Low cell number proteomic analysis using in-cell protease digests reveals a robust signature for cell cycle state classification

Authors:

Van Kelly[1,2], Aymen al-Rawi[1], David Lewis[1], Georg Kustatscher[1,2] and Tony Ly[1,3]*

Addresses:

1. Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Michael Swann Building, Max Born Crescent, Edinburgh EH9 3BF, UK
2. Institute of Quantitative Biology, Biochemistry and Biotechnology, University of Edinburgh, Michael Swann Building, Max Born Crescent, Edinburgh EH9 3BF, UK
3. Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH

* Correspondence should be addressed to Tony Ly (tly@ed.ac.uk)

Running title: Cell state proteomics and classification
Abbreviations: PRIMMUS, AMPL

Abstract

Comprehensive proteome analysis of rare cell phenotypes remains a significant challenge. We report a method for low cell number mass spectrometry (MS)-based proteomics using protease digestion of mildly formaldehyde-fixed cells *in cellulo*, which we call the 'in-cell digest'. We combined this with AMPL (Averaged MS1 Precursor Library Matching) to quantitatively characterise proteomes from low cell numbers of human lymphoblasts. 4,500 proteins were detected from 2,000 cells and 2,500 proteins were quantitated from 200 lymphoblasts. The ease of sample processing and high sensitivity makes this method exceptionally suited for the proteomic analysis of rare cell states, including immune cell subsets and cell cycle subphases.

To demonstrate the method, we characterised the proteome changes across 16 cell cycle states isolated from an asynchronous TK6 human lymphoblast culture, avoiding synchronization. States included late mitotic cells present at extremely low frequency. We identified 119 pseudoperiodic proteins (PsPs) that vary across the cell cycle. Clustering of the PsPs showed abundance patterns consistent with 'waves' of protein degradation in late S, at the G2&M border, mid-mitosis and at mitotic exit. These clusters were distinguished by significant differences in predicted nuclear localization and interaction with the APC/C. The dataset also identifies putative APC/C substrates in mitosis and the temporal order in which they are targeted for degradation.

We demonstrate that a protein signature made of these 119 high confidence cell cycle regulated proteins can be used to perform unbiased classification of proteomes into cell cycle states. We applied this signature to 296 proteomes that encompass a range of quantitation methods, cell types, and experimental conditions. The analysis confidently assigns a cell cycle state for 49 proteomes, including correct classification for proteomes from synchronized cells. We anticipate this robust cell cycle protein signature will be crucial for classifying cell states in single cell proteomes.

Introduction

The proteome is a functional readout of cellular phenotype. Cellular phenotype includes dynamic and persistent features that reflect cell state and cell type, respectively. Rare cell phenotypes play key physiological roles. Quiescent stem cells, while often rare relative to differentiated cell types in a tissue, are essential for tissue homeostasis. Similarly, mitosis is a dynamic cell state that is critical for the accurate propagation of genetic material. Mitotic states are generally short-lived and thus rare in an asynchronous population. Proteomic analysis of these critically important cell phenotypes is a major challenge because typical proteomic workflows require $>10^5$ cells as input.

We previously developed an approach called 'PRIMMUS' or '*P*roteomics of *I*ntracellular I*mmu*nostained *Subsets*" to analyse abundant and rare cell cycle states [1]. Formaldehyde-fixed cells are fractionated into specific cell states by staining cells for intracellular markers and separating them using Fluorescence-Activated Cell Sorting (FACS). Cells grown in asynchronous culture are immediately fixed, thereby minimizing perturbation to physiological processes. This step is critical, as small molecule-based synchronsation can lead to effects on the proteome that are associated with stress responses arising from arrest rather than cell cycle regulation *per se* [2]. PRIMMUS enabled analysis of interphase and mitotic subpopulations, but this approach was limited to relatively abundant subpopulations for which $>10^5$ cells can be collected by FACS within a reasonable time [3].

Low input proteome analysis requires specialized methods for handling low cell number of cells [4,5]. Major improvements have been made by adapting methods used for bulk samples to low cell number samples [6-8]. Recent advances in small volume sample handling to nanoliter volumes have also enabled analysis of <10 cultured human cells, with overall number of proteins detected scaling with cell number [4,9,10]. For example, ~3,000 proteins were identified from 10 HeLa cells using 'nanodroplet processing in one pot for trace samples' [11]. In general, these methods require specialized equipment, ranging from automated robotic sample handling to custom microfabricated chips, which are challenging to satisfy in most labs currently.

102 Cells fixed with formaldehyde introduces additional challenges for bottom-up MS-
103 based proteomics. Formaldehyde crosslinks proteins by forming methylene bridges
104 primarily between lysine residues. Peptide/protein crosslinks are broken with heating
105 to >65 °C. As an example, formalin fixed tissue processing protocols include heating for
106 1 hour at 95 °C. However, the fixative concentration and treatment duration for formalin
107 fixed tissues is much higher (4% formaldehyde for up to several hours). Studies on
108 synthetic peptides demonstrated that protein amino acid residues can be irreversibly
109 modified by formaldehyde, producing chemical modifications and corresponding mass
110 shifts that are not detected in conventional database searches [12,13]. In contrast,
111 formaldehyde fixation for immune cell immunostaining and flow cytometry in clinical and
112 academic research settings is frequently much lower (0.1 – 3%) and carried out under
113 controlled conditions with limited treatment duration (10 – 30 min).
114 Here, we report a methodological advance that eliminates several steps
115 previously required for processing fixed cells for proteomics. We demonstrate that fixed
116 cells in suspension can be directly digested by trypsin without heat-induced crosslink
117 reversal for quantitative proteomics. We call this streamlined approach the *in-cell digest*.
118 The in-cell digest provides major improvements in sensitivity and convenience in
119 performing proteomic analysis on low numbers of fixed cells. To overcome the duty
120 cycle limitations of the Orbitrap Elite instrument, we developed an acquisition method
121 called AMPL. We applied in-cell digest and AMPL with PRIMMUS to analyse the
122 proteomic variation during an unperturbed cell cycle in human lymphoblasts with
123 unparalleled temporal resolution to produce unbiased proteomic definitions of cell cycle
124 state.

125 **Experimental Procedures**

126 *Experimental Design and Statistical Rationale*

127 Four biological replicates of 16 cell cycle populations were collected by FACS, with 2
128 technical replicates of the 64 samples being acquired by LCMS (AMP acquisition
129 strategy) resulting in 128 LCMS analyses, providing 8 pseudotimecoures for periodicity
130 analysis. Three libraries were generated from 12 HPRP fractions of unsorted cells,
131 interphase cells, and mitotic cells. Each library fraction was analysed twice (or thrice for

132 the mitotic library) resulting in a library of 85 LCMS analyses. Libraries were used to
133 increase proteome coverage through MS1 feature matching.
134 Supporting experiments include the analysis of 12 HPRP fractions of formaldehyde
135 fixed, fixed and reversed, and non-fixed control without replicates for a qualitative
136 comparison of peptide modifications. Twelve cell titration samples were also collected in
137 duplicate up to 2,000 cells by FACS, including a zero-cell control, to assess LC-MS
138 sensitivity of the improved processing and AMP acquisition methods. The 24 cell
139 titration samples were analysed by AMP LCMS along with a 12 HPRP fraction library
140 and an unfractionated library of 2000 sorted cells, which were analysed by DDA LC-MS.
141 To assess the impact of peptide filtering on MS1 feature matching FDR, unmodified,
142 dimethylated, and isopropylated peptides were analysed by AMPL and DDA, along with
143 a library of 12 HPRP fractions.
144 *Cell culture*
145 TK6 human lymphoblasts [14] were obtained from the Earnshaw laboratory (University
146 of Edinburgh). Cells were cultured at 37 °C in the presence of 5% CO_2 as a suspension
147 in RPMI-1640 + GlutaMAX (Thermo Scientific) supplemented with 10% v/v fetal bovine
148 serum (FBS, Thermo Scientific). Cell cultures were maintained at densities no higher
149 than 2 x 10^6 cells per ml. MCF10A cells (ATCC) were cultured in phenol red-free
150 F12/DMEM media (Thermo Scientific) supplemented with 5% horse serum, 10 µg/ml
151 insulin (Sigma), 100 ng/ml cholera toxin (Sigma), 20 ng/ml EGF (Sigma), 0.5 µg/ml
152 hydrocortisone (Sigma), 100 units/ml penicillin and 100 µg/ml streptomycin (Thermo
153 Scientific) at 37 °C in the presence of 5% CO_2. Cells were maintained at less than 100%
154 confluency and were discarded when passage number exceed 20 passages. U2OS
155 cells (ATCC) were cultured in DMEM media high glucose + GlutaMAX (Thermo
156 Scientific) supplemented with 10% v/v FBS (Thermo Scientific). Cells were checked for
157 mycoplasma at the point of cryo-storage using a luminescence-based assay (Lonza).
158 *Cell fixation and immunostaining*
159 Cells were washed with Dulbecco's phosphate-buffered saline (DPBS, Lonza) and
160 resuspended in freshly prepared 1% formaldehyde solution (w/v) from a 16% stock (w/v,
161 Thermo Scientific) in DPBS, fixed for 10 min at room temperature with gentle rotation,

162 pelleted, washed with DPBS and permeabilized with cold 90% methanol. Cells were
163 stored at -20°C prior to staining.
164 Cells stored in methanol were washed with DPBS and resuspended in blocking
165 buffer, which is composed of 5% bovine serum albumin (BSA) in 0.1 M Tris-buffered
166 saline, pH 7.4 (TBS). Cells were blocked for 10 min at room temperature, pelleted, and
167 resuspended in primary antibody solution. The rat anti-H3S28ph HTA28 (abcam
168 ab10543), mouse anti-cyclin A2 (Cell Signaling Technologies 4656S), and rabbit anti-
169 cyclin B1 (12231S) were used for staining as 1:200 dilutions in blocking buffer. Cells
170 were stained with primary antibody overnight at 4°C. Stained cells were then washed
171 twice with wash buffer (DPBS + 0.5% BSA) and stained with dye-conjugated secondary
172 antibodies. The donkey anti-rat IgG H&L AlexaFluor 568 preadsorbed (abcam
173 ab175475), Donkey anti Mouse IgG (H+L) Highly Cross Adsorbed Secondary Antibody,
174 Alexa Fluor 488 (Thermo A21202), and Goat Anti-Rabbit IgG H&L (Alexa Fluor® 647)
175 preadsorbed (abcam ab150083) were used as 1:200 dilutions in blocking buffer. Cells
176 were stained stained for 1 hour at room temperature, washed twice with DPBS,
177 pelleted, and stained in DAPI solution (Sigma, 20 µg/ml in DPBS + 0.1% BSA) for at
178 least 1 hour prior to FACS.
179 *FACS and gating strategy*
180 Cells were collected using a BD FACSAria Fusion Cell Sorter equipped with 355nm UV,
181 405nm Violet, 488nm Blue, 561nm YG and 640nm Red lasers, and controlled by BD
182 FACS Diva V8.0.1 software. Cells were first gated into 'narrow' (P1 – P8) and 'wide' (P9
183 – P16) populations based on DAPI fluorescence signal width. The narrow population
184 contains single cells either in interphase, or in mitosis up to late anaphase. These single
185 cells were then separated based on cyclin B into 8 different stages of interphase.
186 Population P1 has low to no cyclin B protein and 2N DNA content, consistent with low to
187 no E2F activity and a G0/early G1 cell state. Cyclin B rises monotonically from P2 to P6
188 and then rises more steeply from P6 to P8. Like cyclin B, cyclin A also increases during
189 interphase, but at a faster rate from P1 to P6 as compared to P6 to P8. P9 to P13 are
190 positive for histone H3 phosphorylation at Ser28 (pH3+). Highest levels of pH3+ are
191 present in prometaphase and metaphase. Rising and declining H3 phosphorylation in
192 early and late mitosis, respectively, result in low to medium levels of pH3+. Cyclin A and

193 cyclin B levels are used to further discriminate mitotic subphases, as they are degraded
194 during prometaphase and the metaphase-to-anaphase transition, respectively.
195 Finally, late mitotic subphases are enriched in the wide population, but so too are
196 doublets. We reasoned that most doublets will have cyclin B signal, as single cells with
197 the exception of P1 are cyclin B positive. Thus, we can further enrich late mitotic stages
198 by selecting wide, 4N, cyclin B negative cells (P14-P16). P14-P16 are then
199 discriminated further by pH3+ levels, which decrease during mitotic exit. We note that
200 P16 may contain doublets of G0/early G1 cells (P1), but P14 and P15 should not as
201 P14 and P15 are pH3+ and G0/early G1 cells are negative for pH3.
202 5000 cells for each gated population were collected using 4-way purity using
203 either a 85 or 100 µm nozzle, into 1.5 ml Eppendorf Protein Lo-Bind tubes. Four
204 biological replicates were collected. An interphase library sample were collected by
205 combining 300,000 cells of G0/G1, S, and G2 populations. A mitotic library sample was
206 composed of 800,000 mitotic cells gated by high DNA content and high Histone H3
207 Ser28 phosphorylation. Samples were centrifuged and supernatant removed before
208 storing at -20 ºC.
209 *In-cell digest*
210 Cell sorted library samples, and unstained unsorted TK6 cells, were resuspended in
211 DPBS at 2 - 5 million cells per ml and incubated with 1 µl (25 - 29 U) benzonase
212 (Millipore) at 37 ºC for a minimum of 1 hr. Trypsin was added to approximately 1:25 w/w
213 and in-cell digested at 37 ºC for ~16 hrs. Digests were acidified with TFA and desalted
214 over Sep-Pak C18 cartridges (Waters) and dried.
215 Individual populations of 5,000 cells were diluted with 40 µl PBS and incubated
216 with 0.25 µl [6 – 7 U] benzonase at 37 ºC for a minimum of 1 hr, then digested with 50
217 ng trypsin (~1:10 w/w) at 37 ºC for ~16 hrs. Samples were acidified with TFA and
218 desalted over self-made C18 columns with 3 Empore C18 disks [10] and eluted directly
219 into Axygen™ 96-well PCR Microplates (Fisher Scientific) and dried.
220 *High pH reverse phase fractionation*
221 Approximately 100 µg interphase, mitotic, and unsorted TK6 cell digests were
222 fractionated by high-pH reverse phase chromatography using an Ultimate 3000 HPLC
223 (ThermoFisher Scientific) and a 1 x 100 mm 1.7 µm Acquity UPLC BEH C18 column

