METHODS

Cell Culturing and Treatments

HeLa S3 cells (ATCC CCL-2.2) were grown in Joklik's modified minimal essential medium. HeLa cells were grown in suspension, whereas B16F10 mouse melanoma cells (ATCC CRL-6475) and H1299 small lung carcinoma cells (ATCC CRL-5803, in DMEM) were grown adherently. Media were supplemented with 10% newborn calf serum and 1% penicillin and

streptomycin. Cells were maintained in 5% CO_2 at 37°C, harvested (after adding 0.05% trypsin and EDTA for B16F10 and H1299 cells) by centrifugation at 200 × *g* for 5 min., and washed twice with PBS.

For experiments using HeLa with intrinsic DNA damage, cells were treated for 1 or 5 h using 1 µM etoposide (4D platform), or for 1 h with 25 or 100 µM etoposide for targeted experiments (3D platform, **Supplementary Fig. 10 and Supplementary Table 6**)[31]. Repair of DNA damage was monitored after placing treated (25 µM) cells into fresh media for 24 h before harvesting. B16F10 cells were treated with 10 µM etoposide for 5 h at 20-30% confluence, and allowed to grow in normal media. Similarly, H1299 cells were treated with 25 nM camptothecin for 24 h. Over several days, these treatments induced stress-associated, accelerated senescence as monitored by[32] a flattened and enlarged cell morphology, expression of senescence-associated β-galactosidase (SA-β-gal), formation of senescence-associated heterochromatic foci (SAHFs) and upregulation of p53 (in B16F10 cells only). Approximately 2×10^7 senescent cells were harvested, lysed, and subjected to the 3D platform.

Preparation of HeLa S3 Cytosolic, Nuclear and Whole Cell Extracts

For large scale mapping using the 4D platform, the HeLa cytosolic and nuclear extracts were prepared through a protocol by Trinkle-Mulcahy *et al.*[33]. After isolation, the pelleted fraction containing nuclei was redissolved using 4% SDS (50 mM Tris, pH 7.5, with protease, phosphatase inhibitors and sodium butyrate). Both cytosolic and nuclear fractions were disrupted using a sonication probe. The nuclear fractions were further homogenized using QIAShredder homogenizer spin columns to reduce viscosity (QIAGEN). All fractions were centrifuged at 14,000 × *g* for 10 min. at 4ºC. Protein concentrations were determined by BCA and stored at -80ºC.

For the 3D platform, whole cell extracts were resuspended in 5 mL of lysis buffer (4% SDS, 100 mM Tris-HCl pH 7.5, 10 mM DTT with protease, phosphatase inhibitors and sodium butyrate). The mixture was vortexed for 5 min. and boiled for 10 min. Immediately after boiling, the samples were alkylated in the dark with 100 mM iodoacetamide for 20 min.

Western Blots, Imaging and Microscopy

Western Blots. Antibodies utilized were Histone H2A.X-pSer139 (Cell Signaling Technology; 2577S), GAPDH (Santa Cruz; sc-47724) and HRP-conjugated secondary antibodies. Chemiluminescence was detected using ChemiDoc XRS+ (Bio-Rad Laboratories) and band densities were calculated using Image Lab software (Bio-Rad Laboratories).

Gamma-H2A.X Imaging by Immunofluorescence. Cells were fixed in 0.1% glutaraldehyde and 3% formaldehyde made fresh from paraformaldehyde followed by permeabilization in 0.5% Triton X-100. Cells were subsequently incubated for 1 h at RT with 3% BSA and then with the primary antibody against Histone H2A.X-pSer139 (1:400). Alexa Fluor Conjugates (Molecular Probes) were used as the secondary antibodies (1:1000). Images were obtained using an inverted Zeiss Axio Observer.Z1 confocal microscope.

β-Gal Staining. Cells were fixed for 5 min. in 2% formaldehyde/0.2% glutaraldehyde in PBS, washed, and incubated at 37°C (no CO_2) overnight with X-Gal (5-bromo-4-chloro-3-indolyl *β*-D-glactosidase) staining solution (1 mg of X-Gal in 40 mM citric acid/sodium phosphate pH 6.0, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 150 mM NaCl, 2 mM $MgCl_2$).

