1 **Increasing the Throughput and Reproducibility of Activity-Based Proteome**

2 **Profiling Studies with Hyperplexing and Intelligent Data Acquisition**

3

4 **Authors: Hanna G. Budayeva[1*], Taylur P. Ma[1], Shuai Wang[2], Meena Choi[1], Christopher M.**

5 **Rose[1*]**

6 **Affiliations: [1] Department of Microchemistry, Proteomics and Lipidomics, Genentech, Inc.**

7 **South San Francisco, CA 94044, [2]Department of Metabolism and Pharmacokinetics,**

8 **Genentech, Inc. South San Francisco, CA 94080**

9 ***Correspondence: Budayeva.Hanna@gene.com, Rose.Christopher@gene.com**

10 ## Abstract

11 Intelligent data acquisition (IDA) strategies, such as real-time database search (RTS), have

12 improved the depth of proteome coverage for experiments that utilize isobaric labels and

13 gas phase purification techniques (i.e., SPS-MS3). While most applications of IDA have been

14 focused on the analysis of protein abundance, these approaches have recently been applied to

15 activity-based proteome profiling (ABPP) studies aimed at characterization of protein site

16 engagement by small molecules. In this work, we extend IDA capabilities offered by vendor

17 software through a program called InSeqAPI. First, we demonstrate robust performance of

18 InSeqAPI in the analysis of biotinylated cysteine peptides from ABPP experiments. Then, we

19 describe PairQuant, a method within InSeqAPI designed for the hyperplexing approach that

20 utilizes protein-level isotopic labeling and peptide-level TMT labeling. PairQuant allows for TMT

21 analysis of 36 conditions in a single sample and achieves ~98% coverage of both peptide pair

22 partners in a hyperplexed experiment as well as a 40% improvement in the number of quantified

23   cysteine sites compared to non-RTS acquisition. We applied this method in ABPP study of

24   ligandable cysteine sites in the nucleus leading to an identification of additional druggable sites

25   on protein- and DNA-interaction domains of transcription regulators and on nuclear ubiquitin

26   ligases.

27

28   Introduction

29   Isobaric mass tags are widely utilized in proteomics research for relative quantification of protein

30   levels and post-translational modifications. For example, tandem mass tag (TMT) reagents

31   allow for analysis of up to 18 conditions in a single mass spectrometry (MS) run.[1] Gas phase

32   purification through synchronous precursor selection MS3 (SPS-MS3) data acquisition improves

33   the accuracy of isobaric label based quantitation, but can decrease proteome depth due to a

34   longer duty cycle as compared to Orbitrap MS2 analysis alone.[2] To improve the depth of SPS-

35   MS3 experiments, a real-time database search (RTS) has been introduced to enable online

36   identification of peptides from ion trap MS2 spectra and limit collection of SPS-MS3 scans to

37   instances where a peptide is confidently identified.[3,4] This approach increases the amount of

38   instrument duty cycle dedicated to collecting quantitative data associated with identified

39   peptides and results in an increased number of proteins quantified by SPS-MS3.[3,4]

40   Recently, RTS-aided acquisition of TMT data was applied in activity-based proteome profiling

41   (ABPP) studies for characterization of accessible and reactive cysteine residues.[5] Cysteine is

42   the most reactive nucleophilic residue in the proteome and cysteine residues are present in the

43   active sites of many enzyme families (such as kinases BTK[6] and EGFR[7]) and near ligandable

44   pockets of "undruggable" proteins (such as KRAS G12C[8]). Their identification and

45   characterization may be used in the discovery of new covalent drug targets (reviewed in[9]).

46   There are approximately 214,000 cysteine sites encoded in the human genome. Combined

47   results from a few recent cysteine profiling studies revealed ~10,000 ligandable cysteine sites

48   against ~300 electrophilic compounds of various reactivities.[5,10,11]

49   While the number of available TMT reagents continues to increase, currently analyzing more

50   than 18 samples (including various cell types, treatments, and replicates) requires performing

51   multiple TMT experiments and combining data through one or more common bridge channels.[12]

52   Due to the stochastic sampling of data dependent acquisition, comparing quantitative data

53   across multiple TMT experiments leads to an increase of missing data as peptides and proteins

54   are not quantified in all multiplexed experiments.[13] This is particularly true for site-level post

55   translational modification (PTM) data which requires peptides containing the PTM to be

56   identified across multiple TMT experiments.[14]

57   It was previously demonstrated that TMT tags can be combined with stable isotope labeling in

58   cell culture (SILAC) in a "hyperplexing" approach to assess temporal abundance of yeast

59   proteins or to profile site-resolved protein turnover.[15,16] Hyperplexing increases the number of

60   samples that can be quantified in a single experiment, but also increases sample complexity

61   potentially decreasing the total number of unique peptides that are quantified. Hyperplexed data

62   acquisition also utilizes data-dependent acquisition and it is not guaranteed that both peptides

63   within a hyperplexed pair will be chosen for identification and quantification. To date, these

64   challenges have limited the broad adoption of hyperplexing in quantitative proteomics

65   experiments.