224 (Waters). Peptides were separated using a constant 10 mM ammonium formate (pH 10)
225 and a gradient of water and 100% acetonitrile. Peptides were loaded at 1% acetonitrile
226 followed by separation by a 48 min multistep gradient of acetonitrile from 3% to 6%,
227 25%, 45% and 80% acetonitrile at 4, 34, 44, 45 minutes, respectively, followed by an
228 80% wash and re-equilibration. Fractions were collected at 30 sec intervals resulting in
229 96 fractions which were concatenated into 12, and 1 µg aliquots dried.
230 *LC-MS/MS*
231 Peptide samples were resuspended in 0.1% TFA. Approximately 0.5 µg of library
232 fractions were injected for DDA LCMS analysis. A volume equal to half the cell
233 population (equivalent to ~2,500 cells) was injected and analysed twice by AMPL to
234 produce two technical replicates for each of the four biological replicates. An Ultimate
235 3000 RSLCnano HPLC (Dionex, Thermo Fisher Scientific) was coupled via electrospray
236 ionisation to an Orbitrap Elite Hybrid Ion Trap-Orbitrap (Thermo Fisher Scientific).
237 Peptides were loaded directly onto a 75 µm x 50 cm PepMap-C18 EASY-Spray LC
238 Column (Thermo Fisher Scientific) and eluted at 250 nl/min using 0.1% formic acid
239 (Solvent A) and 80% acetonitrile/0.1% formic acid (Solvent B). Samples were eluted
240 over 90 min stepped linear gradient from 1% to 30% B over 72 min, then to 45% B over
241 18 min. AMPL analyses included up to 5 MS1 microscans of 1E6 ions in the Orbitrap at
242 120k resolution and with a 250 ms maximum injection time. MS1 scans were acquired
243 over 350-1700 m/z and a 'lock mass' of 445.120025 m/z was used. This was followed
244 by 5 data-dependent MS2 CID events (5E3 target ion accumulation) in the ion trap at
245 rapid resolution with a 2 Da isolation width, a normalised collision energy of 35, 50 ms
246 maximum fill time, a requirement of a 10k precursor intensity, and a charge of 2+ or
247 more. Precursors within 5 ppm were dynamically excluded for 40 sec. DDA analyses
248 were as for AMPL but with a single MS1 microscan with a 75 ms maximum injection
249 time, followed by 20 CID events in the ion trap.
250 Libraries were acquired as for DDA analyses or acquired with 10 data-dependent
251 MS2 HCD events at 30 NCE of 5E4 ions in the Orbitrap at 15k resolution and a
252 maximum fill time of 100 ms, with a precursor intensity required to be at least 50k. For
253 the sample preparation comparisons shown in Fig. 2, a 240 min gradient was used (1%
254 to 30% B for 210 min, then to 42% B over 30 min). MS data was acquired as for DDA

255 analysis described above with the exception that MS1 spectra were acquired at 60k
256 resolution and MS2 events were acquired only on 2+ and 3+ precursors.
257 *MS/MS data analysis*
258 Data was processed using MaxQuant version 1.6.2.6 [15]. LC-MS/MS data was
259 searched against the Human Reference Proteome from UniProt including splice-
260 isoforms (accessed October 23rd, 2017), which contains 93,613 entries, allowing for 2
261 tryptic missed cleavages, allowing for variable methionine oxidation and protein N-
262 terminal acetylation. Carbamidomethyl cysteine modification was allowed only for
263 samples that were alkylated by iodoacetamide. The parameter “Individual peptide mass
264 tolerance” was selected for variable precursor mass tolerances, with 0.5 Da or 20 ppm
265 mass tolerances set for ion trap or orbitrap fragment ions, respectively. A target-decoy
266 threshold of 1% was set for both PSM and protein false discovery rate. Match-between-
267 runs was enabled with identification transfer within 0.5 mins and a retention time
268 alignment within 20 min window. Matching was permitted from the library parameter
269 group, and 'from and to' the unfractionated parameter group. The parameter "Require
270 MS/MS for LFQ comparisons" was deselected, and second peptide search was
271 enabled. Both modified and unmodified unique and razor peptides were used for
272 quantification. ‘Evidence’ and ‘proteinGroups’ output files were used for subsequent
273 analysis in R.
274 *Match-between-runs FDR filtering*
275 A reference sample was generated by lysing TK6 cells in DPBS with 2% SDS and
276 cOMPLETE protease inhibitors without EDTA (Roche, 1x concentration) at 70 ºC,
277 homogenised with a probe sonicator and treated with benzonase. Protein was reduced
278 with 20 mM TCEP for 2 hr before alkylation with 20 mM iodoacetamide at ambient
279 temperature in the dark for 1 hr. Protein was precipitated with 4 volumes cold acetone at
280 -20 ºC overnight, washed with 100% cold acetone and 90% cold ethanol. Protein pellet
281 was air dried before resuspending in DPBS and digesting with 1:50 w/w trypsin for
282 ~16hrs. Peptides were acidified, desalted, aliquoted, and fractionated as previously
283 described. For isopropylation, 50 µg peptides were resuspended in 200 µl 90%
284 acetonitrile containing 0.1% formic acid before addition of 50 µl acetone containing 36
285 µg/µl $NaBH_3CN$. The reaction was conducted at ambient temperature for ~16 hrs before

286 quenching with ammonium bicarbonate, drying off solvent and desalting peptides over
287 C18. For dimethylation, 50 µg peptide was resuspended in 200 µl DPBS before addition
288 of 0.32% formaldehyde and 50 mM $NaBH_3CN$. The reaction was conducted at ambient
289 temperature for ~16 hrs before quenching with ammonium bicarbonate and desalting
290 peptides over C18. 200 ng of unmodified, dimethylated, and isopropylated peptides
291 were analysed by AMPL and DDA, and unmodified fractionated peptide samples were
292 analysis by DDA, as previously described. LCMS data were searched using MaxQuant,
293 as previously described. Note that dimethylation and isopropylation modifications were
294 not specified in in the search parameters.

295 *Cell cycle proteomic data analysis*

296 All subsequent data analysis on the protein intensity table, including the analysis
297 of pseudoperiodicity, was performed using R (v. 3.5.0) within the RStudio integrated
298 development environment. The R scripts are available as Supplementary Data 1. The
299 list of validated APC/C substrates was obtained from the APC/C degron repository
300 (http://slim.icr.ac.uk/apc/). Proteins that contain D box, KEN and ABBA SLIMs in the
301 human proteome were found using SLiMsearch with default settings (Disorder score
302 cut-off: 0.30, Flank length: 5). In order to remove slight variations in total protein amount
303 in each sample, protein intensities were divided by total intensities per sample and
304 multiplied by 10^6 to obtain intensities in parts per million (ppm). There are four biological
305 replicates analysed in technical duplicate. As described above, sample analysis was
306 completely randomized in the second technical repeat. Each technical repeat (i.e. set of
307 four biological replicates) are considered as one 'pseudotimecourse' with samples in
308 each biological replicate arranged in order from P1 to P16. Each of the two
309 pseudotimecourse was then independently subjected to a Fisher's test for periodicity, as
310 implemented in the ptest R library (v. 1.0-8). Fisher's periodicity test p-values were
311 corrected for multiple hypothesis testing using the q value method as implemented in
312 the qvalue R library (2.15.0). Those proteins that showed q values <0.10 in both sets of
313 biological replicates and oscillation frequencies of either 0.0625 (1/16) or 0.125 (1/8)
314 were classified as pseudoperiodic.

315 For clustering, protein ppm values were averaged (mean) to produce a single
316 pseudotimecourse for each protein. These average abundance profiles were scaled

317 using the base R function scale and subjected to hierarchal clustering using the Ward
318 minimum variance algorithm. The appropriate range for cluster number was identified as
319 3 - 6 clusters using the 'elbow method', which involves plotting within-cluster sum of
320 squares versus number of clusters. Bifurcating leaves of the subsequent dendrogram
321 were swapped in order to produce a heatmap that follows a logical, sequential order of
322 peak abundance, i.e. cluster 1 with highest abundance in P0-P8 and cluster 5 with peak
323 abundance in P3-P7, etc.
324 For PCA and cell cycle state classification, scaled pseudotimecourses were
325 used. Cell cycle states were classified using the k-NN model as implemented in the
326 class R library (v. 7.3-15) using k = 6, with k being the number of nearest neighbours for
327 classification. Three biological replicates were used as the training set and the
328 remaining replicate was used as a test set.
329 For the pairwise comparison of the proteomes of P17 with P1 and P16, t-tests
330 were performed on ppm intensities. Uncorrected p-values were plotted against mean
331 fold change in order to identify candidate proteins that were specifically changed in
332 abundance in P17.
333 **Results**
334 *The 'in-cell digest': direct protease digestion of fixed cells*
335 Based on previous work [1,16], we hypothesized that formaldehyde-induced
336 modifications were of low stoichiometry and crosslink reversal may not be required for
337 proteome analysis. Consistent with this idea, deep proteome analysis comparing human
338 epithelial MCF10A cells fixed with 2% formaldehyde for 10 min, fixed and treated with
339 heating to reverse formaldehyde crosslinks, or not fixed (Fig. 1A and Supp. Fig. 1A)
340 showed no significant differences in protein and peptide coverage (Figs. 1B-C). These
341 proteomes were analysed to a depth of ~53,600 peptides and ~7,700 proteins. We next
342 used the error-tolerant MS search to find peptides chemically modified by formaldehyde.
343 The pattern and frequency of detected mass shifts are remarkably similar between
344 control and fixed samples (Suppl. Fig. 1B). From these observations, we concluded that
345 under these controlled and mild fixation conditions, the stoichiometry of crosslinking and
346 chemical modification by formaldehyde is sufficiently low such that the non-detection of

347 modified and crosslinked peptides is not detrimental for characterization of proteomes to
348 a depth of at least 7,700 proteins.

349 We next hypothesized that fixed cells may make suitable substrates for direct
350 protease digestion. Digestion of fixed cells would significantly simplify the sample
351 processing workflow by eliminating several steps, including detergent-based lysis,
352 homogenization, heat treatment and detergent removal. We therefore treated fixed,
353 permeabilized cells suspended in DPBS with either mock treatment (DPBS), or trypsin,
354 and monitored cell morphology by brightfield microscopy. As shown in Fig. 1D,
355 prominent structural features visible in control cells, such as the plasma membrane,
356 nuclei and nucleoli, are degraded in a time-dependent manner by trypsin (see
357 Supplementary Video 1). For LC-MS/MS analysis, fixed cells were also pre-incubated
358 with benzonase to digest RNA and DNA oligonucleotides, which may interfere with
359 downstream sample processing. The peptide-containing supernatant from the digest
360 was then subjected to C18 purification prior to analysis by LC-MS/MS. As the digestion
361 occurs within the fixed cells, we have called this approach an 'in-cell digest' (Fig. 1E).

362 As shown in Fig. 1F, the proteome coverages are similar for fixed cells
363 processed by the in-cell digest method (~4,678 proteins, n = 3), fixed samples that were
364 subjected to the PRIMMUS protocol (~4,446 proteins, n = 3) and extracts from non-fixed
365 cells processed by precipitation (see Methods, ~4,561 proteins, n = 3). We conclude
366 that the proteome coverage from the in-cell digest is similar, or higher, than the other
367 protocols tested.

368 We did not observe a broad bias in quantitation, as label free intensities
369 measured in fixed cells prepared by the in-cell digest and by decrosslinking followed by
370 an in-solution digest showed high correlation (Fig. 2G, $\rho = 0.96$). Similarly, a high
371 correlation was observed between fixed cells prepared by the in-cell digest and non-
372 fixed cells (Fig. 2H, $\rho = 0.97$). Few points lie off-diagonal, indicating that proteins
373 showing a difference in intensity between methods are rare. We then tested if these off-
374 diagonal proteins were enriched in any UniProt keywords or GO annotations using
375 DAVID. The only terms that were significantly enriched in proteins showing lower
376 intensity with the in-cell digest were associated with RNA-binding (FDR < 0.05, Suppl.
377 Fig. 1C). Notably, these RNA-binding proteins are present in cells at high abundance. In

378 contrast, proteins showing higher intensity with the in-cell digest are enriched in
379 membrane proteins (FDR < 0.05, Suppl. Fig. 1D). Improved recovery of membrane
380 proteins using the in-cell digest is consistent with previous results demonstrating that
381 heat treatment can irreversibly precipitate membrane proteins.
382 We conclude that the measurements of protein abundance from the in-cell digest
383 are quantitative, reproducible and broadly comparable to conventional sample
384 preparation methods. We note that each sample preparation method will have its own
385 specific biases. In the case of the in-cell digest, the increased abundance of membrane
386 proteins may more accurately reflect the abundance of these proteins in cells.
387 *Averaged MS1 Precursors with Library matching (AMPL) improves feature detection*
388 To increase the sensitivity and detection speed of the Orbitrap Elite MS instrument, we
389 utilised MS1-based identification and quantitation using accurate mass and retention
390 time matching, as proposed originally by the Smith lab [18]. This approach has been
391 recently demonstrated to be highly sensitive in an implementation called BoxCar [19].
392 The BoxCar method increases the signal-to-noise of trap-based MS by collecting ions
393 using segmented, spaced windows. Peptide identification relies on MS1 feature
394 matching to a reference library generated from a fractionated reference sample using
395 the MaxQuant function 'Match-between-runs' (MBR). The library is analysed separately
396 using data-dependent acquisition (DDA) and peptides are identified by MS2 and
397 database searches.
398 As the BoxCar method cannot be directly implemented on the Orbitrap Elite, we
399 developed a different approach to increase the dynamic range of MS1 feature detection.
400 MS1 spectral averaging is frequently performed in direct infusion MS, but rarely
401 employed in LC-MS bottom-up proteomics. We surmised that averaging several MS1
402 scans would improve signal-to-noise (S/N) and would rapidly plateau as it is known that
403 averaging improves S/N by a factor of sqrt(n) where n is the number of spectra
404 averaged. Features would then be matched between the single shot analyses to a
405 fractionated reference library (Fig. 2A). We call this method Averaged MS1 Precursors
406 with Library matching (AMPL), or AMP if no library is used. Like BoxCar, AMP(L)
407 prioritises MS1 scans over MS2 scans as compared with DDA (Fig. 2B) and includes

408 top-5 DDA MS/MS scans to ensure identification of features for accurate retention time
409 alignment throughout the chromatographic separation.
410 We therefore tested AMPL by analysing 1µg on-column loads of MCF10A tryptic
411 digests. A comparison of different MS1 scans (n = 1, 3, 4, 5) showed that the number of
412 features and peptides identified saturates at n = 4 (Suppl. Figs. 2A-B). AMPL (n = 4)
413 detects ~278,205 features, representing a 20% increase compared to a standard Top
414 20 DDA acquisition using the same gradient (188,928 features). We reasoned that the
415 additional peptides detected by AMPL originate from low-abundance features detected
416 by virtue of the S/N increase due to averaging. Fig. 2C compares the peptide intensity
417 distributions between DDA-L and AMPL. The distributions are bimodal, with MS/MS-
418 dependent identification biased towards higher intensity features (cyan). Consistent with
419 the idea that AMPL improves S/N, AMPL detects a higher number of matched features
420 (pink) in the low abundance regime. Similar to previous MS1-based matching
421 approaches, AMPL shows higher data completeness (4,411 proteins with intensities
422 measured in all 10 replicates) as compared with DDA-L (3,493 proteins) and DDA
423 (2,865 proteins) (Suppl. Fig. 2C).
424 MS1-based matching significantly increases the sensitivity, coverage and data
425 completeness of MS-based proteomics. However, the lack of MS2-based identification
426 for these matched sequences could potentially increase the false discovery rate (FDR).
427 We estimated the matching FDR by using an empirical 'target-decoy' approach where
428 decoy proteomes created by chemical modification (dimethylation and isopropylation)
429 are matched against an unmodified library (Fig. 2D). Whereas matches to the target
430 proteome will contain both true and false positives, matches to the decoy proteomes
431 should contain exclusively false positives (with the rare exception of peptides containing
432 an N-terminal acetyl group a C-terminal arginine, which are not
433 dimethylated/isopropylated). ~32% of the features are assigned a peptide sequence
434 when the target, unmodified proteome is matched against an unmodified library (Suppl.
435 Fig. 2D). By contrast, only ~2% of the features are matched in the decoy samples
436 (Suppl. Fig. 2D). Using this approach, the estimated match FDR is 7.4%. To reduce the
437 FDR <5%, we applied more stringent thresholds for match time, match m/z and match
438 m/z error (2.5 σ for match time, 3 σ for match m/z and match m/z error, Suppl. Fig. 2E-

439 F). Application of these thresholds reduced the estimated FDR to 3.1-3.4% (Fig. 2E)
440 while retaining 96% of the matches in the target dataset.