DAPI Staining. Cells were fixed in 4% paraformaldehyde for 30 min., washed, and incubated with DAPI (4',6'-Diamidino-2-Phenylindole) solution (10 µg/mL) for 10 min. The images of DAPI stained DNA were obtained using the Zeiss microscope noted above.

Sample Handling and Multidimensional Protein Fractionation

Fractionation using 4 dimensional (4D) fractionation (2D-LE and LC-MS). HeLa proteins (0.5-2 mg) were reduced, alkylated, precipitated with cold acetone and resuspended in 3.2 mL sIEF buffer (8 M urea, 2 M thiourea, 50 mM DTT, 1% w/v Biolyte 3/10 carrier ampholytes from Bio-Rad Laboratories). The sample was focused using a custom designed eight-chamber sIEF system as previously described[34]. After complete focusing (~1.5 h at 2 W), the liquid fractions (400 µL) were collected and combined with their respective chamber rinse

solution (100 μL of 1% SDS). Adjacent sIEF fractions (including anode and cathode) were pooled, resulting in about 5 fractions that were precipitated using cold acetone. The precipitated proteins were resuspended in ~50 μL of Laemmli loading buffer[35]. These fractions were then fractionated in parallel using a custom 10 channel mGELFrEE device[12]. Tube gels were cast to 12% T (1 cm length) for the resolving and 4% T for the stacking gels (300 μL volume). Application of 240 V for ~1 h resulted in eight or nine GELFrEE fractions (150 μL) per IEF fraction after elution of the dye front. After complete electroelution, the 2D-LE fractions underwent SDS removal using chloroform/methanol/water precipitation as described previously[36]. Prior to nanocapillary RPLC injection, fractions were resuspended by pipetting vigorously with 15–40 μL solvent A (5% acetonitrile, 0.2% formic acid).

Fractionation using 3 dimensional (3D) fractionation. For experiments using the 3D platform, only GELFrEE coupled to nanocapillary-LC-MS was used for sample fractionation. Whole cell lysates, mitochondrial membrane preparations, or extracts targeting modified proteins were resuspended as described above and fractionated using a single channel GELFrEE device[13]. Nanocapillary-LC-MS conditions for large scale analyses were as described below.

Nanocapillary RPLC-MS. In either the 4D or 3D platform, resuspended fractions were injected (10 μL) onto a trap column (150 μm i.d. × 2 cm) using an autosampler (Eksigent). The nanobore analytical column (75 μm × 10 cm) containing an integral fritted nanospray emitter (PicoFrit, New Objective) was coupled to the trap in a vented column tee setup. Both the analytical and trap columns contain polymeric reversed-phase (PLRP-S, Phenomenex) media (5 μm, 1,000 Å pore size). The Eksigent 1D Plus nano-HPLC system was operated at a flow rate of ~2 μL/min. for 10 min. for loading onto the trap. The proteins were eluted into the mass spectrometer using a flow rate of 300 nL/min. with the following gradient: 5% B (95% acetonitrile + 0.2% formic acid) at 0 min.; 20% B at 5 min.; 55% B at 50 min.; 85% B at 55 min.; 5% B at 65 min.; 5% B at 75 min.

For proteins fractionated from HeLa S3 and B16F10 cells, the nanocapillary RPLC column was coupled online to a 12 Tesla LTQ FT Ultra (Thermo Fisher Scientific) fitted with a digitally controlled nanospray ionization source (PicoView DPV-550, New Objective). For masses up to 25 kDa (as determined from GELFrEE fractions) MS1 data were collected using

the FT-ICR (8 microscans, 170,000 resolving power at *m/z* 400) with a *m/z* range of 500–1,800 and a target value of 1 million charges. For masses either >25 kDa or >50 kDa, MS1 data were obtained using the ion trap (IT) at 20 or 50 microscans, respectively. Data from H1299 cells were obtained on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) using 2-4 microscans, 108,000 or 216,000 resolving power at *m/z* 400, and a target value of 1 million charges.