66   In this study, we present an approach that combines protein-level isotopic labeling (by SILAC or

67   by isotopic cysteine-reactive probes) and peptide-level quantification by TMT to perform

68   hyperplexed ABPP experiments. First, we introduce inSeqAPI - an instrument application

69   interface (iAPI) program that enables creation of custom intelligent data acquisition (IDA)

70   methods, including those that utilize a real-time database search. To address the challenges of

71   hyperplexed experiments we describe a novel IDA method called PairQuant that utilizes a real-

72   time peptide identification to improve the quantitation of both peptides within hyperplexed pairs.

73   Lastly, we utilize PairQuant to perform a 36-plex cysteine reactivity profiling experiment in

74   isolated nuclei treated with a set of electrophilic fragment compounds to investigate ligandability

75   of nuclear targets.

76

77

78

79   **Materials and Methods**

80   **Cell lines**

81   C2C12 mouse myoblasts were obtained from ATCC and maintained in Dulbecco's Modified

82   Eagle Medium (DMEM) supplemented with 10% FBS and Glutamax. MCF7 adenocarcinoma

83   cell line was purchased from ATCC and maintained in RPMI-1640, 10% FBS and Glutamax.

84   For SILAC experiments, C2C12 cells lines were grown in DMEM SILAC "light" ($[^{12}C,^{14}N]$ Lys and

85    $[^{12}C,^{14}N]$ Arg) or SILAC "heavy" ($[^{13}C,^{15}N]$ Lys and $[^{13}C,^{15}N]$ Arg) media for five passages, until

86   >95% incorporation was achieved.

87   **Reagents**

88   Light and heavy desthiobiotin iodoacetamide (LDBIA and HDBIA) reagents were synthesized by

89   Pharmaron and stored at 4.92mg/mL and 5mg/mL in DMSO, respectively. Structures of the

90   DBIA reagents are provided in **Supp. Figure 1**. IodoTMTsixplex^{TM} reagent (Thermo Fisher

91   Scientific) was reconstituted in methanol prior to the experiment, according to the

92   manufacturer's protocol. Electrophilic compound fragments were synthesized in-house (G35,

93   G59, G84, G92), purchased from Sigma (KB05), or Enamine (G51, KB02, KB03) and stored in

94   DMSO at 50mM concentration.

95   **Compound reactivity profiling (cysteine trapping assay)**

96   Incubation was carried out in a 100 mM phosphate buffer (pH 7.4) at 37°C with gentle shaking.

97   Test compounds (1 μM), diphenhydramine (0.1 μM, pre-internal standard), and cysteine (5 mM)

98   were mixed sequentially to start the reaction. Aliquots were taken from 0, 30, 60, 90, 120, 150,

99   and 180 minutes and quenched with 1x volume of ice-cold acetonitrile containing propranolol (2

100   μM, post-internal standard). The resulting samples were quantified via LC-MS/MS. Percentage

101   remaining was calculated using peak area ratios normalized to the 0 minutes time-point sample.

102   Half-life and k_{react} were then determined from fitting the data with one phase decay model

103   ($\% \, remaining = 100 * e^{-kreact * 5 \, mM * time}$). Incubation without cysteine was also carried out to

104   check compound stability in the buffer.

105   **Sample processing for activity-based proteome profiling workflow**

106   For benchmarking an RTS against a non-RTS method in ABPP analysis with DBIA probes,

107   C2C12 mouse myoblasts were collected from cell culture by scraping, washed with PBS and

108   frozen until later or immediately lysed. Lysis was performed on ice in PBS/1% NP-40 buffer and

109   lysates were pulse-sonicated with a probe. Insoluble fraction was removed by 4°C centrifugation

110   at 12000xg for 10 minutes and protein concentration in the soluble fraction was measured by

111   Pierce^{TM} BCA protein assay kit (Thermo Fisher Scientific). Protein concentration was adjusted

112   to 1-2 mg/mL for LDBIA treatment.

113   For whole-cell proteome profiling with electrophilic fragments, lysates were treated with

114   indicated compounds at 500µM concentration for 1 hr at room temperature. Unreacted cysteine

115   sites were labeled with 500µM LDBIA reactive probe for 1hr at room temperature. Upon

116   labeling, protein mixtures were reduced with 5mM Pierce^{TM} DTT (Thermo Fisher Scientific), 30

117   minutes at room temperature and alkylated with 20mM Pierce^{TM} iodoacetamide (Thermo Fisher

118   Scientific) for 30 minutes in the dark at room temperature. Protein mixtures were cleaned up by

119   chloroform-methanol precipitation. In brief, four volumes of methanol, one volume of chloroform

120   and three volumes of water were added to the mixture and centrifuged for 2 minutes at

121   20,000xg. Top layer was removed and the remaining protein disk was washed twice with

122   methanol. After final wash, all liquid was removed and the protein pellet was air dried for 15

123   minutes.