441 The improvements in detecting low abundance features suggest that AMPL may
442 be well suited to analysis of low sample loads. AMP (i.e. no library) consistently detects
443 more features than DDA (Fig. 2F), which leads to significant improvements in peptide
444 coverage (Fig. 2G). For example, at 10 ng loading, 21,483 unique peptides are
445 quantitated by AMPL versus 14,702 by DDA-L, representing a 46% increase in
446 coverage. AMPL provides 150-535% improvement relative to conventional DDA with no
447 library and 24-46% improvement relative to DDA-L for protein coverage at all tested
448 column loads with greatest gains observed at low column load.

449 As shown in Fig. 4G, AMPL detects a slightly higher number of peptides in 10 ng
450 on-column load as DDA with 1 µg load, demonstrating a >100x increase in sensitivity. A
451 10 ng on-column load is equivalent to the protein content of ~67 cells based on the
452 protein per cell measured in bulk assays. However, the effective number of cells
453 required for proteome analysis is frequently much higher. This is due to losses during
454 sample preparation. We reasoned that these losses are significantly reduced using the
455 streamlined in-cell digest.

456 We combined the in-cell digest with AMPL to analyse FACS collected TK6 cells,
457 a human lymphoblastoid cell line with a stable near-diploid karyotype. Notably, TK6
458 cells are smaller than typical adherent human cell lines, such as HeLa and MCF10A.
459 Being cultured in suspension, TK6 cells are amenable towards cell separation
460 techniques, including fluorescence-activated cell sorting (FACS) and centrifugal
461 elutriation, without requiring cell dissociation, which can induce physiological
462 perturbations.

463 Fig. 2H shows the result of a cell titration analysis of S-phase cells performed in
464 duplicate whereby two aliquots at each indicated cell number (2,000 cells to 0 cells)
465 were collected by FACS from the same starting cell population. ~4,500 proteins were
466 quantitated with 2,000 cells, with 4,480 proteins reproducibly quantitated in two
467 technical repeats. At the lower end of the cell titration, 2,933 proteins on average were
468 quantitated from 200 cells. We note that below this number of cells, we observe a
469 higher variability in proteome coverage, which will need to be addressed by further

470 optimization. Indeed, while approx. 30 proteins were detected in single cells, with 17
471 reproducibly detected, nearly all of these proteins were also detected in the background
472 samples ('0 cells').
473 We conclude that combining in-cell digest and AMPL enables characterization of
474 proteomes of 2,000 cells to a protein depth comparable to conventional single shot DDA
475 analysis of 1 µg on-column loads. The advanced PRIMMUS method presented here
476 significantly reduces the number of cells required, i.e. ~10^3 versus ~10^5 with low
477 estimated match FDR (<3.5%).
478 *High temporal resolution analysis of an unperturbed cell cycle using PRIMMUS*
479 The process of normal cell division requires linear progression through several cellular
480 states i.e. S- and M-phases) in which DNA replication and mitosis must occur in
481 sequential order. These states can be further resolved. DNA replication, for example,
482 occurs in a temporally and spatially patterned manner, with euchromatic genomic
483 regions replication first before heterochromatin-dense regions. Similarly, M-phase can
484 be resolved into prophase, prometaphase, metaphase, anaphase I, anaphase II and
485 telophase based on cytological features. Some of these phases, including telophase are
486 exceptionally rare in asynchronous cells and are not amenable for collection by FACS in
487 numbers required for typical proteomic analysis. We therefore developed an advanced
488 PRIMMUS workflow incorporating the in-cell digest to target these rare cell states and
489 carry out a high temporal resolution analysis of proteome variation across 16 cell cycle
490 subpopulations, including 8 interphase and 8 mitotic states. This fractionation-based
491 approach to separating cell cycle phases relies on continuous cell trajectories, such as
492 cell cycle progression in asynchronous populations that are unperturbed by drug-based
493 synchronisation.
494 TK6 cells were immunostained for DNA content, cyclin B, cyclin A and histone
495 H3 phosphorylation (Ser28), which are all markers of cell cycle progression. Cells were
496 then separated into 16 cell cycle populations (P1 – P16) (see Suppl. Fig. 3 for the full
497 gating strategy). Biochemical differences are used as a surrogate for time and cell cycle
498 progression. Based on past literature [20,21] and our previous data [1], we have
499 correlated these biochemical changes with specific cell cycle states (as illustrated in Fig.
500 3B, top). For example, cyclin A and cyclin B levels are used to discriminate mitotic

501 subphases, as they are degraded during prometaphase and the metaphase-to-
502 anaphase transition, respectively.
503 The rarest target population are cells in late anaphase of mitosis, which are
504 present in 0.01% of an asynchronous TK6 culture. Proteome characterisation of these
505 cells, previously challenging due to lack of sensitivity, is now possible with the in-cell
506 digest. Four separate cultures of TK6 cells were independently FACS separated into 16
507 populations. For each population, 5,000 cells were collected and processed using the
508 in-cell digest. Collection of 5,000 cells provided sufficient material for duplicate
509 injections for LC-MS/MS analysis by AMPL with DDA libraries generated from
510 interphase, mitotic and asynchronous cells. The data were then processed by
511 MaxQuant with MBR and filtered by match parameters as discussed above to generate
512 a dataset with 6,899 quantitated proteins (Suppl. Table 1).
513 Next, to identify cell cycle regulated proteins, we treated each set of 16
514 populations as an ordered series of related biochemical states. These states were
515 projected onto a temporal axis (i.e. cell cycle progression). A single replicate series of
516 ordered cell states constitutes a 'pseudotimecourse' (Fig. 3A, bottom). We then applied
517 the Fisher's periodicity test to identify '*ps*eudoperiodic *p*roteins' (PsPs), i.e. proteins
518 abundance patterns that showed periodic behavior across the four pseudoperiodic
519 timecourses. In order to increase robustness, the periodicity test was separately
520 performed on each technical repeat and only those proteins showing periodicity in both
521 were designated as PsPs. Fig. 3A (bottom) shows the abundance profiles for heat
522 shock protein HSP90AA1 and ATPase AAA domain-containing protein ATAD2 as an
523 example non-PsP and PsP, respectively. ATAD2 shows highly reproducible abundance
524 variation in all 8 pseudotimecourses, with peak abundance in S-phase populations (P5-
525 P6), consistent with previous reports [22]. In total, 119 PsPs were identified using these
526 criteria (Fig. 3A, bottom, Suppl. Table 2).
527 Hierarchal clustering of the 119 PsPs identified five major classes of protein
528 abundance patterns (Fig. 3B). Each cluster shows peak abundance at different times
529 during cell cycle progression. The gene ontology terms enriched for each cluster reflects
530 key processes and/or compartments associated with that period during cell cycle
531 progression. We also assessed enrichment in short linear (sequence) motifs (SLIMs).

532 SLIMs mediate protein-protein interactions that lead to changes in post-translational
533 modification, stability and/or subcellular localization of a protein. Using the eukaryotic
534 linear motif (ELM) database [23], we identified SLIMs that are enriched in each cluster
535 ($p < 0.01$, Fisher's exact test, Suppl. Table 3).
536 Cluster 1 proteins show high abundance in interphase, which decreases in early
537 mitosis (P8-P10) and recovers slightly in late mitotic populations (P15-P16). This cluster
538 is highly enriched in proteins with a monopartite nuclear import signal sequence (Fig.
539 3D) and in contrast to other clusters, do not show any enrichment for the Crm1-
540 mediated nuclear export signal sequence (Fig. 3E). Most proteins in this cluster are
541 either RNA- or DNA-binding (26 / 33). For example, several mRNA splicing factors are
542 in this group, including serine/arginine-rich proteins (SRRM2, SRSF2, SRSF3, SRSF5,
543 SRSF6, SRSF10/TRAB). These proteins reproducibly decrease in abundance in
544 mitosis, but with a small fold change (< 2-fold) than key cell cycle regulators, e.g. cyclin
545 B1 (> 4-fold). The stability of the SR proteins is regulated by nucleocytoplasmic
546 shuttling. For example, SRSF1 is stable in the nucleus but has a short half-life in the
547 cytoplasm [24]. Proteasome-dependent degradation of SR proteins is dependent on the
548 RS domain, which is shared among SR proteins [25]. Cluster 1 is also enriched in
549 poly(A)-binding proteins in the nucleus that are involved in pre-mRNA and ribosomal
550 RNA processing, e.g. XRN2, NOLC1. The remaining proteins with no known or
551 anticipated oligonucleotide-binding properties are enriched in cytoskeleton-binding
552 factors, e.g. the actin-binding protein MARCKS, CCDC6, CEP89 and DBNL.
553 Cluster 2 proteins peak in late G1/S. Nearly all proteins in this cluster are directly
554 involved in DNA replication, establishment of nascent chromatin, or the G1/S transition
555 (Fig. 3C). In this cluster are three members of the MCM helicase (MCM2, MCM5,
556 MCM6), the replication-dependent histone chaperone (CHAF1B) and the histone mRNA
557 stem-loop binding factor SLBP, which is essential for the synthesis of histones for
558 incorporation into newly synthesized DNA in S-phase. This cluster also includes the
559 DNA damage checkpoint kinase ATM, which is important in resolving endogenous
560 replication stress [26].
561 Cluster 3 shows peak abundance in late S, G2 (P6 to P8) and decreased
562 abundance in early-mid mitosis (P9 to P11). Three proteins show >5-fold decrease in

563 abundance by mid-mitosis with low or undetectable levels in late mitosis: GMNN, RRM2
564 and PAF/KIAA0101. All three are targeted for degradation in late mitosis and G1 by the
565 APC/C-Cdh1. The remaining proteins in the cluster show an increase in S/G2 phase
566 and a decrease in prophase/prometaphase (P9-P12), followed by a slight recovery in
567 abundance in late mitosis. These include sororin/CDCA5, which functions in sister
568 chromatid cohesion establishment, and MIS18BP1, which facilitates loading of the
569 centromere-specific histone in late mitosis and G1. This cluster is enriched in chromatin-
570 binding factors, including TRIM28/KAP1, EXO1, sororin, PAF and MIS18BP1.
571 Clusters 4 and 5 show peak abundance during mitosis and contain the largest
572 proportion of proteins with either known direct roles in mitotic progression or targeted for
573 degradation in mitosis (9/12 for cluster 4, 38/46 for cluster 5). The feature that
574 distinguishes clusters 4 and 5 is the mitotic abundance pattern. Cluster 4 proteins show
575 decreased abundance in earlier mitotic populations, particularly in P11 – P12,
576 coincident with the onset of cyclin A2 and cyclin B1 degradation (c.f. Fig. 5B). The three
577 mitotic cyclins detected (A2, B1 and B2), the spindle assembly checkpoint kinases
578 BUB1 (Fig. 6F) and BUB1B (BubR1), the kinesin-8 family member KIF18B, securin
579 (PTTG1) and shugoshin (SGO2) are in this cluster. Functionally, this cluster is
580 characterized by proteins that protect sister chromatid cohesion (securin, shugoshin)
581 and the spindle assembly checkpoint that prevents anaphase (Bub kinases) while
582 proper microtubule attachment and biorientation of chromosomes takes place.
583 By contrast, cluster 5 proteins show a significant increase in abundance at the
584 end of interphase (P7 – P8) with peak abundance throughout mitosis (P9 – P15) and a
585 significant decrease only in the last population (P16), i.e. cells undergoing mitotic exit.
586 Example proteins include the catalytic E2 subunits of the APC/C (UBE2C, UBE2S), the
587 chromosome passenger complex (AURKB, INCENP, BIRC5 – Survivin, CDCA8 –
588 Borealin), polo kinase (PLK1) and the spindle-associated protein FAM83D. Both aurora
589 kinases (Aurora A and Aurora B) are known to relocalise to the central spindle after
590 anaphase onset. Aurora B activity is crucial for cytokinesis, the final step in cell division.
591 Clusters 4 and 5 are strongly enriched in the Crm1-mediated nuclear export
592 signal (NES, Fig. 3E). 8/12 proteins in cluster 4 match the NES consensus. Notably,
593 cluster 4 includes cyclins B1 and B2, and constitutive export of cyclin B-CDK from the