MS Data Acquisition for Targeted Monitoring of Intact Isoforms/Species. For 3D experiments on targets up to 25 kDa, MS1 data were collected with FT-ICR parameters as described above. Data-dependent "zoom mapping" was performed using a top 3 (no fragmentation) acquisition strategy with 60 *m/z* isolation window, 3 microscans at 85,000 resolving power (target value of 2 million charges with 60 *m/z* isolation at MS2 zero collision energy, or SIM mode). Dynamic exclusion was enabled with a repeat count of 2, an exclusion duration of 5,000 s, and a repeat duration of 240 s. "Mass mode" was enabled in the Xcalibur software to ensure that each zoom map scan detected a different protein species.

Preparation of Mitochondrial Membrane Proteins for Top Down Mass Spectrometry

A HeLa S3 cell pellet consisting of ~10^9 cells was resuspended in ~16 mL STM buffer (250 mM sucrose, 50 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 10 mM sodium butyrate, 1 mM DTT, 1% protease and phosphatase inhibitors (Sigma-Aldrich)). Cells were lysed using a glass Dounce homogenizer. The lysate was centrifuged at 800 × *g* for 15 min. to remove nuclei and cellular debris. Mitochondria membrane isolation was performed as described previously[37]. Briefly, the cell lysate was centrifuged at 6,000 × *g* for 15 min. The pellet was washed with STM buffer and the centrifugation step repeated. Mitochondria were resuspended in 2 mL lysis buffer (10 mM HEPES, pH 7.9, 10 mM sodium butyrate, 1 mM DTT, 1% protease and phosphatase inhibitors) and stirred at 4°C for 30 min. prior to sonication. The suspension was centrifuged at 9,000 × *g* for 30 min. The pellet was resuspended in 0.5 mL extraction buffer (20 mM Tris, pH 7.8, 0.4 M NaCl, 15% glycerol, 5% SDS, 10 mM sodium butyrate, 1 mM DTT, 1% protease and phosphatase inhibitors) and vortexed for 30 min. The sample was centrifuged at 9,000 × *g* for 30 min. and the supernatant was collected, aliquoted, flash frozen, and stored at -80°C until use.

GELFrEE Separation and LC-MS of Integral Membrane Proteins. Mitochondrial membrane proteins (400 µg as determined by BCA) were acetone precipitated, resuspended in 100 µL loading buffer, reduced with 20 mM DTT, and alkylated with 100 mM iodoacetamide. After a GELFrEE separation as reported previously[13], SDS was removed[36] and the fractions resuspended in 30 µL of fresh 60% formic acid. Each 10 µL fraction was injected onto a PLRP analytical column (75 µm × 10 cm), heated to 45°C using a nanocapillary column heater (New Objective, Inc.). A solvent system developed for integral membrane proteins was used in nanocapillary format here: A: 60% formic acid in water, B: 100% isopropanol. The gradient used was 0% B at 0 min.; 25% B at 5 min.; 60% B at 50 min.; 95% B at 56 min. Ten GELFrEE samples were also analyzed by the standard acetonitrile solvent and gradient system described above.

Transmembrane Domains. TMHMM v.2.0. (http://www.cbs.dtu.dk/services/TMHMM) was used for the transmembrane domain prediction of identified proteins. The SwissProt accession number for each identification was entered into the server to determine the number and location of all transmembrane domains. Differences between the database sequence and that identified by top down MS, were accounted for by interrogating the identified sequence with TMHMM.

Protein Identification and Characterization

MS fragmentation data were acquired in three different modes, depending on protein mass range. Below 17 kDa, data-dependent CID was used (FT/FT), whereas source induced dissociation (SID) was used for masses between 17-25 kDa (FT/FT) and above 25 kDa (IT/FT). Based on preliminary analyses, SID of 15 V was optimal for ion trap scans for the dissociation of weakly bound non-covalent adducts, while the 75 V SID for fragmentation was standardized as described[15]. For data-dependent fragmentation (top 2 MS/MS, 15-25 *m/z* isolation window), dynamic exclusion was enabled with a repeat count of 2, an exclusion duration of 5,000 s, and a repeat duration of 240 s. Both CID and SID were collected using 8 microscans and 85,000 resolving power (at *m/z* 400) with a target value of 2 million charges. Data for H1299 cells

obtained on the Orbitrap Elite were acquired using either HCD, CID or ETD during top 3 data dependent MS/MS fragmentation using mass mode, 4 microscans, 108,000 resolving power, and a target value of 1 million charges (*cf.* **Fig. 4i** and **4j**, **Supplementary Fig. 12**, **Supplementary Fig. 13**, and **Supplementary Table 7**).