124   Protein pellets were redissolved in 20mM HEPES, pH 8.0 and digested for 4 h at 37°C with lysC

125   (1:50 enzyme:protein ratio), followed by overnight digestion at 37°C with trypsin (1:100 ratio).

126   TMTPro^{TM} (Thermo Fisher Scientific) labeling was performed as described in the manufacturer's

127   protocol. TMT tag incorporation of >95% and relative protein abundances between channels

128   were checked by MS prior to quenching with 5% hydroxylamine and mixing at 1:1 ratio.

129   Streptavidin enrichment was performed on AssayMap Bravo (Agilent) automation using affinity

130   enrichment protocol. Briefly, lyophilized TMT-labeled peptide mixtures were resuspended in

131   PBS. Streptavidin cartridges (5uL capacity) were equilibrated with PBS and loaded with the

132   samples. Upon loading, cartridges were washed sequentially with PBS, 0.1% PBS and HPLC

133   water. Biotinylated peptides were eluted with 50% acetonitrile/0.1% TFA. Eluents were dried

134   down prior and desalted by C18 PhyTip columns (Biotage) using PhyNexus MEA robot

135   (Biotage) prior to LC-MS analysis.

136   For iodoTMT experiment, MCF cell pellets were lysed with HENS buffer (Thermo Scientific) by

137   pipette mixing before sonicating in a cold water bath, followed by 10 minute incubation on ice.

138   Protein lysate obtained by high speed centrifugation, 14,000xg at 4°C for 15 minutes, and

139   separation of the protein solute from the pelleted cellular debris was collected. Protein

140   concentration was brought to 1 mg/mL and reduced with a final concentration of 10 mM TCEP

141   for 1 hour at 50°C. IodoTMT reagent was prepared as per manufacturer's protocol (Thermo

142   Scientific) and labeling was carried out for 1 hour at 37°C in the dark. Labeling was quenched

143   with a final concentration of 20 mM DTT (Sigma), incubating at 37°C in the dark for 15 minutes.

144   Samples labeled with iodoTMT were combined after quenching. Protein cleanup was performed

145   via a methanol chloroform precipitation as described above. Subsequently, the pellet was

146   dissolved with 2 M UREA (in 20 mM HEPES), sonicated before adding trypsin (Promega) at a

147   1:60 ratio of enzyme/substrate and digested overnight at 37°C. Following digestion, peptides

148   were acidified with 1% TFA before desalting with a 3 cc C18 Sep-Pak columns (Waters, Milford,

149   MA) eluting with 60% ACN/0.1% TFA. Samples were lyophilized in preparation for enrichment.

150   For affinity enrichment of iodoTMT-labeled peptides, anti-TMT resin slurry (Prod#90076,

151   ThermoFisher Scientific) was washed 3 times with TBS. Peptides were then re-suspended with

152   TBS before addition to the rinsed resin. Enrichment was performed by incubating for 2 hours at

153   room temperature with end-over-end mixing. The supernatant was removed and the resin

154   washed 4 times with 2 M UREA in TBS, followed by 4 times with just TBS, and 3 times with

155   water. Peptides were eluted off the beads with 50% ACN/0.5% TFA after incubation for 10

156   minutes with shaking and elution was performed twice. All centrifugation steps were performed

157   at 2,500 xg for 1 minute at room temperature. After drying to completion, peptides were cleaned

158   up using a 5 μl C18 Phytip (Biotage, Inc) and injected for LC-MS/MS acquisition.

159   **Mass Spectrometry analysis for ABPP workflow**

160   For DBIA labeling experiment, desalted samples were reconstituted in 2% acetonitrile/0.1%

161   formic acid (buffer A) at ~1ug/uL peptide concentration for liquid chromatography-tandem mass

162   spectrometry (LC-MS) analysis on Dionex Ultimate 3000 RSLCnano system (Thermo Fisher

163   Scientific, Inc) and Orbitrap Eclipse Tribrid MS (Thermo Fisher Scientific, Inc). Peptides (~1ug)

164   were resolved on 25 cm x 75 μm Aurora column packed with 1.6 μm C18 (Ion Opticks) by 135

165   minutes linear gradient of 4% to 30% buffer B (98% acetonitrile/0.1% formic acid) in buffer A

166   (2% acetonitrile/0.1% formic acid) at a flow rate of 300 nL/min.