594 nucleus is important in preventing premature mitotic entry. Exclusion from the nucleus
595 of other proteins within these two clusters (Fig. 3E) may also be important in preventing
596 premature activation of processes that are normally restricted to mitosis.
597 We identified PsPs that have no reported function in cell cycle control. These
598 novel cell cycle regulated proteins may, like many of the other proteins identified in this
599 manner, have significant roles in cell cycle progression. These candidates include
600 EXO1, the DNA helicase PIF1, the guanine-exchange factor NET1 and the serine
601 protease FAM111B.
602 *Analysis of mitotic protein abundance dynamics in unperturbed cells*
603 A major difference between the clusters is the timing of protein abundance decrease
604 (Fig. 4A). A critical regulator of protein abundance during the cell cycle is the anaphase
605 promoting complex/cyclosome (APC/C). The APC/C is an E3 ubiquitin ligase and is
606 active during the mitotic and G0/G1 phases of the cell cycle [27,28]. Its substrates
607 include key regulators of the cell cycle, including cyclin A2 and cyclin B1. Ubiquitination
608 of APC/C substrates is tightly temporally controlled, with APC/C substrate specificity
609 changing during the cell cycle (Fig. 4B). This is mediated through changes in the APC/C
610 co-activators and substrate recognition factors, Cdc20 and Cdh1. While APC/C-Cdc20
611 is active in early mitosis, the substrate receptor changes to Cdh1 in late mitosis, thereby
612 conferring a temporal order to substrate degradation. Cdc20 is itself a substrate of the
613 APC/C-Cdh1, allowing for switch-like handover in substrate receptor control.
614 25 PsPs (out of 119) are experimentally validated APC/C substrates [29], and of
615 these, 24 are found in clusters 3, 4 and 5. Substrate recognition by APC/C-Cdc20 and
616 APC/C-Cdh1 is mediated by the interaction between WD40 domains on the APC/C-
617 (Cdc20/Cdh1) and SLIMs found on substrates. The KEN and D-box (RxxL) degrons are
618 well documented SLIMs that bind both APC/C-Cdc20 and APC/C-Cdh1, with APC/C-
619 Cdh1 having a preference for the KEN degron. A third SLIM called the ABBA motif was
620 shown to be important in substrate recognition by APC/C-Cdc20 [30].
621 Fig. 4C shows the enrichment profile of these SLIMs across the six clusters. The
622 KEN motif is comparably enriched in 4 out of the 5 clusters (Fig. 4C, top), with highest
623 enrichments for the mitotic phase-peaking clusters (clusters 4 and 5). The frequencies
624 range from 25% of the proteins in a cluster having the KEN motif (cluster 2) to 43%

625 (cluster 5), representing a 3-5-fold enrichment over the background frequency (8%). All
626 four clusters show low to non-detectable abundance in P16, P1 and P2, i.e. mitotic exit
627 and G0/early G1 when APC/C-Cdh1 is active. In total, 35 cell cycle regulated proteins
628 contain a KEN SLIM, approximately 50% (18 proteins) that have been experimentally
629 characterized as APC/C substrates. The remaining uncharacterized 17 proteins are
630 excellent candidates to be APC/C-Cdh1 substrates. Consistent with this prediction,
631 cluster 1, which is the only cluster showing no enrichment for the KEN motif, contains
632 proteins that have on average, higher abundance in G0/early G1.
633 Six out of 12 proteins that peak in mid-mitosis (cluster 4) contain the RxxL D-box
634 sequence. The 50% frequency is ~8-fold higher than the background frequency (6%).
635 By contrast, the fold-enrichment is considerably lower in the other clusters (Fig. 4C).
636 Similarly, 5 out of 12 proteins contain the ABBA motif (42%, Fig. 4C), representing a ~9-
637 fold enrichment over the background frequency (5%). D-box and ABBA motif-containing
638 proteins in this cluster are mostly mutually exclusive (Fig. 4D). Of the D-box and ABBA
639 motif containing proteins, two have not been previously experimentally characterized as
640 APC/C substrates: MVP and CLEC16A.
641 Cluster 4 is highly enriched in proteins containing more than one SLiM (KEN/D-
642 box/ABBA, Fig. 4C, bottom) and two proteins in this cluster contain all three SLIMs:
643 BubR1 (BUB1B) and shugoshin-2 (SGOL2). KIF20B is the only other PP that has all
644 three SLiMs and is in cluster 5. BubR1 has been demonstrated to interact with APC/C
645 through these three SLiMs and acts as a pseudosubstrate to inhibit APC/C activities in
646 spatiotemporally controlled manner [31]. It would be interesting to test the role of these
647 SLiMs in the other two proteins (SGOL2 and KIF20B). For example, SGOL2 is
648 important in protecting sister chromatid cohesion and the spindle assembly checkpoint
649 [32].
650 *Proteomic assignment of cell cycle states*
651 MS-based single cell proteome analysis is an emerging area. Recent advances in
652 miniaturized sample preparation [11],[5,9,10] suggest that routine proteome analysis of
653 single somatic mammalian cells will be possible in the near future. In comparison, single
654 cell transcriptomics as a mature field with commercial kits now available. In single cell
655 RNA-seq (sc-RNAseq) analysis [33], the deconvolution of cell cycle state has been

656 critical [34,35] This is because cell cycle variation contributes significantly to the
657 variation observed in a cell population. For example, to identify cell fate trajectories
658 during differentiation, researchers relied on reference cell cycle regulated genes in order
659 to identify the effect of cell cycle variation in the gene expression differences observed
660 [36]. A validated reference set of cell cycle regulated proteins will be important for the
661 biological interpretation of single cell proteomic datasets.

662 We tested whether the abundances of the PsPs determined in this study were
663 sufficient to assign specific cell cycle states to cellular proteomes (Fig. 5A). The
664 abundance patterns for the 119 proteins for each sample (16 timepoints x 8 replicates =
665 128 samples) were subjected to principal component analysis (PCA). The two major
666 principal components, PC1 and PC2, explain 53% and 20.5% of the variance,
667 respectively, as shown in Fig. 5B. Interphase (circles) and mitotic (triangle) are
668 separated predominantly along PC1. To a lesser extent, subphases within each (for
669 example, see arrows indicating P1 and P2) are separated along both principal
670 components. Moving counterclockwise, starting from the top right for P1, the samples
671 clearly follow a trajectory that reflects the position of each sample in the cell cycle,
672 starting from early G1 (P1 and P2) to mitosis (left side, triangles). Telophase/cytokinesis
673 populations (P16, pink triangles) are situated between the other mitotic populations and
674 P1. To ease visualization, the PCA was repeated using mean values per population
675 (Fig. 5C). Using unbiased, unsupervised methods, the PCA has arranged the
676 populations into a cell cycle 'wheel', suggesting a largely continuous process with the
677 major separation along PC1 correlated with interphase (P1 - P8) versus mitosis (P9 -
678 P16). It is less clear what is the major correlate for PC2. We note however that there is
679 a correlation with APC/C activity, with active APC/C in populations with positive values
680 along PC2 (early G1 and end of mitosis) and inactive APC/C in populations with
681 negative values (S and G2).

682 Detection of relevant features is essential as PCA analysis of the entire proteome
683 dataset does not result in cell cycle separation. Repeating the PCA analysis with cyclin
684 A2 and cyclin B1 removed essentially produces identical results, which indicates that
685 the relationships produced by using ~119 cell cycle marker proteins are robust towards
686 the absence of individual proteins, including key proteins that drive cell cycle

687 progression. This robustness will be important in assigning cell cycle states in diverse
688 datasets, as described below.
689 We then asked whether the PCA classification could be used to assign cell cycle
690 states to cellular proteomes obtained in published cell cycle fractionation and arrest
691 experiments. Human promyelocytic leukemia cells (NB4) were fractionated by
692 centrifugal elutriation into different cell cycle populations (Fig. 5A, middle) [37]. There
693 are seven fractions (F0 – F6), which correspond to asynchronous (F0), and samples
694 enriched in G1 (F1 – F2), S (F3 – F4), and G2&M (F5 and F6). In a separate
695 experiment, NB4 cells were arrested in G0-phase, S-phase, and G2 phase,
696 respectively, using serum starvation (SS), hydroxyurea (HU), and the drug RO-3306
697 (RO) [2]. Label free quantitation (LFQ) intensities were normalised to asynchronous
698 cells and these ratios were combined with mean-normalised data from this dataset prior
699 to PCA.
700 Figs. 5D and E shows the combined PCA plots for the elutriation and arrest
701 datasets, respectively. The NB4 cell populations are broadly separated according to the
702 appropriate cell cycle phase. For example, as shown in Fig. 5D, F1 and F2 are
703 positioned nearby P1 (early G1). F3 is in-between P7 and P8 (late S/G2) and F4 is near
704 P9 (late G2/early mitosis). F5 is closest to P9 whereas F6 is in-between P9 and P10
705 (late G2/early mitosis). The nearest population for F1 and F2 is P1, which is an early G1
706 population. F3 is in-between P7 and P8, which are late S/G2 populations. In Fig. 5E, the
707 SS samples are nearest the early G1 populations, P1-P4. The HU samples are in-
708 between P7 and P8, which are late S/G2 populations. The RO samples are positioned
709 near P9 – P11, which are late G2/early mitotic populations. We conclude from these
710 data that this signature can be used to classify cell cycle enriched label free proteomes.
711 We next tested if the cell cycle signature can be broadly applicable to assign cell
712 cycle state to a proteome. To do this, we made use of a large set of stable isotope
713 labelling by amino acids in cell culture (SILAC) datasets curated in proteomeHD [38].
714 Incomplete synchrony and/or cell cycle enrichment will generally lead to much poorer
715 purities compared to FACS. This lowers the resolution of classification for bulk
716 population samples, which likely contain mixtures of different phases unless purified by

717 FACS. This will not be the case for single cell proteomes, which will be by definition in a
718 single cell state.
719 To facilitate assignment of cell cycle states to partially or completely
720 asynchronous bulk populations, we first used k-means clustering to reduce the number
721 of classes from 16 populations to 8 cell cycle states (CCS) (Fig. 6A, Suppl. Table 4).
722 PCA using these 8 CCS also shows the cell cycle 'wheel' (Fig. 6B). We then mapped
723 chromatin proteomes (nascent chromatin capture, NCC and chromatin enrichment
724 proteomics, ChEP) from synchronized cells, arrested with thymidine (G1/S), 3 hr
725 thymidine release (NCC), RO-3306 (G2), or nocodazole (M) (Fig. 6B). Although these
726 samples were from a different cell type than our cell cycle signature data (HeLa vs
727 TK6), and had been processed differently (chromatin-enriched vs in-cell digest) as well
728 as quantitated differently (SILAC vs label-free), these samples group according to the
729 appropriate cell cycle phase. For example, the G2 and M-phase samples are grouped
730 between CCS6 and CCS5, which are early-to-mid mitotic states. By contrast, G1/S and
731 NCC samples are grouped with CCS2 and CCS3, which are G1/S states.
732 One challenge for the systematic classification of a heterogeneous set of
733 proteomics data are missing values, because not all of our 119 signature proteins were
734 detected in all experiments in ProteomeHD. We therefore employed Spearman rank
735 correlation to correlate the abundance of the signature proteins in these chromatin
736 proteomes with the 8 CCS. For example, the M/G1S proteome shows the highest
737 correlation with CCS6 (Figs. 6C-D), which is a mitotic state.
738 We subsequently applied this correlation approach systematically to all 294
739 experiments in ProteomeHD. We found that~15% of the experiments in ProteomeHD
740 (47 out of 294) showed a high and significant correlation with one or more CCS (Suppl.
741 Table 5). Many of these experiments involve a cell cycle perturbation, including the
742 NCC and ChEP experiments described above (Figs. 6B-D). These experiments also
743 include other types of perturbations, including differentiation, where cell cycle arrest is
744 an expected direct consequence. For example, proteomes from THP-1 monocytic cells
745 treated with phorbol myristate acetate (PMA) ester are highly correlated with G1 cell
746 cycle states. PMA treatment induces terminal differentiation of these cells and leads to

747 cessation of cell proliferation. In total, ~50% of the proteomeHD experiments highly
748 correlated with a CCS can be linked directly to cell cycle arrest.
749 From these data, we conclude that the signature robustly and accurately assigns
750 CCS across far ranging experimental contexts, cell types and quantitation strategies.
751 The remaining experiments with high correlation have less obvious links to cell
752 cycle. For example, Jurkat T cells are treated with the HSP90 inhibitor geldanamycin for
753 either 6 h or 20 h (Fig. 6E). Proteomes from 6 h treatment are highly correlated with
754 CCS1 (early G1). By contrast, proteomes from 20 h treatment are highly correlated with
755 the G2/mitotic states, CCS4 and CCS6. Geldanamycin has been reported to arrest cells
756 in G1 or G2 phases of the cell cycle. Interestingly, flow cytometry analysis of cells
757 treated with geldanamycin for 20 h shows an accumulation of 4N DNA content cells,
758 corresponding to G2&M phase cells [39].
759 We also detect significant CCS signatures in experiments that have no apparent
760 link to cell cycle arrest, direct or indirect. In a study comparing untransformed breast
761 epithelial cells with breast cancer cell lines, three untransformed breast epithelial lines,
762 MCF10A, HMT-3522 and HMEC1, showed significant correlation with one or more
763 CCS. Cell lines were compared using a super-SILAC approach against MCF7, which is
764 an ER+ breast cancer line. Both HMT-3522 and HMEC1 show strong correlation with
765 early G1 states (CCS1). By contrast, MCF10A were correlated with S-phase (CCS4).
766 Interestingly, ER+ MDA-MB-453 cultures also showed correlation with CCS1. These
767 data suggest that the cell cycle distributions of these cell cultures are shifted compared
768 to MCF7. In a separate study, 16 out of 62 lymphoblastoid cell lines (LCLs) analysed by
769 proteomics to identify quantitative trait loci (QTL) were significantly CCS correlated (Fig.
770 6E). Interestingly, they were correlated in different states: 12 correlated with CCS1
771 and/or CCS2 (G1 phase) and the remaining four correlated with CCS5 (G2/early
772 mitosis). These data suggest there is significant heterogeneity in cell cycle distribution,
773 impacting at least 25% of the LCLs compared. How much of the heterogeneity in cell
774 cycle state correlation observed has a genetic basis or is due to technical variation in
775 cell culture handling will be important to assess.

776 **Discussion**

777 A major challenge with the comprehensive analysis of proteomes from low cell number
778 samples is sample preparation. An on-column load of 200 ng peptide, the equivalent to
779 the protein content of approximately 2,000 TK6 cells, is sufficient material to obtain
780 proteome coverage of >4,000 proteins with current instrumentation. Removal of
781 detergents used to produce soluble cell extracts by use of membrane filters [40],
782 organic precipitation (with or without the aid of magnetic beads) [41][42] or SDS-PAGE
783 gel extraction are protocols involving several steps and repeated exposure to new
784 plastic surfaces that introduce opportunities for non-specific peptide and protein
785 adsorption. Here, we have presented a minimalistic approach for preparing cells for
786 proteomics called the 'in-cell digest'. Cells are fixed with formaldehyde and methanol to
787 effectively trap them in biochemical states, then directly digested with trypsin and
788 desalted prior to LC-MS/MS analysis.

789 We show that the in-cell digest enables reproducible and quantitative analysis of
790 proteomes from 2,000 TK6 and MCF10A cells using AMPL analysis. The AMPL
791 approach overcomes the low duty cycle of the Orbitrap Elite to enable proteome
792 analysis with a sensitivity comparable with current instruments. Newer instrumentation
793 with higher duty cycles, including the TIMS-TOF Pro and Exploris 480, is expected to
794 enable conventional DDA and DIA analyses of proteomes at a similar depth with 2,000
795 TK6 cells, or alternatively, improve proteome depth further using MS1-based matching
796 methods.