Software and Data Analysis

Much of the software for intact mass determination (KDECON), generation of visual outputs (PROTEOME DISPLAY), and species differentiation (PTMCRAWLER) has been published recently[17]. Briefly, KDECON provides average mass information for charge state resolved distributions using the ion trap. PTMCRAWLER traverses a list of intact monoisotopic masses to find mass differences corresponding to PTMs such as methylations, acetylations, and phosphorylations (**Fig. 2b**). PROTEOME DISPLAY was devised for visualization of nanocapillary-LC-MS/MS data, allowing graphical viewing of 4-dimensional proteome runs and generation of plots for specific PTMs (**Fig. 2**).

The RAW files collected were first processed with an algorithm called CRAWLER to assign masses. Using a version of this program, scans collected by FTMS were deisotoped using the Xtract or THRASH algorithm[38]. Scans collected in the ion trap were deconvoluted using the KDECON algorithm (minimum intensity cutoff: 1,000, mass range: 10-70 kDa). Xtract or THRASH processing generated monoisotopic neutral masses, while KDECON processing provided average neutral masses. Data-dependent MS2 scans were summed within a retention time tolerance of 1 min. and a precursor tolerance of 0.05 *m/z* whereas SID scans were summed in 0.3 min. In both cases, multiplexed fragmentation was considered. Fragmentation data were filtered by selecting the top 3 most intense neutral fragment masses within a 100 Da window below 2,000 Da, and the top 5 above 2,000 Da. One or more precursor masses and one or more fragment masses resulting from each summed unit were grouped as ProSightPC experiment, and written to a ProSight Upload Format (PUF) file. If a precursor mass could not be determined, the experiment was written out to a separate file with a placeholder value, for separate analysis.

The PUF files output by CRAWLER were searched against a human proteome database using a custom implementation of ProSightPC 2.0 with iterative search logic[39-41] on a 168-node Rocks[42] cluster. Four types of analyses were run, depending on the type of data: FT/FT CID,

FT/FT SID, IT/FT SID or SID data where the precursor could not be determined ('No-Hi-SID'). The iterative search trees were designed to take advantage of high mass accuracy, while retaining the option to run less specific searches if a result of sufficient quality could not be obtained by more specific searches. All searches used 10 part-per-million tolerance for the fragment ions, all of which were obtained at high resolving power. For each search, the top 10 hits were returned; if the top hit had an E-value $\leq 1\times10^{-2}$, the analysis moved on to the next experiment, while otherwise, the next search in the tree was run. All searches were in absolute mass mode. The FT-FT-CID tree consisted of searches at 200, 2,000 Da, and "entire database" precursor tolerances; the FT-FT-SID at 2.3, 2,000, 20,000, and entire database; and IT-FT-SID trees used 2,000, 20,000 Da, and "entire database" precursor tolerances; the No-Hi-SID tree just searched against the entire database. Searches were against two different human proteome databases built against UniProt Release 2011_04, encompassing known alternative splices, modifications, peptide cleavage events, potential initial methionine cleavage and N-terminal acetylation. A complex database was created encompassing combinations of annotated alternative splice and peptide cleavage events to generate 54,190 base sequences. A maximum of 2^{13} protein forms for each base sequence make ~8,450,000 theoretical protein species. This database was used for all searches where the precursor tolerance was less than 2,000 Da. All other searches used a simplified database consisting of the same 54,190 base sequences, modified with N-terminal acetylation and initial methionine cleavage (where applicable) creating a total of ~160,000 forms.

Data were run against both forward and scrambled databases (see FDR estimation below) separately with identical search parameters. Upon completion, all search results were loaded into a ProSight data repository and a report was produced, returning the top hits for each experiment. The hit with the best *q-value* was then chosen as the exemplar for each gene product cluster. If two member hits had the same *q-value*, the member with the lowest absolute mass difference to the theoretical hit was chosen, and if this still produced a set with more than one member, the form with the "most-terminal" PTMs (*i.e.*, closest to the N- or C-terminus) was chosen. For a desired FDR cutoff (*e.g.*, 5%), a list of accession numbers and species is produced.

