., Svinkina, T., Udeshi, N.D., Thoudam, T., Kwak, C., Rhee, H.-W., Lee, I.-K., Carr, S.A. *et al.* (2020) Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 12143–12154.
95. Cho, K.F., Branon, T.C., Udeshi, N.D., Myers, S.A., Carr, S.A. and Ting, A.Y. (2020) Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nat. Protoc.*, **15**, 3971–3999.
96. Abbasi, S. and Schild-Poulter, C. (2019) Mapping the ku interactome using proximity-dependent biotin identification in Human cells. *J. Proteome Res.*, **18**, 1064–1077.
97. Scott, W.A., Dhanji, E.Z., Dyakov, B.J.A., Dreseris, E.S., Asa, J.S., Grange, L.J., Mirceta, M., Pearson, C.E., Stewart, G.S., Gingras, A.-C. *et al.* (2021) ATRX proximal protein associations boast roles beyond histone deposition. *PLoS Genet.*, **17**, e1009909.
98. Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B. and Vitek, O. (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*, **30**, 2524–2526.
99. Sowa, M.E., Bennett, E.J., Gygi, S.P. and Harper, J.W. (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell*, **138**, 389–403.
100. Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K. *et al.* (2015) The BioPlex Network: a systematic exploration of the Human interactome. *Cell*, **162**, 425–440.
101. Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C. and Nesvizhskii, A.I. (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat. Methods*, **8**, 70–73.
102. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.C. and Choi, H. (2014) SAINTexpress: improvements and additional features in Significance Analysis of INteractome software. *J. Proteomics*, **100**, 37–43.
103. Verschueren, E., Von Dollen, J., Cimermancic, P., Gulbahce, N., Sali, A. and Krogan, N.J. (2015) Scoring large-scale affinity purification mass spectrometry datasets with MiST. *Curr. Protoc. Bioinformatics*, **49**, 8.19.1–8.19.16.
104. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C. *et al.* (2022) The reactome pathway knowledgebase 2022. *Nucleic Acids Res.*, **50**, D687–D692.
105. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. and Ishiguro-Watanabe, M. (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.*, **51**, D587–D592.
106. Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D.N., Hanspers, K., Miller, R.A., Digles, D., Lopes, E.N., Ehrhart, F. *et al.* (2021) WikiPathways: connecting communities. *Nucleic Acids Res.*, **49**, D613–D621.

107. Gene Ontology Consortium (2021) The Gene ontology resource: enriching a GOld mine. *Nucleic Acids Res.*, **49**, D325–D334.
108. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.
109. Goeman, J.J. and Bühlmann, P. (2007) Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, **23**, 980–987.
110. Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D. and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
111. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P. *et al.* (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.
112. Wu, G., Feng, X. and Stein, L. (2010) A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.*, **11**, R53.
113. Cowen, L., Ideker, T., Raphael, B.J. and Sharan, R. (2017) Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.*, **18**, 551–562.
114. Eckhardt, M., Zhang, W., Gross, A.M., Von Dollen, J., Johnson, J.R., Franks-Skiba, K.E., Swaney, D.L., Johnson, T.L., Jang, G.M., Shah, P.S. *et al.* (2018) Multiple routes to oncogenesis are promoted by the Human papillomavirus-host protein network. *Cancer Discov.*, **8**, 1474–1489.
115. Wang, R., Simoneau, C.R., Kulsuptrakul, J., Bouhaddou, M., Trivisano, K.A., Hayashi, J.M., Carlson-Stevermer, J., Zengel, J.R., Richards, C.M., Fozouni, P. *et al.* (2021) Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. *Cell*, **184**, 106–119.
116. Calabrese, A.N. and Radford, S.E. (2018) Mass spectrometry-enabled structural biology of membrane proteins. *Methods*, **147**, 187–205.
117. Schneider, M., Belsom, A. and Rappsilber, J. (2018) Protein tertiary structure by crosslinking/mass spectrometry. *Trends Biochem. Sci.*, **43**, 157–169.
118. Chavez, J.D. and Bruce, J.E. (2019) Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Curr. Opin. Chem. Biol.*, **48**, 8–18.