797 The in-cell digest is compatible with other approaches of low cell number sample
798 preparation for MS-based proteomics. In-cell digested samples can be efficiently
799 labelled by isobaric tags, e.g. TMT and iTRAQ, and therefore compatible with use of
800 carrier channels to boost the signal of rare or single cell channels (e.g. iBASIL [43]). The
801 protocol requires no specialized humidified sample handling chambers or direct loading
802 onto premade analytical nanoLC columns, such as those described in the nanoPOTS
803 workflow [11]. While the proteome coverages obtained by nanoPOTS is higher than
804 reported here, it is possible that a new workflow combining aspects of the in-cell digest
805 and nanoPOTS could improve both generalizability and performance compared to either
806 method.

807 Each sample preparation method will have its unique advantages and potential
808 biases, which we evaluated by quantitatively comparing the in-cell digest with a more
809 conventional in-solution digest. This analysis revealed an overrepresentation of
810 membrane proteins amongst those proteins with higher abundance measured in the in-
811 cell digest samples. These proteins include mitochondrial membrane proteins (e.g.
812 TOMM7) and proteins that are known to be localized to the cell surface (ADAM15).
813 Membrane proteins have been shown to irreversibly aggregate in soluble extracts when
814 heat-treated and precipitated. Delipidation by methanol, which is used to increase cell
815 permeability, could also play an important role in increasing digestion efficiency of
816 membrane proteins by trypsin. We suggest that the higher abundances measured for
817 membrane proteins is unlikely to be an artefact of the in-cell digest; in contrast, the
818 measurements are likely to more accurately reflect the abundances of these proteins in
819 cells.
820 Feature matching FDR is controlled in our approach by implementing stringent
821 cutoffs for retention time difference, m/z difference and match m/z error. Using a
822 chemically modified 'decoy' proteome, we demonstrate that these cutoffs reduce the
823 false positive rate with minimal impact on true positives. Elution time filtering provided
824 greater discrimination between true and false positives than mass accuracy, suggesting
825 that further improvements in chromatographic precision will benefit FDR control. We
826 detect a higher estimated FDR compared to previous published models using mixed
827 species [44]. However, our analysis differs in two significant aspects: 1) unlike matching
828 between individual ‘single shot’ analyses, our experimental approach assesses match
829 FDR from a fractionated library to a single shot analysis, and 2) unlike a mixed species
830 proteome, our decoy proteome lacks true positives that could prevent assignment to
831 false positive features. The latter means our reported FDR is likely an overestimate, but
832 does provide a metric for assessing the relative FDR when filtering on feature match
833 parameters. Additionally, models based on mixed-species suggest that matching FDR
834 increases at low sample loads. It will be important in future to assess this with AMPL. In
835 this study, comparable on-column loads between FDR estimation and cell cycle
836 analysis and therefore, we are confident in the performance of false positive removal in
837 the cell cycle dataset.

838 We identify novel proteins whose cell cycle function has not been previously
839 characterized. FAM111B is a pseudoperiodic protein in cluster 1 (Fisher's $p_1 < 0.001$, p_2
840 $= 0.06$), showing peak levels in S-phase populations, followed by a decrease in G2
841 populations. FAM111B is poorly characterized despite its expression being associated
842 with poor prognosis in pancreatic and liver cancers (Human Protein Atlas [45]) and
843 mutation causative for a rare inherited genetic syndrome (hereditary fibrosing
844 poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis). FAM111A,
845 the only other member of the FAM111 gene family, localizes to newly synthesized
846 chromatin during S-phase, interacts with PCNA via its PCNA-interacting protein (PIP)
847 box and its depletion reduces base incorporation during DNA replication [46]. FAM111B
848 also contains a PIP box (residues 607 – 616). Data from HeLa S3 cells also suggest
849 that FAM111B is a cell cycle regulated protein with peak levels in S-phase [47].
850 FAM111B contains D-box and KEN-box motifs that are recognized by the APC/C E3
851 ligase to ubiquitinate targets for proteasomal degradation. Due to the similarity with
852 FAM111A in sequence, predicted interactions with PCNA and peak protein abundance
853 in S-phase, we propose that FAM111B also is likely to play a key role in DNA
854 replication.
855 We present an unbiased pseudotemporal analysis of protein abundance changes
856 across 8 biochemically resolved mitotic states with a resolution extremely challenging to
857 obtain with high precision using arrest and release methodologies. The frequency of
858 PsPs identified (1.7%, 119 / 6,899) compares well with a recent antibody-based screen
859 for cell cycle regulated proteins (2.6%, 331 / 12,390) [48]. Included in 331 hits are
860 proteins that vary in subcellular localization but not abundance across the cell cycle,
861 consistent with other datasets using biochemical fractionation [49]. PsPs identified in
862 this study will be limited to proteins that change in abundance. However, these PsPs
863 are critical for robust cell state classification of proteomes obtained by MS, most of
864 which do not involve subcellular fractionation.
865 A high proportion of proteins in clusters 4 and 5 (24/69, 35%) are experimentally
866 validated APC/C substrates, which represents a 70-fold overrepresentation in these two
867 clusters compared to non-pseudoperiodic proteins (0.5%). The high mitotic phase
868 resolution and purity obtained in this study enabled characterisation of protein

869 abundance variation of APC/C substrates in mitosis. We identify two waves of mitotic
870 degradation, one coinciding with the destruction of cyclins A and B (cluster 4) and the
871 second at mitotic exit (cluster 5). The unbiased clustering failed to separate cyclin A and
872 cyclin B, which are degraded in prometaphase and at the metaphase-to-anaphase
873 transition, respectively. This can be explained by the relatively few proteins detected
874 that correlate with cyclin A and is consistent with the idea that prometaphase
875 degradation by the APC/C is highly selective. 44 proteins in clusters 4 and 5 have not
876 been previously experimentally validated as APC/C substrates [29] and are candidates
877 for future follow-up analysis as novel, uncharacterized substrates. These include
878 proteins (e.g. PRC1, KIF23, KIF20A) that were not identified as APC/C-Cdh1 and
879 APC/C-Cdc20 substrates by bioinformatic analysis of co-regulation [50] and by chemical
880 biology approaches [51,52].
881 High resolution classification of cell cycle state is an important prerequisite to
882 obtaining meaningful biological insights into single cell 'omics' data. However, datasets
883 on the cell cycle regulated transcriptome and proteome generally provide low time
884 resolution, particularly in mitosis. Mitotic time resolution will be crucial for interpreting
885 single cell proteomes. Whereas transcriptional and translational activity are dampened
886 during mitosis, there are major changes in protein phosphorylation and protein
887 abundance, which will contribute towards single cell proteome variation.
888 Here we have identified a robust cell cycle signature composed of the
889 abundances from 119 PsPs that can be used to classify the cell cycle state of a cell
890 population by virtue of its cellular proteome. We apply this signature to assign cell cycle
891 states to hundreds of published proteomic datasets that range in cell type and
892 experimental condition. We have not tested if this signature can be used to assign
893 proteomes from species other than human. We note that many of these proteins are
894 well conserved, with several conserved to yeast (e.g. cyclin, REC8, Aurora kinase, Polo
895 kinase). We anticipate that this high-resolution cell cycle signature here will be important
896 to understand the biological implications of emerging single cell proteomics datasets
897 [9,10], particularly in systems where cell cycle phase differences are an underlying
898 source of variation, as is frequently the case.

899 Formaldehyde fixation is used frequently as a precursor to intracellular
900 immunostaining for cellular analysis and for inactivating cells that potentially harbor
901 infectious agents, e.g. viruses. We have shown that mild formaldehyde treatment is
902 compatible with comprehensive and quantitative proteomics with low cell numbers. We
903 anticipate that the in-cell digest will be broadly applicable to characterise the proteomes
904 of formaldehyde fixed, virally-infected cells. Recently published data suggest that
905 formaldehyde crosslinks can be directly detected from MS data [53]. We anticipate the
906 in-cell digest would enhance the sensitivity of crosslink detection and lead to an
907 increase in identified protein-protein interactions.

908 **Acknowledgements**

909 This work was supported by a Sir Henry Dale Fellowship to TL (Wellcome Trust & Royal
910 Society 206211/Z/17/Z), a Darwin Trust PhD Studentship to ASA, an EASTBIO PhD
911 Studentship to DAL, an MRC Career Development Award to GK, the Wellcome Centre
912 for Cell Biology (WCB) core facilities (Wellcome Trust 203149), and funding for
913 instrumentation, including equipment grants to the WCB Proteomics Core (091020) and
914 the flow facilities. We thank valuable feedback and discussions with colleagues in the
915 WCB and the University of Edinburgh, including Fiona Rossi (Scottish Centre for
916 Regenerative Medicine), Christos Spanos (WCB) and Shaun Webb (WCB).

917 **Data Availability**

918 Raw MS data and processed MaxQuant output files are available on
919 ProteomeXchange/PRIDE. These data can be accessed using the project accession
920 number PXD020006. Username: reviewer45620@ebi.ac.uk Password: AUSLTsvu.

921

922 **Figures and Figure Legends**

Figure 1

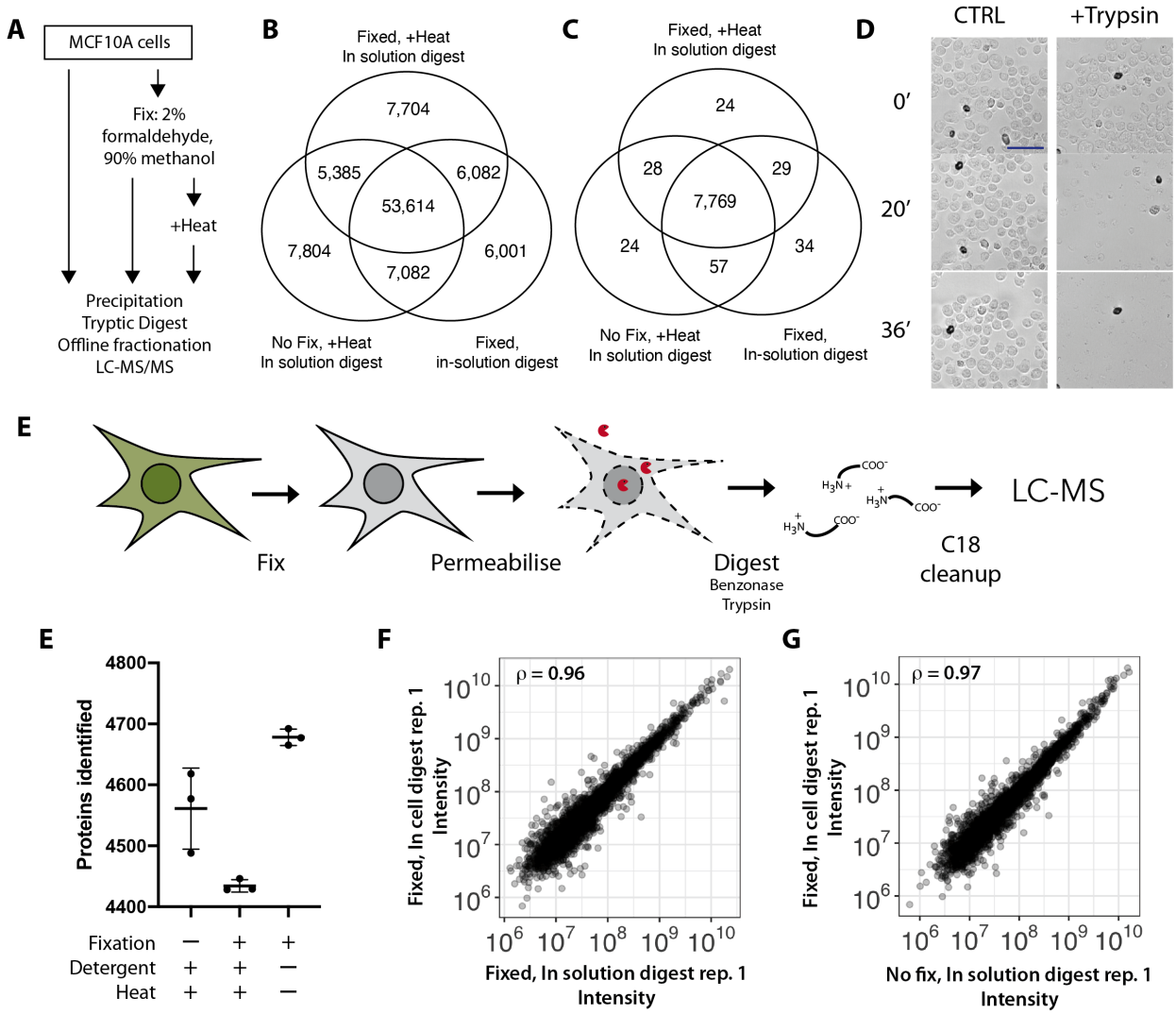


Figure 1. A simplified workflow for proteomics of formaldehyde fixed cells using in cell digestion. A) Proteomics workflow to assess impact of formaldehyde-induced modifications on proteome analysis. B, C) The number of peptides (B) and proteins (C) identified. D) Fixed and permeabilized cells treated either with DPBS (left) or with trypsin (right) were imaged at the indicated times in minutes. Scale bar is 50 μm. E) Schematic of the in-cell digest workflow. F-H) Comparison of the identification (F) and quantitative reproducibility (G, H) between in-solution and in-cell digests.