167   Where indicated, standard SPS-MS3/Multi-Notch-MS3 TMT method[2] was performed with the

168   following parameters. Each duty cycle included an MS1 scan in the Orbitrap at 120,000

169   resolution across 350-1350 m/z range with normalized automatic gain control (AGC) target of

170   250%, 50 ms maximum injection time. Data dependent ion trap MS2 scans were performed on

171   the top 10 peptides with CID activation, 0.5 m/z isolation window in quadrupole, turbo scan rate,

172   150% normalized AGC target, 35 ms maximum injection time. Eight MS2 fragment ions were

173   selected for Orbitrap SPS-MS3 scans with isolation widths of 2 m/z using isolation waveforms

174   with multiple frequency notches. These ions were fragmented during MS3 by high energy

175   collision-induced dissociation (HCD) and analyzed by Orbitrap at 50,000 resolution, 400%

176   normalized AGC target, 150 ms maximum injection time.

177   For LDBIA-based experiments benchmarking inSeqAPI method against Thermo RTS and

178   standard SPS-MS3 method, MS1 and MS2 scans were acquired with the same settings as

179   standard SPS-MS3 TMT method. For Thermo RTS and inSeqAPI RTS analysis of mouse cell

180   lines, the following parameters were used: Uniprot mouse database October 2022 version,

181   including 25440 entries of common contaminants, Swissprot sequences of canonical and

182   protein isoforms, plus decoys. Static modifications included Cys carbamidomethylation

183   (+57.0215), K and n-term TMTZero™Pro (+295.1896); variable modifications included Met

184   oxidation (+15.9949), Cys with light desthiobiotin label (+239.1633). Other parameters were set

185 to 0 maximum missed cleavages, 3 variable modification per peptide, FDR filtering enabled, 35

186 ms maximum search time. Comet scoring threshold: 0.5 Xcorr, 0.05 dCn, 10 ppm precursor

187 mass error, +2 charge state. InSeqAPI modification node allowed for inclusion of only

188 biotinylated peptide for MS3 analysis. MS3 HCD settings were as in standard SPS-MS3 mode,

189 except 600% normalized AGC target and 400 ms maximum injection time.

190 For iodoTMT sample analysis, data was collected on an Orbitrap Eclipse Tribrid (Thermo Fisher

191 Scientific, San Jose, CA USA). Peptides loaded onto a IonOptick Aurora GEN3 C18 column (1.7

192 μM, 75 μM x 25 cm) via a Nanospray Flex Ion-Source (Thermo Scientific) at a voltage of 1.6 kV

193 with a flow rate of 0.4 μL/min in Solvent A (98% water/2% acetonitrile/0.1% formic acid).

194 Separation was achieved at a flow rate of 0.3 μL/min with a linear gradient of 5% solvent B

195 (98% acetonitrile/2% water/0.1% formic acid) to 35% solvent B over 158 minutes with a total run

196 time of 180 minutes. For SPSMS^3 acquisition, full ms scans were collected in the orbitrap at

197 120,000 resolution, scan ranging from 350 to 1600 m/z, automatic gain targeting (AGC) of 1 x

198 10^6, and a maximum injection time of 50 ms. MS^2 ions were selected using a top speed data

199 dependent mode and fragmented with CID energy of 35, AGC of 1.5 x 10^4, and a maximum

200 injection time of 100 ms. Utilizing the Orbitrap Fusion Tribrid's SPS mode for isolating MS^3 ions,

201 the top 8 MS^2 precursor ions are selected and fragmented via higher collision dissociation

202 energy (HCD) of 55, AGC of 7.5 x 10^4, a maximum injection time of 350 ms, isolation width of

203 1.2 Da, and a resolution of 50,000 at 200 m/z. For ThermoRTS acquisition all parameters

204 matched that of the SPSMS3 method with the exception of an increased AGC 1 x 10^5 target for

205 MS3 ions and the addition of Thermo's real-time search module enabled. Real time search

206 parameters include a human proteome with a decoy addendum from the October 2022 version

207 database for tryptic peptides (42374 total entries). Variable modifications of methionine

208 oxidation (+15.9949 Da), and variable modifications of cysteine (+329.2266) for IodoTMT. Other

209 parameters include, allowance of 1 missed cleavage, 1 maximum variable mod/peptide,

210 enabling FDR filtering, TMT SPS MS3 Mode and a maximum search time of 35 ms. Scoring

211 Thresholds include 0.5 Xcorr, 0.05 dCN, 10 ppm of the precursor and a 2+ charge state.

212 InSeqAPI method acquisition functions by enabling Xcalibur method triggering of MS1 and MS2

213 levels of which parameters match that of the SPSMS3 method indicated above, whilst all MS3

214 specific parameters are controlled via the API.