Estimation of FDR for Top Down Proteomics. The Poisson-based model as published in

2001[43] had been previously modified with a Bonferroni correction that enables probability-based scoring (*i.e.*, use of the E-value noted above) for searches done on a database created by Shotgun Annotation[16]. To validate the process used in high throughput operations, an FDR analysis was carried out to correct for multiple hypothesis testing using the method of Benjamini and Hochberg[44] as applied by Storey[45]. For each Poisson-based p-value, a corresponding Bayesian posterior p-value is calculated, termed the *q-value*, which is a measure of the FDR for that particular identification event (also called an instantaneous FDR).

To calculate *q-values*, a separate decoy database of scrambled sequences was created equal in size to the forward (real) database[46]. Searches were done separately on both the forward and decoy (scrambled) databases using all data in a set of 4D or 3D proteome runs. A histogram was created for the decoy database results using the log p_{id}-value (where p_{id} is the Poisson probability of an incorrect protein identification) of these false identifications (**Supplementary Fig. 14**). These data were modeled against a Gamma distribution and fit with a shape (k) of 10.26 ± 0.04 and rate (θ) of 2.25 ± 0.01. The distribution of scrambled hits is taken as an empirical estimate of the distribution of scores under the null hypothesis that the match was due to chance. Thus, the area under the scrambled score distribution to the right of the observed forward score is the probability of getting this good a forward score, or better, by chance (abbreviated hereafter p_{Γ}). From here, all data are rank-ordered by their corresponding p_{Γ}-values and q-values are calculated as in Storey[45]. The final results were generated using a *q-value* cut off of 0.05, thus achieving a protein level FDR of 5%. For comparison, bottom up studies typically use a 1% FDR cutoff at the peptide level (so-called PSM level, for peptide spectrum match), which typically rolls up into a 5-8% FDR at the protein level[47]. Extensive comparison with other FDR estimation techniques (including generating ROC curves for reversed decoy databases of concatenated reversed sequences) showed the *q-value* approach to be 10-25% more stringent in terms of number of identified proteins and species.

31 Soubeyrand, S., Pope, L. & Hache, R. J. G. Topoisomerase II alpha-dependent induction of a persistent DNA damage response in response to transient etoposide exposure. *Mol. Oncol.* **4**, 38-51 (2010).

32 Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). *Cell* **88**, 593-602 (1997).

33 Trinkle-Mulcahy, L. *et al.* Identifying Specific Protein Interaction Partners Using Quantitative Mass Spectrometry and Bead Proteomes. *J. Cell Biol.* **183**, 223-239 (2008).
34 Tran, J. C. & Doucette, A. A. Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool. *J. Proteome Res.* **7**, 1761-1766 (2008).
35 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of Bacteriopage T4. *Nature* **227**, 680-685 (1970).
36 Wessel, D. & Flugge, U. I. A Method for the Quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141-143 (1984).
37 Cox, B. & Emili, A. Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nat. Protoc.* **1**, 1872-1878 (2006).
38 Horn, D. M., Zubarev, R. A. & McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **11**, 320-332 (2000).
39 Boyne, M. T. *et al.* Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J. Proteome Res.* **8**, 374-379 (2009).
40 LeDuc, R. D. *et al.* ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.* **32**, W340-W345 (2004).
41 Zamdborg, L. *et al.* ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **35**, W701-W706 (2007).
42 Philip M. Papadopoulos, Mason J. Katz, Greg Bruno. NPACI Rocks: Tools and techniques for easily deploying manageable Linux clusters. *Third IEEE International Conference on Cluster Computing,* pp. 258, http://www.computer.org/portal/web/csdl/doi/10.1109/CLUSTR.2001.959986 (2001).
43 Meng, F. Y. *et al.* Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat. Biotechnol.* **19**, 952-957 (2001).
44 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289-300 (1995).
45 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440-9445 (2003).
46 Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207-214 (2007).
47 Reiter, L. *et al.* Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **8**, 2405-2417 (2009).