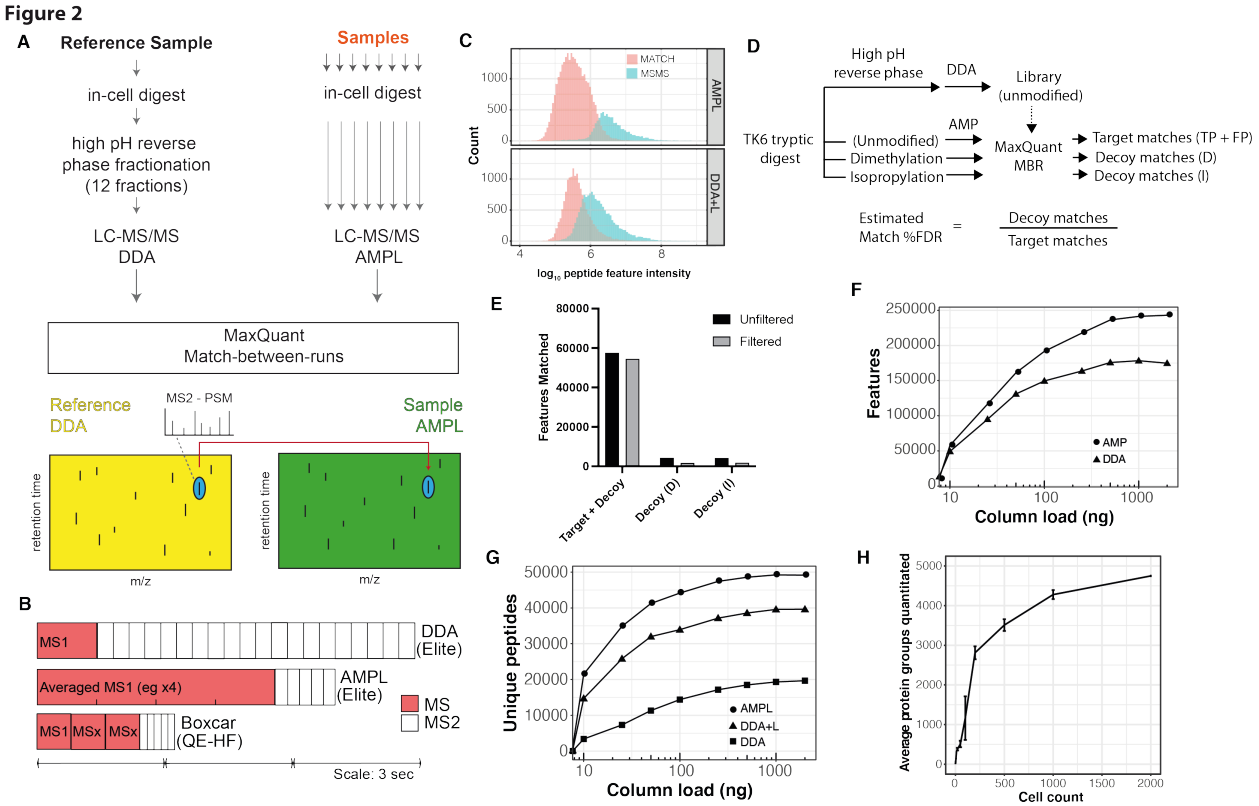


Figure 2. Averaged MS1 precursors with library matching (AMPL) increases peptide detection sensitivity. A) Schematic outlining the AMPL experimental design. B) Both the AMPL and BoxCar acquisition methods prioritise MS time to enhance MS1 scan quality. Schematic comparing duty cycles for DDA, AMPL and Boxcar acquisition methods on the indicated MS instruments (Orbitrap Elite, Orbitrap HF). The median peak width using our chromatographic setup with the Orbitrap Elite is ~38 seconds. C) A comparison between AMPL and DDA+L, showing intensity distributions of peptide features identified by MS/MS (blue) and matching to identified library features (red). D) Schematic outlining experimental workflow for assessing match-between-runs FDR. E) Features matched in target and decoy proteomes before and after additional filtering based on match retention time difference, match m/z difference and match m/z error. F, G) Features (F) and unique peptides (G) detected in AMP(L) vs DDA. DDA+L is DDA with matching to a library. H) Proteome coverage versus cell number. The cell titration was performed in duplicate.

Figure 3

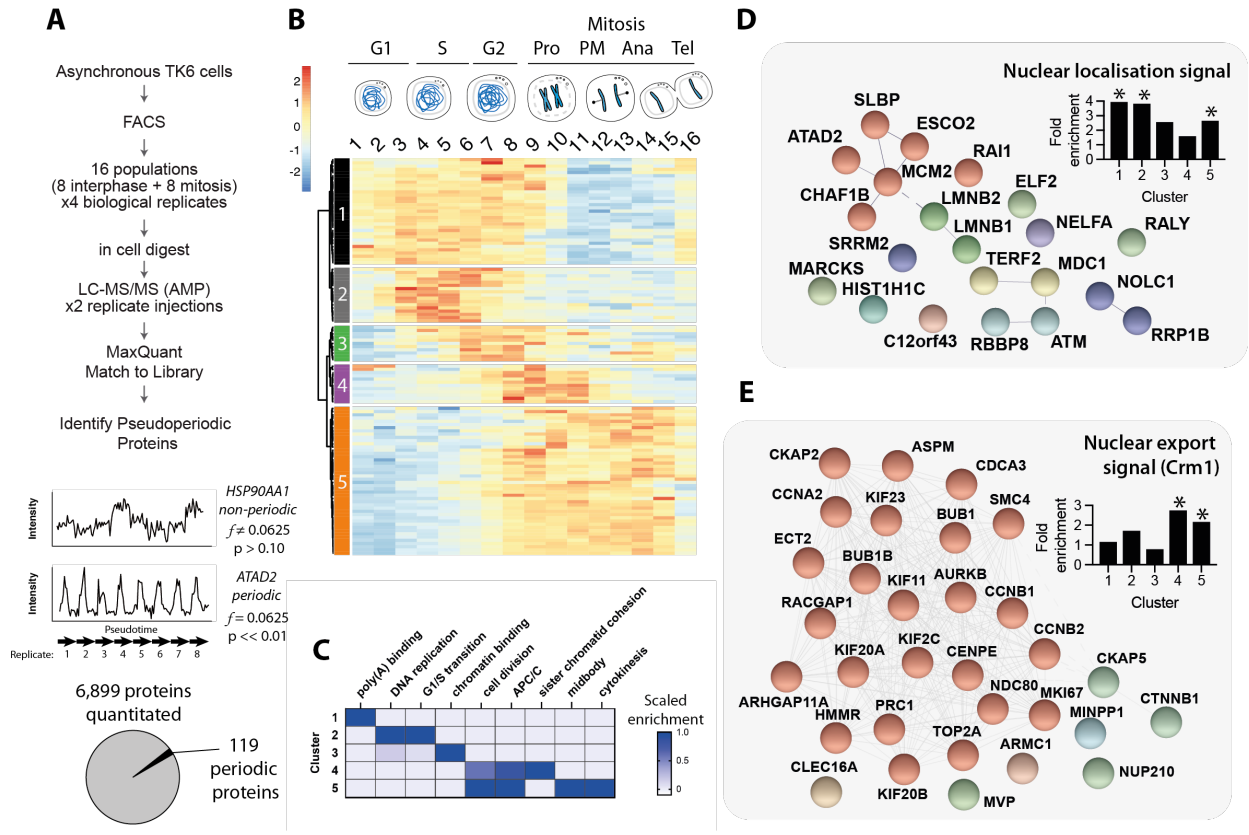


Figure 3. High resolution proteomic analysis of an unperturbed cell cycle. A) Schematic describing the experimental design and workflow. Details for protein abundance normalization and pseudoperiodicity analysis can be found in the Experimental methods section. B) Heatmap of the 119 identified pseudoperiodic proteins (PsPs) organized by cluster. Schematic above heatmap shows indicative cell cycle stages isolated by FACS (full gating strategy shown in Suppl. Fig. 3). C) Enriched UniProt keywords by cluster. D) Proteins in clusters 1 and 2 containing a putative nuclear localization signal (NLS). E) Proteins in clusters 4 and 5 containing a putative nuclear export signal (NES). * indicate $p < 0.01$ (Fisher's exact test).

Figure 4

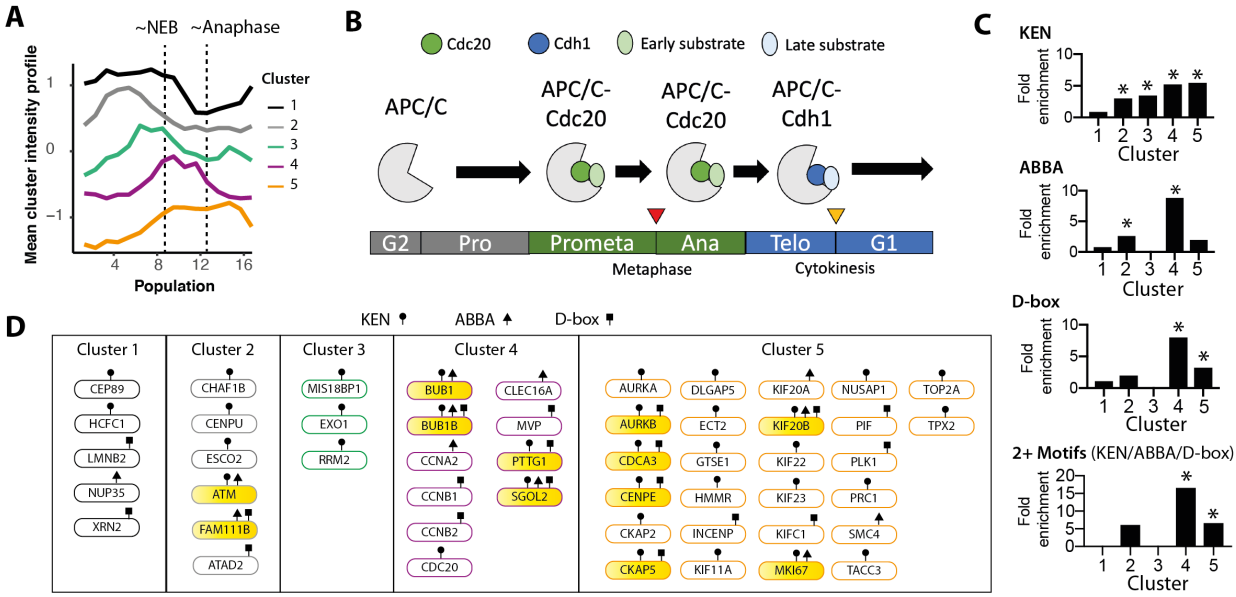


Figure 4. Characterisation of potential APC/C substrates. A) Mean intensity profile for five clusters shown in Fig. 3B. B) Schematic illustrating regulation of APC/C substrate choice switch during mitosis by co-activators Cdc20 and Cdh1. C) Enrichment analysis of SLiMs that control interaction with the APC/C. D) Proteins with at least one APC/C SLiM grouped by cluster. The yellow fill indicates proteins that contain 2 or more SLiMs.

Figure 5

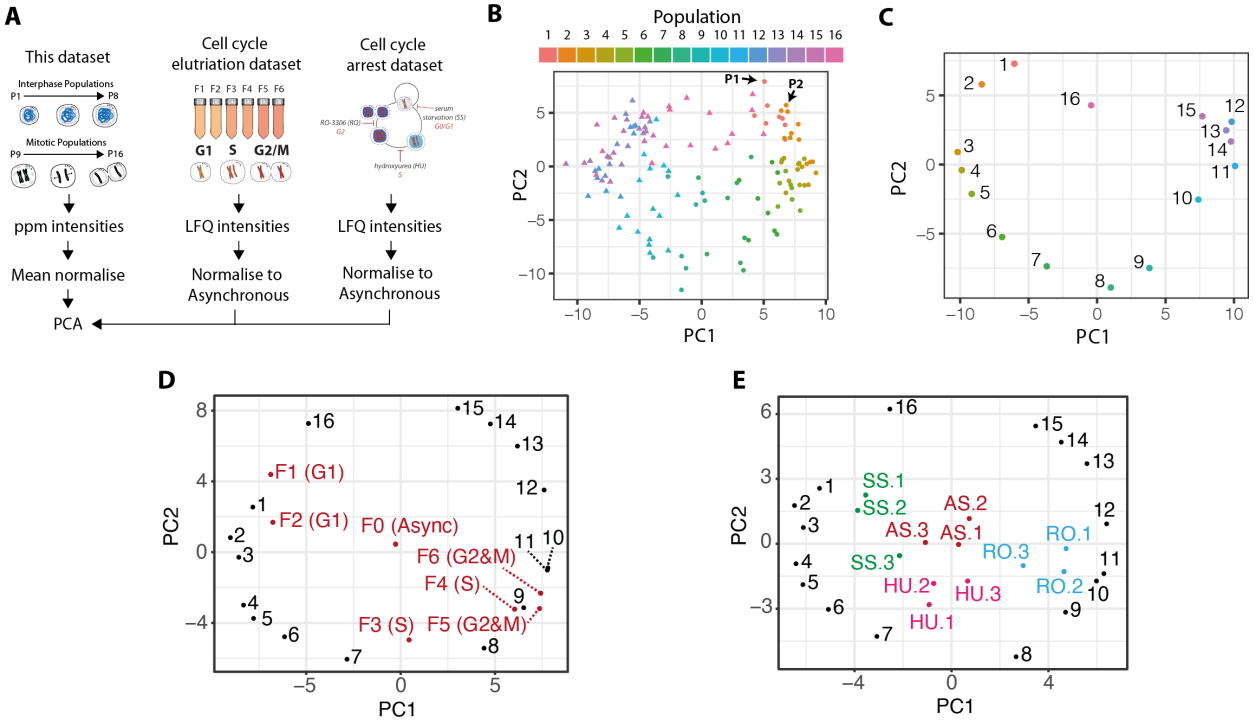
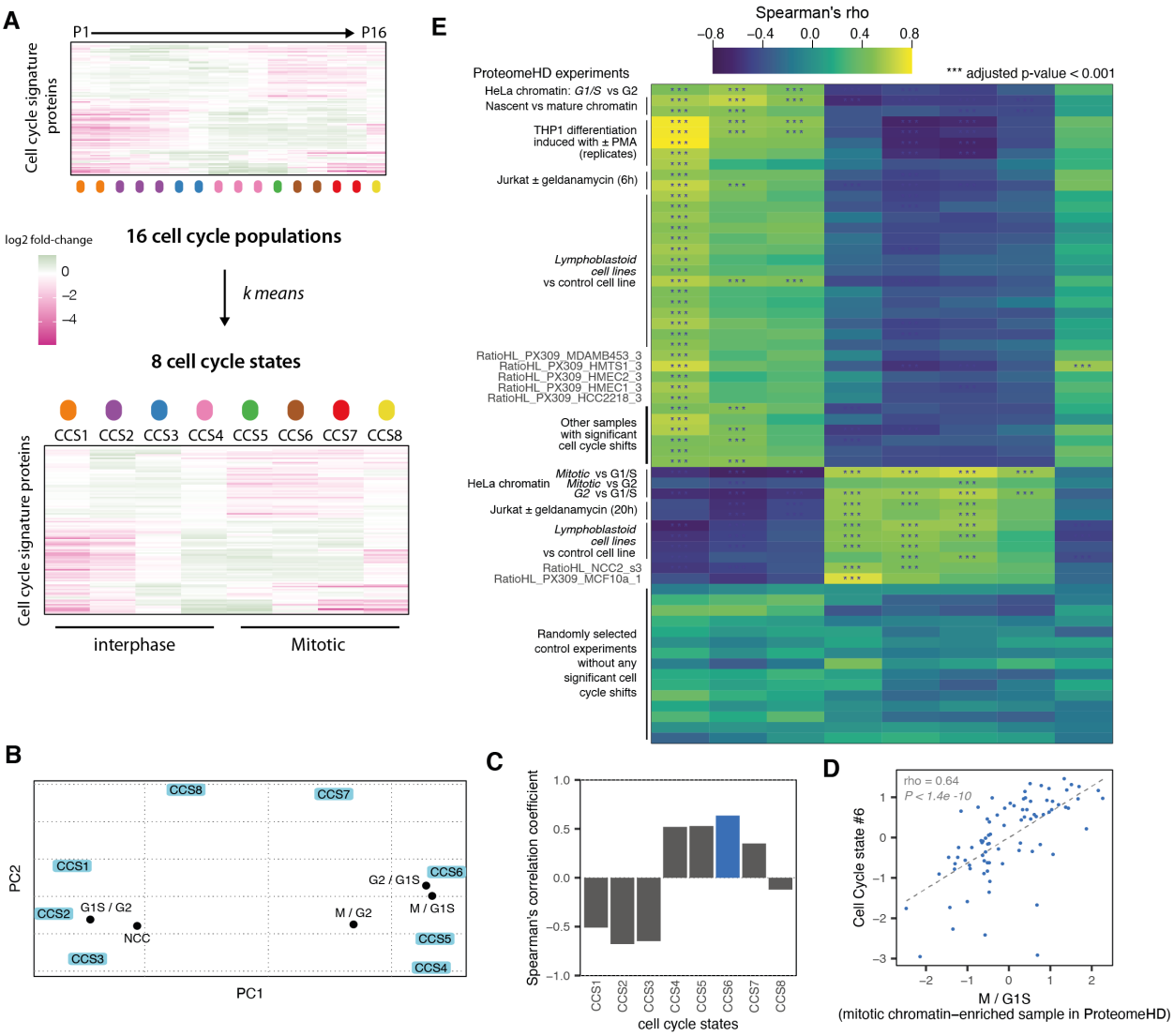


Figure 5. The 119 PsPs form a signature that classifies cell cycle states. A) Schematic illustrating the approach to compare the signature across datasets. B, C) PCA of the 16 cell cycle populations in this study using the 119 PsPs as features using either individual replicates (B), or mean abundances (C). D, E) PCA as in (C) with samples from elutriated (D) or from cell cycle arrest (E) label free datasets using NB4 cells.