119. Iacobucci, C., Götze, M. and Sinz, A. (2020) Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr. Opin. Biotechnol.*, **63**, 48–53.
120. Piotrowski, C. and Sinz, A. (2018) Structural investigation of proteins and protein complexes by chemical cross-linking/mass spectrometry. *Adv. Exp. Med. Biol.*, **1105**, 101–121.
121. 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.
122. Lin, D.H. and Hoelz, A. (2019) The structure of the nuclear pore complex (An Update). *Annu. Rev. Biochem.*, **88**, 725–783.
123. Fernandez-Martinez, J., Kim, S.J., Shi, Y., Upla, P., Pellarin, R., Gagnon, M., Chemmama, I.E., Wang, J., Nudelman, I., Zhang, W. *et al.* (2016) Structure and function of the nuclear pore complex cytoplasmic mRNA export platform. *Cell*, **167**, 1215–1228.
124. Tüting, C., Iacobucci, C., Ihling, C.H., Kastrius, P.L. and Sinz, A. (2020) Structural analysis of 70S ribosomes by cross-linking/mass spectrometry reveals conformational plasticity. *Sci. Rep.*, **10**, 12618.
125. Klykov, O., van der Zwaan, C., Heck, A.J.R., Meijer, A.B. and Scheltema, R.A. (2020) Missing regions within the molecular architecture of human fibrin clots structurally resolved by XL-MS and integrative structural modeling. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 1976–1987.
126. Qin, H. and Wang, Y. (2009) Exploring DNA-binding proteins with in vivo chemical cross-linking and mass spectrometry. *J. Proteome Res.*, **8**, 1983–1991.
127. Li, C.G., Mahon, C., Sweeney, N.M., Verschuere, E., Kantamani, V., Li, D., Hennings, J.K., Marciano, D.P., Diebold, I., Abu-Halawa, O. *et al.* (2019) PPAR γ interaction with UBR5/ATMIN promotes DNA repair to maintain endothelial homeostasis. *Cell Rep.*, **26**, 1333–1343.
128. Holzer, S., Degliesposti, G., Kilkenny, M.L., Maslen, S.L., Matak-Vinkovic, D., Skehel, M. and Pellegrini, L. (2017) Crystal structure of the N-terminal domain of human Timeless and its interaction with Tipin. *Nucleic Acids Res.*, **45**, 5555–5563.
129. Fasci, D., van Ingen, H., Scheltema, R.A. and Heck, A.J.R. (2018) Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei. *Mol. Cell. Proteomics*, **17**, 2018–2033.
130. Hvidt, A. and Nielsen, S.O. (1966) Hydrogen exchange in proteins. *Adv. Protein Chem.*, **21**, 287–386.
131. Englander, S.W. and Kallenbach, N.R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.*, **16**, 521–655.
132. Konermann, L., Pan, J. and Liu, Y.-H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.*, **40**, 1224–1234.
133. Masson, G.R., Burke, J.E., Ahn, N.G., Anand, G.S., Borchers, C., Brier, S., Bou-Assaf, G.M., Engen, J.R., Englander, S.W., Faber, J. *et al.* (2019) Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods*, **16**, 595–602.
134. Harrison, R.A. and Engen, J.R. (2016) Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry. *Curr. Opin. Struct. Biol.*, **41**, 187–193.
135. Chalmers, M.J., Busby, S.A., Pascal, B.D., West, G.M. and Griffin, P.R. (2011) Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev. Proteomics*, **8**, 43–59.
136. Englander, J.J., Del Mar, C., Li, W., Englander, S.W., Kim, J.S., Stranz, D.D., Hamuro, Y. and Woods, V.L. Jr (2003) Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 7057–7062.
137. Englander, S.W. and Mayne, L. (2014) The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 15873–15880.
138. Balasubramanian, D. and Komives, E.A. (2013) Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta*, **1834**, 1202–1209.
139. Vadas, O., Jenkins, M.L., Dornan, G.L. and Burke, J.E. (2017) Using hydrogen-deuterium exchange mass spectrometry to examine protein-membrane interactions. *Methods Enzymol.*, **583**, 143–172.