215 **Sample processing for activity-based proteome profiling workflow with**

216 **hyperplexing**

217

218 For optimization of RTS acquisition of hyperplexed samples, two alternative workflows were

219 tested. In HLDBIA/TMT hyperplexing workflow, MCF7 lysates at 2 mg/mL protein concentration

220 were labeled with 500 µM LDBIA or HDBIA reagent, as described in regular ABPP workflow

221 above. For hyperplexed samples with SILAC labeling, "heavy" SILAC or "light" SILAC protein

222 mixtures from C2C12 lysate at 2 mg/mL concentrations were labeled with 500 µM LDBIA. For

223 both types of hyperplexed samples, "heavy" and "light" protein mixtures were reduced,

224 alkylated, and combined 1:1 into a single sample (HLDBIA or SILAC). Peptide mixtures were

225 cleaned by chloroform/methanol precipitation and digested with lysC/trypsin mixture at 1:50

226 enzyme:protein ratio as described above. Peptides were labeled with TMTPro^{TM}Zero following

227 the manufacturer's protocol. Prior to quenching, TMT labeling efficiency was checked to ensure

228 >95% labeling. Streptavidin enrichment was performed on AssayMap Bravo, as described

229 above.

230

231 To fractionate hyperplexed samples, eppendorf tips were loaded with SDB-XC Empore disks

232 and activated with 100% acetonitrile, followed by 2 washes with 20 mM NH_4HCO_2. Lyophilized

233 peptides were resuspended in 20 mM NH_4HCO_2 and loaded onto the disks, followed by two

234   washes, first elution with 12% acetonitrile/20 mM NH_4HCO_2 and second elution with 50%

235   acetonitrile/20 mM NH_4HCO_2. Elutions were dried down prior to injection.

Nuclear lysate treatment

237   For nuclear fractionation, MCF7 cells were propagated to six 15 cm plates and collected by

238   scraping into ice-cold PBS when at ~90% confluency. Nuclear extraction was performed using

239   REAP protocol. [17] Briefly, cell pellets were resuspended in 0.1% NP40 and triturated 5 times.

240   After 10 sec pulse-spin, the supernatant (cytosol) was removed, followed by an additional wash

241   with 0.1% NP40. Pelleted nuclei were resuspended in PBS and sonicated on tip with thirty 1 sec

242   pulses, 20% AMP. Insoluble fraction was removed by centrifugation at 14,000xg for 10 minutes

243   at 4°C. Soluble nuclear fraction was treated with indicated compounds at 100 μM for 1hr at

244   room temperature. Unlabeled reactive cysteine sites were probed with 500 μM LDBIA or HDBIA

245   for 1 hr at room temperature. Samples were reduced with 5 mM DTT for 30 minutes at room

246   temperature and LDBIA/HDBIA samples pairs were combined 1:1. Combined samples were

247   alkylated with 20 mM iodoacetamide for 30 minutes at room temperature. Sample clean-up,

248   digestion and TMTPro labeling was performed as described above.

Mass Spectrometry analysis for ABPP workflow with hyperplexing

250   Hyperplexed samples were reconstituted in buffer A for liquid chromatography-tandem mass

251   spectrometry (LC-MS) analysis on Dionex Ultimate 3000 RSLCnano system (Thermo Fisher

252   Scientific, Inc) and Orbitrap Eclipse Tribrid MS (Thermo Fisher Scientific, Inc). Peptides (~1 ug)

253   were resolved on 25 cm x 75 μm Aurora column packed with 1.6 μm C18 (Ion Opticks) by 135

254   minutes linear gradient of 4% to 30% buffer B (98% acetonitrile/0.1% formic acid) in buffer A

255   (2% acetonitrile/0.1% formic acid) at a flow rate of 300 nL/min.

256   Where indicated, standard SPS-MS3/Multi-Notch-MS3 TMT method[2] was performed with the

257   same parameters as in regular ABPP workflow described above. In addition, a "Targeted Mass

258   Difference" node was used for acquisition of precursor isotopic peptide pairs with the following

259   parameters: for SILAC experiment, delta M1 was set to 8.0142 and 10.0082 and for HLDBIA

260   experiment, delta M1 was set to 6.0319; partner intensity range set to 10-100%; subsequent

261   scans performed on both ions in the pair, with the same charge state.

262   For SILAC/TMT hyperplexed sample analysis using Thermo RTS and InSeqAPI RTS

263   acquisition, the following RTS parameters were used: Uniprot mouse database October 2022

264   version, including 25,367 Swissprot sequences of canonical and protein isoforms, plus common

265   contaminants and decoys; static modifications included Cys carbamidomethylation (+57.0215),

266   K and n-term TMTZero™Pro (+295.1896); variable modifications included Met oxidation

267   (+15.9949), Cys with light desthiobiotin label (+239.1633), Lys +8.0142, Arg +10.00827. Other

268   parameters were set to 0 missed cleavages, 3 variable modifications per peptide, FDR filtering

269   disabled, 35 ms maximum search time, and protein filter set to exclude decoy matches from