Figure 6



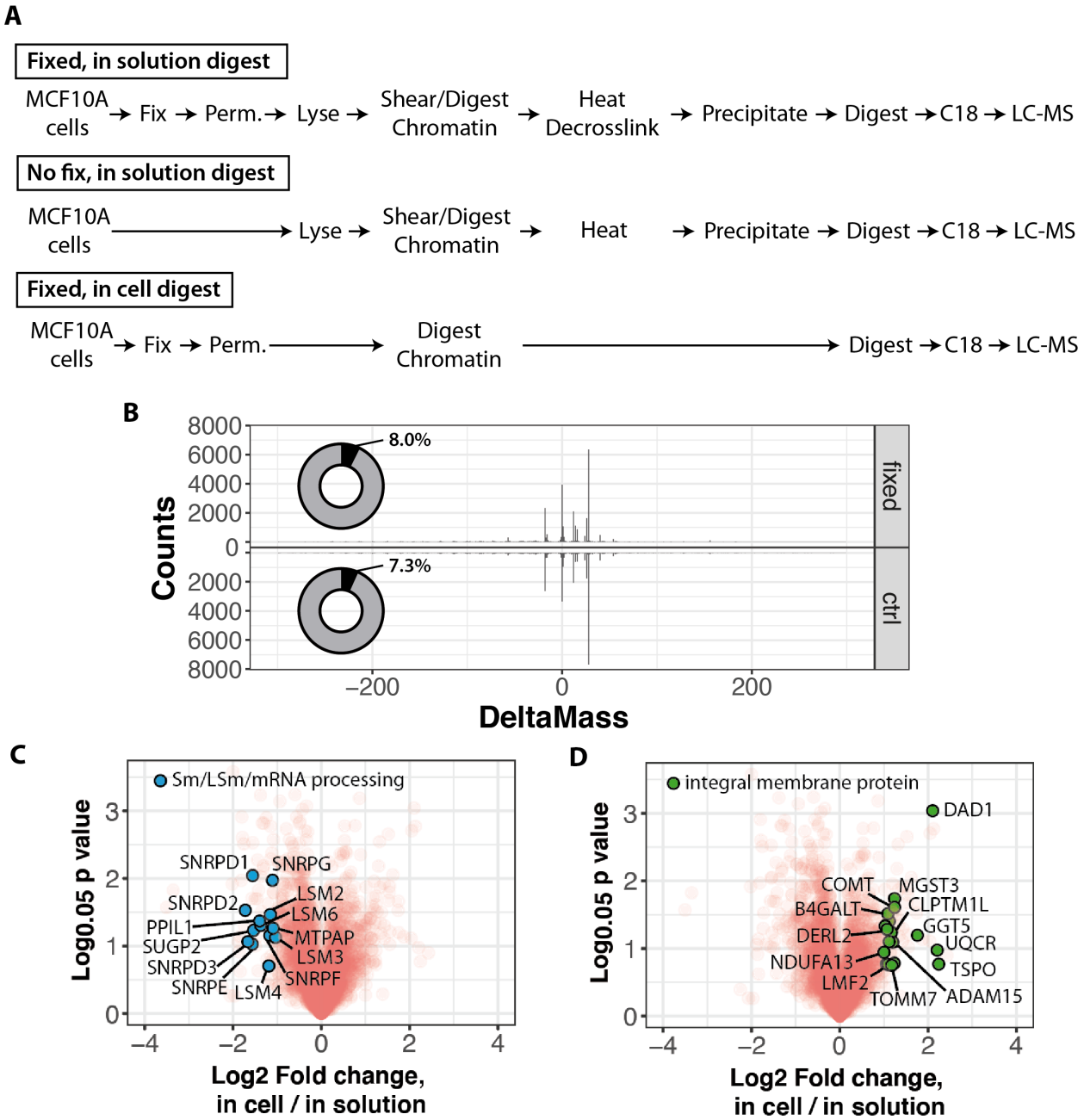
977

978 **Figure 6. Unbiased assignment of cell cycle states across hundreds of proteomes**
979 **in ProteomeHD.** A) Schematic of how the 16 cell cycle populations were aggregated
980 into 8 cell cycle states (CCS). B) Assignment of chromatin proteomes from nascent
981 chromatin capture (NCC) and chromatin enrichment proteomics (ChEP) SILAC
982 experiments. NCC samples were enriched in S-phase. G1S, G2 and M samples were
983 treated with thymidine, RO-3306, and nocodazole, respectively. C) Spearman rank
984 correlation coefficient for the 8 cell cycle states for M / G1S. This sample has the
985 highest correlation with CCS6 (blue fill). D) Correlation scatterplot between CCS6 and M
986 / G1S. E) Heatmap showing Spearman correlation coefficients for 47 experiments (out

987 of 294) that show an enriched CCS and 15 randomly selected experiments that have no
988 CCS enrichment.
989
990 **Supplementary Tables**
991 Supplementary Table 1. Error tolerant search results
992 Supplementary Table 2. Raw MaxQuant proteomic output for the 16 cell cycle
993 populations
994 Supplementary Table 3. Quantitative and pseudoperiodic analysis of the 16 cell cycle
995 proteomes.
996 Supplementary Table 4. Pseudoperiodic proteins (PsPs) identified in this dataset.
997 Supplementary Table 5. Short linear motif (SLiM) enrichment analysis.
998 Supplementary Table 6. 8-state cell cycle signature.
999 Supplementary Table 7. Application of cell cycle signature to ProteomeHD.
1000
1001 **Supplementary Data**
1002 Supplementary Data 1. R scripts used for the evidence FDR filtering, pseudoperiodicity
1003 analysis and proteomeHD analysis
1004

1005 **Supplementary Figures**

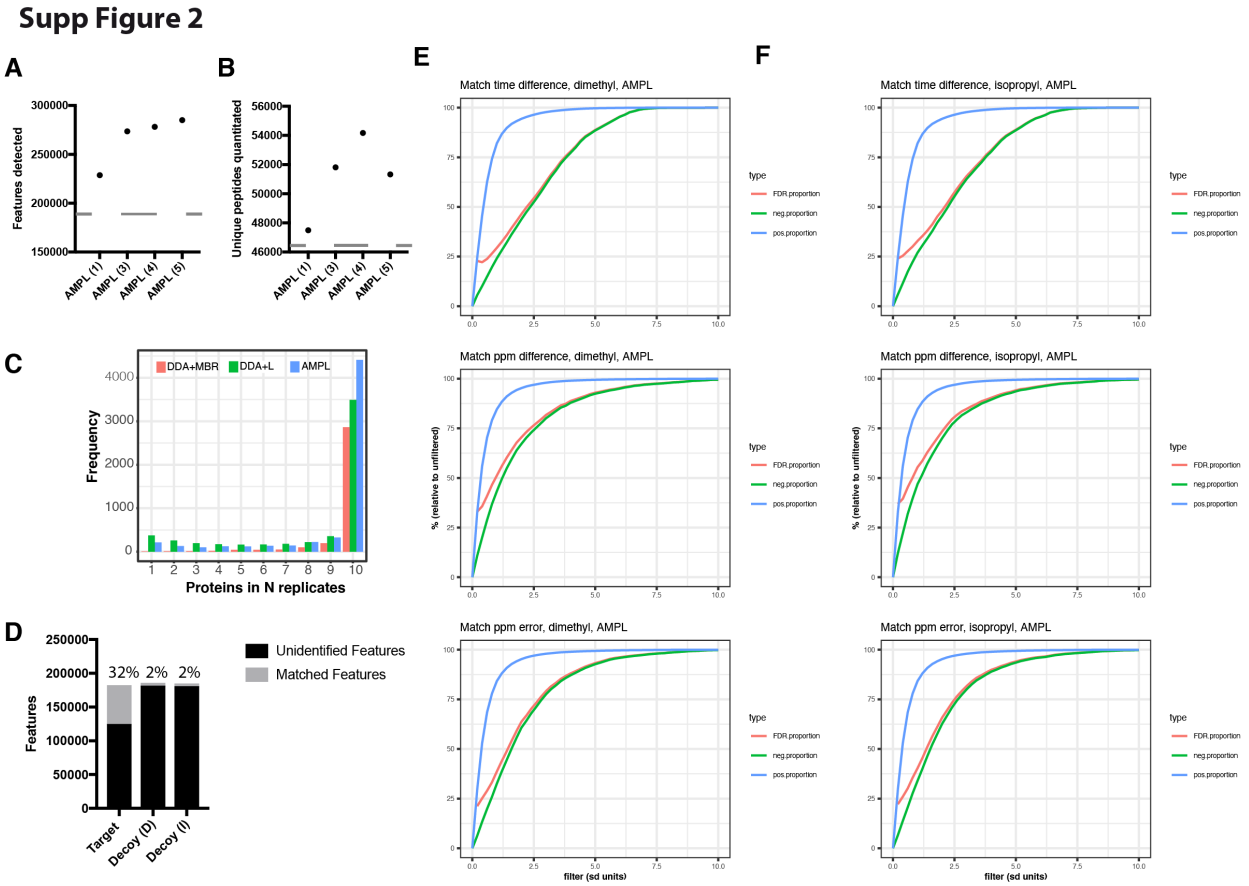
Supp Figure 1



1006

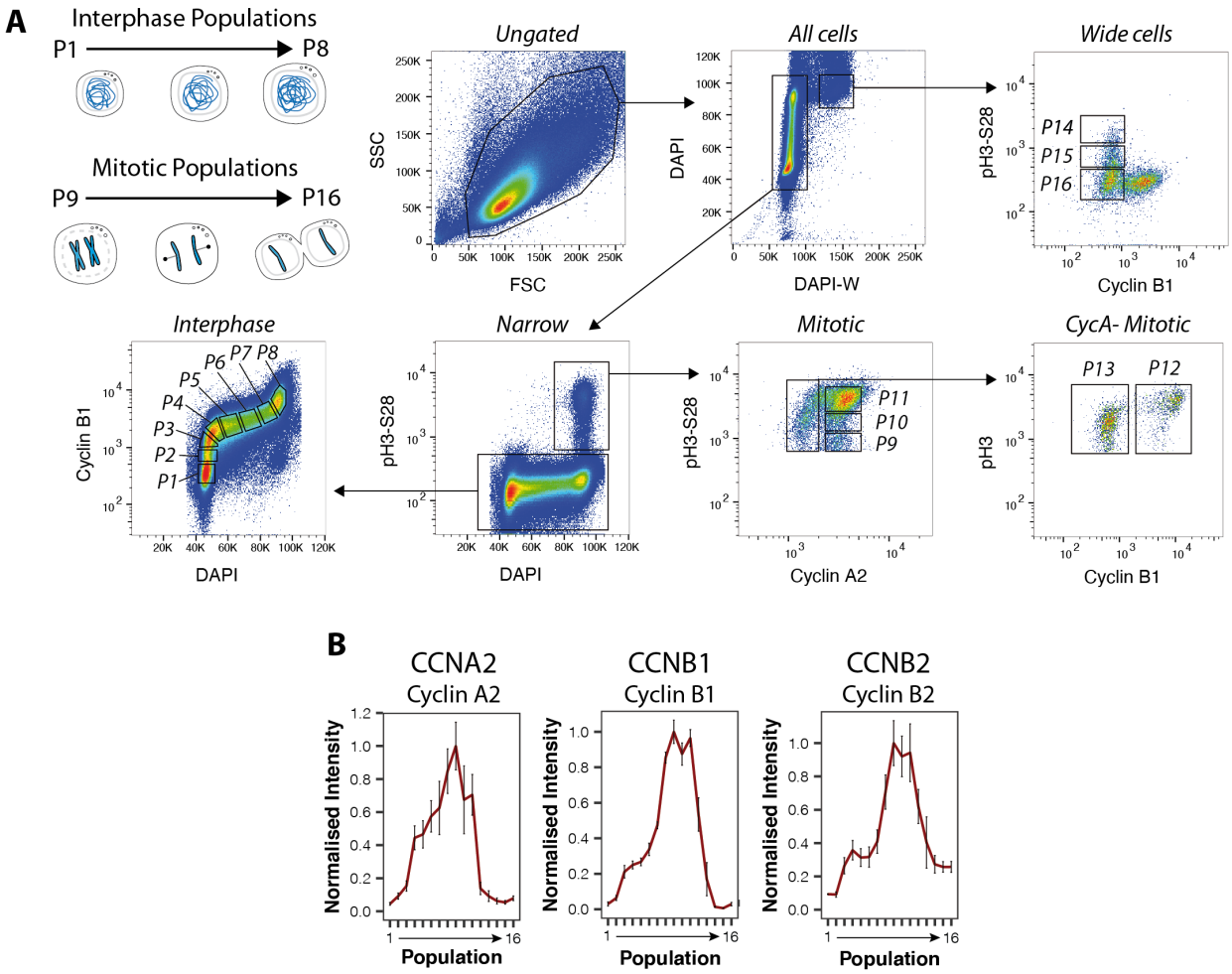
1007 **Supplementary Figure 1.** A) Scheme in Fig. 1A expanded to include details. B) Results
1008 from an error-tolerant dependent peptide search (MaxQuant) comparing control (cells
1009 without fixative) and fixed cells. C,D) Volcano plots comparing triplicate in cell and in
1010 solution digests.