140. Rand, K.D., Adams, C.M., Zubarev, R.A. and Jørgensen, T.J.D. (2008) Electron capture dissociation proceeds with a low degree of intramolecular migration of peptide amide hydrogens. *J. Am. Chem. Soc.*, **130**, 1341–1349.
141. Zehl, M., Rand, K.D., Jensen, O.N. and Jørgensen, T.J.D. (2008) Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution. *J. Am. Chem. Soc.*, **130**, 17453–17459.
142. Jørgensen, T.J.D., Gårdsvoll, H., Ploug, M. and Roepstorff, P. (2005) Intramolecular migration of amide hydrogens in protonated peptides upon collisional activation. *J. Am. Chem. Soc.*, **127**, 2785–2793.
143. Ferguson, P.L., Pan, J., Wilson, D.J., Dempsey, B., Lajoie, G., Shilton, B. and Konermann, L. (2007) Hydrogen/deuterium scrambling during quadrupole time-of-flight MS/MS analysis of a zinc-binding protein domain. *Anal. Chem.*, **79**, 153–160.
144. Sterling, H.J. and Williams, E.R. (2010) Real-time hydrogen/deuterium exchange kinetics via supercharged electrospray ionization tandem mass spectrometry. *Anal. Chem.*, **82**, 9050–9057.
145. Resetca, D. and Wilson, D.J. (2013) Characterizing rapid, activity-linked conformational transitions in proteins via sub-second hydrogen deuterium exchange mass spectrometry. *FEBS J.*, **280**, 5616–5625.
146. Pan, J., Wilson, D.J. and Konermann, L. (2005) Pulsed hydrogen exchange and electrospray charge-state distribution as complementary probes of protein structure in kinetic experiments: implications for ubiquitin folding. *Biochemistry*, **44**, 8627–8633.
147. Resetca, D., Haftchenary, S., Gunning, P.T. and Wilson, D.J. (2014) Changes in signal transducer and activator of transcription 3 (STAT3) dynamics induced by complexation with pharmacological

- inhibitors of src homology 2 (SH2) domain dimerization. *J. Biol. Chem.*, **289**, 32538–32547.
148. Zhu, S., Shala, A., Bezginov, A., Sljoka, A., Audette, G. and Wilson, D.J. (2015) Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble. *PLoS One*, **10**, e0120416.
 149. Ahmad, F., Patterson, A., Deveryshetty, J., Mattice, J.R., Pokhrel, N., Bothner, B. and Antony, E. (2021) Hydrogen-deuterium exchange reveals a dynamic DNA-binding map of replication protein A. *Nucleic Acids Res.*, **49**, 1455–1469.
 150. Dawicki-McKenna, J.M., Langelier, M.-F., DeNizio, J.E., Riccio, A.A., Cao, C.D., Karch, K.R., McCauley, M., Steffen, J.D., Black, B.E. and Pascal, J.M. (2015) PARP-1 activation requires local unfolding of an autoinhibitory domain. *Mol. Cell*, **60**, 755–768.
 151. Sharon, M., Witt, S., Glasmacher, E., Baumeister, W. and Robinson, C.V. (2007) Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome. *J. Biol. Chem.*, **282**, 18448–18457.
 152. Bakhtiari, M. and Konermann, L. (2019) Protein ions generated by native electrospray ionization: comparison of gas phase, solution, and crystal structures. *J. Phys. Chem. B*, **123**, 1784–1796.
 153. Santambrogio, C., Natalello, A., Brocca, S., Ponzini, E. and Grandori, R. (2019) Conformational characterization and classification of intrinsically disordered proteins by native mass spectrometry and charge-state distribution analysis. *Proteomics*, **19**, e1800060.
 154. Kaltashov, I.A. and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal. Chem.*, **77**, 5370–5379.
 155. Testa, L., Brocca, S. and Grandori, R. (2011) Charge-surface correlation in electrospray ionization of folded and unfolded proteins. *Anal. Chem.*, **83**, 6459–6463.
 156. Beveridge, R., Covill, S., Pacholarz, K.J., Kalapothakis, J.M.D., MacPhee, C.E. and Barran, P.E. (2014) A mass-spectrometry-based framework to define the extent of disorder in proteins. *Anal. Chem.*, **86**, 10979–10991.