270   MS3 analysis. Comet scoring threshold: 0.5 Xcorr, 0.05 dCn, 10 ppm precursor mass error, +2

271   charge state. MS3 HCD settings were as in standard SPS-MS3 mode, with 600% normalized

272   AGC target and 400 ms maximum injection time.

273   For "heavy" SILAC/TMT sample analysis, the following RTS parameters were used: Uniprot

274   mouse database (October 2022 version, including 25,367 Swissprot sequences of canonical

275   and protein isoforms, plus common contaminants and decoys; static modifications included Cys

276   carbamidomethylation (+57.0215), n-term TMTZero™Pro (+295.1896), Lys TMTZero™Pro

277   +295.1896, Arg +10.00827; variable modifications included Met oxidation (+15.9949), Cys with

278   light desthiobiotin label (+239.1633). Other parameters were set to 0 missed cleavages, 3

279   variable modifications per peptide, FDR filtering enabled, 35 ms maximum search time. Comet

280   scoring threshold: 0.5 Xcorr, 0.05 dCn, 10 ppm precursor mass error, +2 charge state. MS3

281   HCD settings were as in standard SPS-MS3 mode, with 600% normalized AGC target and 400

282   ms maximum injection time.

283   For "light" SILAC/TMT sample analysis, the following RTS parameters were used: Uniprot

284   mouse database October 2022 version, including 25,367 Swissprot sequences of canonical

285   and protein isoforms, plus common contaminants and decoys; static modifications included Cys

286   carbamidomethylation (+57.0215), Lys and n-term TMTZero^{TM}Pro (+295.1896); variable

287   modifications included Met oxidation (+15.9949), Cys with light desthiobiotin label (+239.1633).

288   Other parameters were matched to "heavy" SILAC/TMTPro sample analysis.

289   For HLDBIA/TMT hyperplexed sample analysis using Thermo RTS and InSeqAPI RTS

290   acquisition, the following parameters were used for RTS: Uniprot human database (October

291   2022 version, includes 42,352 Swissprot sequences of canonical and protein isoforms, plus

292   decoys); static modifications: Cys carbamidomethylation (+57.0215), K and n-term

293   TMTZero^{TM}Pro (+295.1896); variable modifications: Met oxidation (+15.9949), Cys with light

294   desthiobiotin label (+239.1633) and Cys with heavy desthiobiotin label (+245.1952). Other

295   parameters were set to 1 maximum missed cleavages, 2 max variable modification per peptide,

296   FDR filtering disabled, 35 ms maximum search time, and protein filter set to exclude decoy

297   entries from MS3 analysis. Comet scoring threshold: 0.5 Xcorr, 0.05 dCn, 10 ppm precursor

298   mass error, +2 charge state. MS3 HCD settings were as in standard SPS-MS3 mode, except

299   600% normalized AGC target and 400 ms maximum injection time.

300   For LDBIA/TMT and HDBIA/TMT sample analysis using Thermo RTS and InSeqAPI RTS

301   acquisition, the following parameters were used for RTS: Uniprot human database October

302   2022 version (includes 42,352 Swissprot sequences of canonical and protein isoforms, plus

303   decoys); static modifications included Cys carbamidomethylation (+57.0215), K and n-term

304   TMTProZero (+295.1896); variable modifications included Met oxidation (+15.9949) and Cys

305   with light desthiobiotin label (+239.1633) or Cys with heavy desthiobiotin label (+245.1952).

306   FDR filtering was enabled, whereas other parameters were matched to the hyperplexed sample

307   analysis.

308   **InSeqAPI Instrument Control**

309   inSeqAPI was written in C# (.Net Framework 4.6.1) and interfaces with Orbitrap Tune versions

310   3.4 and higher on Orbitrap Eclipse and Ascend mass spectrometers. Search functionality within

311   inSeqAPI is enabled by a previously described version of Comet (v.2020013, comet-

312   ms.sourceforge.net/).[4] Real-time access to spectral data was enabled by the Thermo Scientific

313   instrument API (https://github.com/thermofisherlsms/iapi). For inSeqAPI methods, the native

314   instrument method was used to control the nano-LC as well as MS1 scans, filters guiding

315   selection of precursor ions for MS2, and MS2 scans (as described above). inSeqAPI received

316   the MS2 scans collected by the instrument and analysis was triggered by specific MS2 scan

317   descriptions.

318

319   The inSeqAPI 'real-time search' method contained a single node comprising a real-time search

320   filter to match the spectra to a peptide sequence, PPM adjustment filter to account for

321   systematic mass shifts, an LDA filter to perform real-time false discovery estimations, a decoy

322   filter to ensure decoy identifications are not quantified, a modification filter to only quantify

323   peptides that contained a modified cysteine residue, a mass range filter to remove y1 ions, a

324   precursor range exclusion filter to exclude peaks around the precursor, a tag loss exclusion filter

325   to exclude fragment ions that lost the TMT tag (i.e., TMTc ions), a targeted SPS ion filter to only

326   select fragment ions that contain the quantitative tag, and finally a topN filter to select the most

327   abundant fragment ions for SPS-MS3 analysis. If all of these filters were passed, instructions for

328   an SPS-MS3 scan were sent to the instrument.