1011



Supplementary Figure 2. A) Identified features in target and decoy proteomes, including proportion of features matched to library. B, C) Optimising the number of MS1 spectra to average by measuring impact on the number of features and peptides detected. D) An analysis of data completeness comparing DDA, DDA with a library (DDA+L) and MPL. E, F) ROC plots for the dimethyl (E) and isopropyl (F) decoy proteomes filtering against match time difference, match ppm difference and match ppm error. Proportions are given in %s. FDR proportion is the estimated FDR at indicated filtering threshold over FDR with no filter.

Supp Fig. 3



Supplementary Figure 3. A) Scheme of FACS isolation and pseudocolour plots showing the gating strategy to isolate the 16 cell cycle populations. B) Mean normalised intensities across the 16 cell cycle populations as measured by MS for cyclin A2, cyclin B1 and cyclin B2.

References

1. Ly T, Whigham A, Clarke R, Brenes-Murillo AJ, Estes B, Madhessian D, Lundberg E, Wadsworth P, Lamond AI (2017) Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *eLife* **6**: e27574.
2. Ly T, Endo A, Lamond AI (2015) Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* **4**:.

1036 3. Ly T, Whigham A, Clarke R, Brenes-Murillo AJ, Estes B, Wadsworth P, Lamond
1037 Al (2017) Proteomic analysis of cell cycle progression in asynchronous cultures,
1038 including mitotic subphases, using PRIMMUS. *bioRxiv* 125831.
1039 4. Kelly RT (2020) Single-cell Proteomics: Progress and Prospects. *Mol Cell*
1040 *Proteomics* **19**: 1739–1748.
1041 5. Specht H, Emmott E, Petelski AA, Huffman RG, Perlman DH, Serra M,
1042 Kharchenko P, Koller A, Slavov N (2021) Single-cell proteomic and transcriptomic
1043 analysis of macrophage heterogeneity using SCoPE2. *Genome Biol* **22**: 50–27.
1044 6. Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample
1045 preparation method for proteome analysis. *Nat Meth* **6**: 359–362.
1046 7. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J (2014)
1047 Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular*
1048 *Systems Biology* **10**: 757–757.
1049 8. Zhang Z, Dubiak KM, Huber PW, Dovichi NJ (2020) Miniaturized Filter-Aided
1050 Sample Preparation (MICRO-FASP) Method for High Throughput, Ultrasensitive
1051 Proteomics Sample Preparation Reveals Proteome Asymmetry in Xenopus laevis
1052 Embryos. *Anal Chem* **92**: 5554–5560.
1053 9. Brunner A-D, Thielert M, Vasilopoulou CG, Ammar C, Coscia F, Mund A,
1054 Hoerning OB, Bache N, Apalategui A, Lubeck M, et al. (2021) Ultra-high
1055 sensitivity mass spectrometry quantifies single-cell proteome changes upon
1056 perturbation. *bioRxiv* 2020.12.22.423933.
1057 10. Hartlmayr D, Ctortecka C, Seth A, Mendjan S, Tourniaire G, Mechtler K (2021) An
1058 automated workflow for label-free and multiplexed single cell proteomics sample
1059 preparation at unprecedented sensitivity. *bioRxiv* 2021.04.14.439828.
1060 11. Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, Shukla AK, Petyuk
1061 VA, Campbell-Thompson M, Mathews CE, et al. (2018) Nanodroplet processing
1062 platform for deep and quantitative proteome profiling of 10–100 mammalian cells.
1063 *Nat Comms* **9**: 882.
1064 12. Metz B, Kersten GFA, Hoogerhout P, Brugghe HF, Timmermans HAM, de Jong
1065 A, Meiring H, Hove ten J, Hennink WE, Crommelin DJA, et al. (2004) Identification
1066 of Formaldehyde-induced Modifications in Proteins REACTIONS WITH MODEL
1067 PEPTIDES. *Journal of Biological Chemistry* **279**: 6235–6243.
1068 13. Toews J, Rogalski JC, Clark TJ, Kast J (2008) Mass spectrometric identification of
1069 formaldehyde-induced peptide modifications under in vivo protein cross-linking
1070 conditions. *Anal Chim Acta* **618**: 168–183.
1071 14. Skopek TR, Liber HL, Penman BW, Thilly WG (1978) Isolation of a human
1072 lymphoblastoid line heterozygous at the thymidine kinase locus: Possibility for a
1073 rapid human cell mutation assay. *Biochemical and Biophysical Research*
1074 *Communications* **84**: 411–416.
1075 15. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates,
1076 individualized p.p.b.-range mass accuracies and proteome-wide protein
1077 quantification. *Nat Biotechnol* **26**: 1367–1372.
1078 16. Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS
1079 (2016) Rapid immunoprecipitation mass spectrometry of endogenous proteins
1080 (RIME) for analysis of chromatin complexes. *Nat Protoc* **11**: 316–326.

1081 17. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI (2017)
1082 MSFragger: ultrafast and comprehensive peptide identification in mass
1083 spectrometry–based proteomics. *Nat Meth* **14**: 513–520.
1084 18. Pasa-Tolic L, Masselon C, Barry RC, Shen Y, Smith RD (2004) Proteomic
1085 analyses using an accurate mass and time tag strategy. *BioTechniques* **37**: 621–
1086 639.
1087 19. Meier F, Geyer PE, Winter SV, Cox J, Mann M (2018) BoxCar acquisition method
1088 enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat*
1089 *Meth* **15**: 440–448.
1090 20. Pines J, Clute P (1999) Temporal and spatial control of cyclin B1 destruction in
1091 metaphase. *Nat Cell Biol* **1**: 82–87.
1092 21. Elzen den N, Pines J (2001) Cyclin a Is Destroyed in Prometaphase and Can
1093 Delay Chromosome Alignment and Anaphase. *The Journal of Cell Biology* **153**:
1094 121–136.
1095 22. Koo SJ, Fernández-Montalván AE, Badock V, Ott CJ, Holton SJ, Ahsen von O,
1096 Toedling J, Vittori S, Bradner JE, Gorjánácz M (2016) ATAD2 is an epigenetic
1097 reader of newly synthesized histone marks during DNA replication. *Oncotarget* **7**:
1098 70323–70335.
1099 23. Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J,
1100 Diakogianni A, Valverde JA, Bukirova D, Čalyševa J, et al. (2019) ELM—the
1101 eukaryotic linear motif resource in 2020. *Nucleic Acids Res* **48**: D296–D306.
1102 24. Gonçalves V, Jordan P (2015) Posttranscriptional Regulation of Splicing Factor
1103 SRSF1 and Its Role in Cancer Cell Biology. *Biomed Res Int* **2015**: 287048.
1104 25. Breig O, Baklouti F (2013) Proteasome-mediated proteolysis of SRSF5 splicing
1105 factor intriguingly co-occurs with SRSF5 mRNA upregulation during late erythroid
1106 differentiation. *PLoS ONE* **8**: e59137.
1107 26. Maréchal A, Zou L (2013) DNA damage sensing by the ATM and ATR kinases.
1108 *Cold Spring Harb Perspect Biol* **5**: a012716.
1109 27. Kernan J, Bonacci T, Emanuele MJ (2018) Who guards the guardian?
1110 Mechanisms that restrain APC/C during the cell cycle. *Biochimica et Biophysica*
1111 *Acta (BBA) - Molecular Cell Research* **1865**: 1924–1933.
1112 28. Chang L, Barford D (2014) Insights into the anaphase-promoting complex: a
1113 molecular machine that regulates mitosis. *Current Opinion in Structural Biology*
1114 **29**: 1–9.
1115 29. Davey NE, Morgan DO (2016) Building a Regulatory Network with Short Linear
1116 Sequence Motifs: Lessons from the Degrons of the Anaphase-Promoting
1117 Complex. *Molecular Cell* **64**: 12–23.
1118 30. Di Fiore B, Davey NE, Hagting A, Izawa D, Mansfeld J, Gibson TJ, Pines J (2015)
1119 The ABBA Motif Binds APC/C Activators and Is Shared by APC/C Substrates and
1120 Regulators. *Developmental Cell* **32**: 358–372.
1121 31. Lara-Gonzalez P, Scott MIF, Diez M, Sen O, Taylor SS (2011) BubR1 blocks
1122 substrate recruitment to the APC/C in a KEN-box-dependent manner. *Journal of*
1123 *Cell Science* **124**: 4332–4345.
1124 32. Hellmuth S, Gómez-H L, Pendás AM, Stemmann O (2020) Securin-independent
1125 regulation of separase by checkpoint-induced shugoshin-MAD2. *Nature* **580**:
1126 536–541.

1127 33. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical
1128 challenges in single-cell transcriptomics. *Nat Rev Genet* **16**: 133–145.
1129 34. Scialdone A, Natarajan KN, Saraiva LR, Proserpio V, Teichmann SA, Stegle O,
1130 Marioni JC, Buettner F (2015) Computational assignment of cell-cycle stage from
1131 single-cell transcriptome data. *Human Pluripotent Stem Cells* **85**: 54–61.
1132 35. Liu Z, Lou H, Xie K, Wang H, Chen N, Aparicio OM, Zhang MQ, Jiang R, Chen T
1133 (2017) Reconstructing cell cycle pseudo time-series via single-cell transcriptome
1134 data. *Nat Comms* **8**: 22.
1135 36. Cuomo ASE, Seaton DD, McCarthy DJ, Martinez I, Bonder MJ, Garcia-Bernardo
1136 J, Amatya S, Madrigal P, Isaacson A, Buettner F, et al. (2020) Single-cell RNA-
1137 sequencing of differentiating iPS cells reveals dynamic genetic effects on gene
1138 expression. *Nat Comms* **11**: 810.
1139 37. Ly T, Ahmad Y, Shlien A, Soroka D, Mills A, Emanuele MJ, Stratton MR, Lamond
1140 AI (2014) A proteomic chronology of gene expression through the cell cycle in
1141 human myeloid leukemia cells. *eLife* **3**: e01630.
1142 38. Kustatscher G, Grabowski P, Schrader TA, Passmore JB, Schrader M, Rappsilber
1143 J (2019) Co-regulation map of the human proteome enables identification of
1144 protein functions. *Nat Biotechnol* **37**: 1361–1371.
1145 39. Fierro-Monti I, Echeverria P, Racle J, Hernandez C, Picard D, Quadroni M (2013)
1146 Dynamic impacts of the inhibition of the molecular chaperone Hsp90 on the T-cell
1147 proteome have implications for anti-cancer therapy. *PLoS ONE* **8**: e80425.
1148 40. Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal
1149 sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–
1150 362. http://doi.org/10.1038/nmeth.1322
1151 41. Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., &
1152 Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead
1153 technology. *Molecular Systems Biology*, *10*(10), 757–757.
1154 http://doi.org/10.15252/msb.20145625
1155 42. Batth, T. S., Tollenaere, M. X., Rüther, P., Gonzalez-Franquesa, A., Prabhakar, B.
1156 S., Bekker-Jensen, S., et al. (2019). Protein Aggregation Capture on
1157 Microparticles Enables Multipurpose Proteomics Sample Preparation. *Molecular &*
1158 *Cellular Proteomics : MCP*, *18*(5), 1027–1035.
1159 http://doi.org/10.1074/mcp.TIR118.001270
1160 43. Yu, F., Haynes, S. E., & Nesvizhskii, A. I. (2021). IonQuant Enables Accurate and
1161 Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs.
1162 *Molecular & Cellular Proteomics : MCP*, *20*, 100077.
1163 http://doi.org/10.1016/j.mcpro.2021.100077
1164 44. Tsai, C.-F., Zhao, R., Williams, S. M., Moore, R. J., Schultz, K., Chrisler, W. B., et
1165 al. (2020). An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL)
1166 Strategy for Precise Quantitative Single-cell Proteomics. *Molecular & Cellular*
1167 *Proteomics : MCP*, *19*(5), 828.
1168 45. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif
1169 M, Liu Z, Edfors F, et al. (2017) A pathology atlas of the human cancer
1170 transcriptome. *Science* **357**: eaan2507.
1171 46. Alabert C, Bukowski-Wills J-C, Lee S-B, Kustatscher G, Nakamura K, de Lima
1172 Alves F, Menard P, Mejlvang J, Rappsilber J, Groth A (2014) Nascent chromatin

1173 capture proteomics determines chromatin dynamics during DNA replication and
1174 identifies unknown fork components. *Nat Cell Biol* **16**: 281–291.
1175 47. Aviner R, Shenoy A, Elroy-Stein O, Geiger T (2015) Uncovering Hidden Layers of
1176 Cell Cycle Regulation through Integrative Multi-omic Analysis. *PLoS Genet* **11**:
1177 e1005554.
1178 48. Mahdessian D, Cesnik AJ, Gnann C, Danielsson F, Stenström L, Arif M, Zhang C,
1179 Le T, Johansson F, Shutten R, et al. (2021) Spatiotemporal dissection of the cell
1180 cycle with single-cell proteogenomics. *Nature* **590**: 649–654.
1181 49. Herr P, Boström J, Rullman E, Rudd SG, Vesterlund M, Lehtiö J, Helleday T,
1182 Maddalo G, Altun M (2020) Cell cycle profiling reveals protein oscillation,
1183 phosphorylation, and localization dynamics. *Mol Cell Proteomics*
1184 mcp.RA120.001938.
1185 50. Franks JL, Martinez-Chacin RC, Wang X, Tiedemann RL, Bonacci T, Choudhury
1186 R, Bolhuis DL, Enrico TP, Mouery RD, Damrauer JS, et al. (2020) In silico APC/C
1187 substrate discovery reveals cell cycle-dependent degradation of UHRF1 and other
1188 chromatin regulators. *PLoS Biol* **18**: e3000975.
1189 51. Bakos G, Yu L, Gak IA, Roumeliotis TI, Liakopoulos D, Choudhary JS, Mansfeld J
1190 (2018) An E2-ubiquitin thioester-driven approach to identify substrates modified
1191 with ubiquitin and ubiquitin-like molecules. *Nat Comms* **9**: 4776–15.
1192 52. Manohar S, Yu Q, Gygi SP, King RW (2020) The Insulin Receptor Adaptor IRS2
1193 is an APC/C Substrate That Promotes Cell Cycle Protein Expression and a
1194 Robust Spindle Assembly Checkpoint. *Mol Cell Proteomics* **19**: 1450–1467.
1195 53. Tayri-Wilk T, Slavin M, Zamel J, Blass A, Cohen S, Motzik A, Sun X, Shalev DE,
1196 Ram O, Kalisman N (2020) Mass spectrometry reveals the chemistry of
1197 formaldehyde cross-linking in structured proteins. *Nat Comms* **11**: 3128.
1198