 157. Wang, G., Chaihu, L., Tian, M., Shao, X., Dai, R., de Jong, R.N., Ugarlar, D., Gros, P. and Heck, A.J.R. (2020) Releasing nonperipheral subunits from protein complexes in the gas phase. *Anal. Chem.*, **92**, 15799–15805.
 158. Stiving, A.Q., VanAernum, Z.L., Busch, F., Harvey, S.R., Sarni, S.H. and Wysocki, V.H. (2019) Surface-induced dissociation: an effective method for characterization of protein quaternary structure. *Anal. Chem.*, **91**, 190–209.
 159. Zhou, M. and Wysocki, V.H. (2014) Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Acc. Chem. Res.*, **47**, 1010–1018.
 160. Monti, M.C., Cohen, S.X., Fish, A., Winterwerp, H.H.K., Barendregt, A., Friedhoff, P., Perrakis, A., Heck, A.J.R., Sixma, T.K., van den Heuvel, R.H.H. *et al.* (2011) Native mass spectrometry provides direct evidence for DNA mismatch-induced regulation of asymmetric nucleotide binding in mismatch repair protein MutS. *Nucleic Acids Res.*, **39**, 8052–8064.
 161. Papp-Kádár, V., Balázs, Z., Vékey, K., Ozohanics, O. and Vértessy, B.G. (2019) Mass spectrometry-based analysis of macromolecular complexes of Staphylococcus aureus uracil-DNA glycosylase and its inhibitor reveals specific variations due to naturally occurring mutations. *FEBS Open Bio*, **9**, 420–427.
 162. van de Waterbeemd, M., Fort, K.L., Boll, D., Reinhardt-Szyba, M., Routh, A., Makarov, A. and Heck, A.J.R. (2017) High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat. Methods*, **14**, 283–286.
 163. Tamara, S., den Boer, M.A. and Heck, A.J.R. (2022) High-resolution native mass spectrometry. *Chem. Rev.*, **122**, 7269–7326.
 164. Konermann, L., Ahadi, E., Rodriguez, A.D. and Vahidi, S. (2013) Unraveling the mechanism of electrospray ionization. *Anal. Chem.*, **85**, 2–9.
 165. Garza, K.Y., Feider, C.L., Klein, D.R., Rosenberg, J.A., Brodbelt, J.S. and Eberlin, L.S. (2018) Desorption electrospray ionization mass spectrometry imaging of proteins directly from biological tissue sections. *Anal. Chem.*, **90**, 7785–7789.
 166. Ifa, D.R., Wu, C., Ouyang, Z. and Cooks, R.G. (2010) Desorption electrospray ionization and other ambient ionization methods: current progress and preview. *Analyst*, **135**, 669–681.
 167. Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P. *et al.* (2018) Integrative structure and functional anatomy of a nuclear pore complex. *Nature*, **555**, 475–482.
 168. Sali, A. (2021) From integrative structural biology to cell biology. *J. Biol. Chem.*, **296**, 100743.
 169. Echeverria, I., Braberg, H., Krogan, N.J. and Sali, A. (2022) Integrative structure determination of histones H3 and H4 using genetic interactions. *FEBS J.*, <https://doi.org/10.1111/febs.16435>.
 170. Gutierrez, C., Chemmama, I.E., Mao, H., Yu, C., Echeverria, I., Block, S.A., Rychnovsky, S.D., Zheng, N., Sali, A. and Huang, L. (2020) Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 4088–4098.
 171. Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M.P. and Chait, B.T. (2014) Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics*, **13**, 2927–2943.
 172. Robinson, P.J., Trnka, M.J., Pellarin, R., Greenberg, C.H., Bushnell, D.A., Davis, R., Burlingame, A.L., Sali, A. and Kornberg, R.D. (2015) Molecular architecture of the yeast Mediator complex. *Elife*, **4**, e08719.
 173. Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebbersold, R., Sali, A. and Baumeister, W. (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 1380–1387.