329

330   The inSeqAPI 'PairQuant' method contained two nodes: 1) PairQuant node, 2) PairSeq node

331   (**Supp. Fig. 6**). The PairQuant node comprised the same filters as the 'real-time search' method

332   described above with the addition of a PairQuant filter that keeps a record of which peptides

333   have been identified and saves spectral information relating to which SPS ions would be

334   selected for SPS-MS3 and a PairSeq filter that keeps a record of which peptides have been

335   identified. If all filters are passed the PairQuant node will send instructions for two SPS-MS3

336   scans corresponding to the light and heavy partners of the hyperplexed pair. The PairSeq node

337   contains only a PairSeq filter (the same PairSeqFilter object that is referenced in the PairQuant

338   node) that checks to see if both peptides within a hyperplexed pair have been identified. The

339   'Targeted Mass Difference' filter within the vendor method isolates and fragments each peptide

340   within a pair in successive scans. Due to this the PairSeq filter assumes that both peptides

341   within a hyperplexed pair will be sequenced within 2 scans. If only one peptide is sequenced,

342   the PairSeq node will send instructions to perform an ion trap MS2 on the missing peptide in an

343   attempt to sequence it. This MS2 will be analyzed within the PairQuant node just as MS2 scans

344   which are triggered by the vendor method.

345   **Data analysis for ABPP workflow**

346   Offline search was performed using comet v.2019.01 with parameters matched to the RTS

347   search. Linear Discriminator Algorithm was used for peptide filtering to <2% FDR, as previously

348   described.[18] The PSMs were filtered out if they were from peptides with length less than 7. For

349   quantification of site ligandability, total reporter ion intensity across all 18 channels / plex was

350   filtered to >10,000 and individual reporter ion intensity less than 2^8 noise estimate. In the case

351   of redundant PSMs (i.e. multiple PSMs in one MS run corresponding to the same peptide ion),

352   only the single PSM with the least missing values or highest isolation specificity or highest

353   maximal reporter ion intensity was retained for subsequent analysis. If the same PSM was

354   identified in multiple fractions, only the single PSM with the highest isolation specificity or least

355   missing values or highest sum of reporter ion intensity were kept. The reporter ion intensities of

356   all the peptide ions mapped to a cysteine site were summarized into a single site level intensity

357   in each channel and plex by MSstats (v4.8.3).[19] For each cysteine site, the local normalization

358   on the summaries was performed to reduce the systematic bias between plexes. For the local

359   normalization, we created an artifact reference channel by averaging over all the 18 channels

360   for each cysteine site and plex. The cysteine site abundances in the reference channel of each

361   plex were equalized to their median across two plexes. MSstatsTMT R package (v2.8.0)[20]

362   estimated log2(fold change) and the standard error by linear mixed effect model for each

363   cysteine site. The inference procedure was adjusted by applying an empirical Bayes shrinkage.

364   To test two-sided null hypothesis of no changes in abundance, the model-based test statistics

365   were compared to the Student t-test distribution with the degrees of freedom appropriate for

366   each site and each dataset. The resulting P values were adjusted to control the FDR with the

367   method by Benjamini–Hochberg.

368   For iodoTMT analysis, MS/MS spectra were searched using comet against Uniprot human

369   database October 2022 version (includes 42,352 Swissprot sequences of canonical and protein

370   isoforms, plus decoys). Search parameters included trypsin cleavage with allowance of up to 2

371   missed cleavage events, a precursor ion tolerance of 50 ppm, and a fragment ion tolerance of

372   0.8 Da. Searches permitted variable modifications of methionine oxidation (+15.9949 Da), and

373   variable modifications of cysteine (+329.2266) for iodoTMT. Reporter ions produced by the

374   IodoTMT tags were quantified with an in-house software package known as Mojave.

375   MSstatsTMT R package was used for statistical analysis as described above.

376   **Gene ontology enrichment analysis**

377 A list of Uniprot accession entries identified in the nuclear ABPP experiment were uploaded to

378 DAVID.[21,22] UniprotKB Keywords Cellular Component enrichment analysis was performed

379 against proteins identified in the whole-cell ABPP experiment as a background.

380

Data and Software availability

382 Mass spectrometry raw data files, metadata, quantification data and statistical analysis results

383 have been uploaded to the UCSD MassIVE mass spectrometry data repository

384 (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) using identifier MSV000092113 and

385 password "hyperplexing". The inSeqAPI software enabling inSeqAPI RTS and PairQuant data

386 collection is available upon request and requires users to have signed the Thermo IAPI

387 agreement as well as a distribution agreement with Genentech, Inc.