 174. Luo, J., Cimermancic, P., Viswanath, S., Ebmeier, C.C., Kim, B., Dehecq, M., Raman, V., Greenberg, C.H., Pellarin, R., Sali, A. *et al.* (2015) Architecture of the Human and yeast general transcription and DNA repair factor TFIIH. *Mol. Cell*, **59**, 794–806.
 175. Molnar, K.S., Bonomi, M., Pellarin, R., Clinthorne, G.D., Gonzalez, G., Goldberg, S.D., Goulian, M., Sali, A. and DeGrado, W.F. (2014) Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure*, **22**, 1239–1251.
 176. Zhou, C.Y., Stoddard, C.I., Johnston, J.B., Trnka, M.J., Echeverria, I., Palovcak, E., Sali, A., Burlingame, A.L., Cheng, Y. and Narlikar, G.J. (2017) Regulation of Rvb1/Rvb2 by a domain within the INO80 chromatin remodeling complex implicates the yeast rvbs as protein assembly chaperones. *Cell Rep.*, **19**, 2033–2044.
 177. Thomas, M.C. and Chiang, C.-M. (2006) The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.*, **41**, 105–178.
 178. Compe, E. and Egly, J.-M. (2012) TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.*, **13**, 343–354.
 179. Manuguerra, M., Saletta, F., Karagas, M.R., Berwick, M., Veglia, F., Vineis, P. and Matullo, G. (2006) XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. *Am. J. Epidemiol.*, **164**, 297–302.
 180. Wang, F., Chang, D., Hu, F.-L., Sui, H., Han, B., Li, D.-D. and Zhao, Y.-S. (2008) DNA repair gene XPD polymorphisms and cancer risk: a meta-analysis based on 56 case-control studies. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 507–517.
 181. Chang, W.H. and Kornberg, R.D. (2000) Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell*, **102**, 609–613.
 182. Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D. and Egly, J.M. (2000) Molecular structure of human TFIIH. *Cell*, **102**, 599–607.
 183. Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J. and Kornberg, R.D. (2012) Subunit architecture of general transcription factor TFIIH. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 1949–1954.
 184. He, Y., Fang, J., Taatjes, D.J. and Nogales, E. (2013) Structural visualization of key steps in human transcription initiation. *Nature*, **495**, 481–486.
 185. Labib, M. and Kelley, S.O. (2020) Single-cell analysis targeting the proteome. *Nat. Rev. Chem.*, **4**, 143–158.
 186. Huffman, R.G., Leduc, A., Wichmann, C., Di Gioia, M., Borriello, F., Specht, H., Derks, J., Khan, S., Khoury, L., Emmott, E. *et al.* (2023) Prioritized mass spectrometry increases the depth, sensitivity and

- data completeness of single-cell proteomics. *Nat. Methods*, **20**, 714–722.
187. Eldrid,C. and Thalassinou,K. (2020) Developments in tandem ion mobility mass spectrometry. *Biochem. Soc. Trans.*, **48**, 2457–2466.
188. Ewing,M.A., Glover,M.S. and Clemmer,D.E. (2016) Hybrid ion mobility and mass spectrometry as a separation tool. *J. Chromatogr. A*, **1439**, 3–25.
189. Konijnenberg,A., Butterer,A. and Sobott,F. (2013) Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta*, **1834**, 1239–1256.
190. Wongkongkathep,P., Han,J.Y., Choi,T.S., Yin,S., Kim,H.I. and Loo,J.A. (2018) Native top-down mass spectrometry and ion mobility MS for characterizing the cobalt and manganese metal binding of α -synuclein protein. *J. Am. Soc. Mass Spectrom.*, **29**, 1870–1880.
191. 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.
192. Chantada-Vázquez,M.D.P., García Vence,M., Serna,A., Núñez,C. and Bravo,S.B. (2021) SWATH-MS protocols in Human diseases. *Methods Mol. Biol.*, **2259**, 105–141.
193. Meier,F., Brunner,A.-D., Frank,M., Ha,A., Bludau,I., Voytik,E., Kaspar-Schoenefeld,S., Lubeck,M., Raether,O., Bache,N. *et al.* (2020) diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods*, **17**, 1229–1236.
194. Zhang,F., Ge,W., Ruan,G., Cai,X. and Guo,T. (2020) Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics*, **20**, e1900276.