Results

Development of intelligent data acquisition approaches for TMT profiling of

reactive cysteine sites

391 In a typical activity-based proteome profiling study of ligandable cysteine sites, cells or native

392 cell lysates are treated with an electrophilic compound that reacts with ligandable sites (**Figure**

393 **1A**). Subsequent incubation with a reactive probe, such as desthiobiotin-modified

394 iodoacetamide (DBIA), labels all of the remaining reactive sites. For relative quantification

395 between conditions, peptide mixtures are labeled with tandem mass tags (TMT), mixed and

396 biotinylated peptides are enriched by streptavidin. Site ligandability can be quantified based on

397 the relative abundance of peptides containing DBIA labeled cysteine residues in compound-

398 treated and vehicle-treated samples.

399   Real-time search (RTS) MS data acquisition methods that greatly improve the depth of TMT-

400   based quantification studies were recently introduced within the vendor software for select

401   instruments (Orbitrap Eclipse and Orbitrap Ascend).[3,4] In this approach, ion trap MS2 spectra

402   containing precursor fragment ions, are collected by the instrument and searched in real time

403   against a database of tryptic peptides. Only spectra matched to a peptide with high confidence

404   are sent for acquisition of Orbitrap SPS-MS3 spectra that record relative abundances of TMT

405   report ions (**Figure 1A**). This approach improves the depth of SPS-MS3 analysis by limiting the

406   collection of slow SPS-MS3 scans to peptides that have been identified and dedicating a larger

407   percentage of instrument duty cycle to the collection of SPS-MS3 scans that will be used in the

408   final quantitative analysis. However, these intelligent data acquisition approaches provided by

409   instrument vendors have been optimized for measurements of protein abundance. These

410   methods typically make assumptions that may not be applicable to experiments quantifying

411   PTM bearing peptides. For example quantitative scans are triggered for all peptide

412   identifications, even if users only want to quantify peptides that contain a particular PTM. This is

413   an important consideration for reactive cysteine profiling where we wished to limit collection of

414   quantitative scans to peptides that contained DBIA labeled cysteine residues.

415   Due to these limitations, we set out to tailor the RTS acquisition method to the analysis of ABPP

416   samples through the development of an intelligent data acquisition (IDA) instrument application

417   program interface (iAPI) called inSeqAPI (**Figure 1B**). Like other iAPI programs[3,4], inSeqAPI

418   captures data that are collected by the mass spectrometer in real-time, analyzes the data

419   through a number of user-defined filters, and creates instructions for new scans that are sent

420   back to the mass spectrometer for execution. The differentiating characteristic of inSeqAPI is

421   the ability to develop customized methods through modular method design that reflects the

422   construction of methods within the vendor software. Methods within inSeqAPI are broken down

423   into separate nodes that each contain a list of filters that are successively applied to the spectra

424 (**Figure 1B**). If all filters are passed the instructions to perform the scan(s) specified within the

425 node are sent to the instrument. Importantly, a method can contain multiple nodes, enabling a

426 unique set of filters to be applied to different types of scans as indicated by the scan

427 descriptions. Methods can be easily edited by re-ordering filters and changing filter parameters

428 within the graphical user interface (GUI) (**Figure 1B**). Lastly, new filters are implemented

429 through the creation of a single C# class, quickly enabling new functionality such as the

430 'Modification Filter' used to ensure only identified peptides that contain a modified cysteine

431 residue are quantified (**Figure 1B**).

432 To benchmark InSeqAPI against non-RTS and vendor RTS (Thermo RTS) acquisition methods

433 for ABPP studies, native cell lysates of C2C12 myoblasts were treated with "light" DBIA (LDBIA)

434 reactive probe, followed by TMT labeling and streptavidin enrichment of biotinylated peptides.

435 For MS data acquisition, all MS1, MS2 and MS3 parameters were matched for accurate

436 comparison of the methods. However, it was observed that the RTS acquisition method

437 sampled more low-intensity precursor ions than the non-RTS method (**Supp. Figure 2A**).

438 Therefore, MS3 AGC target and Max IT were optimized in RTS acquisition approach to boost

439 TMT intensity and signal/noise (**Supp. Figure 2B**). Overall, a 15% increase in the number of

440 quantified biotinylated cysteine peptides was observed with Thermo RTS acquisition compared

441 to non-RTS method (**Figure 1C**). Furthermore, an additional 7% increase was yielded by the

442 InSeqAPI RTS method. A boost in the number of quantified cysteine-containing peptides by

443 InSeqAPI RTS was largely attributed to a greater number of total spectra acquired (**Supp.**

444 **Figure 2C**). We further tested InSeqAPI performance in iodoTMT cysteine labeling approach.

445 As expected, we observed that RTS approaches allow for deeper proteome coverage compared

446 to a non-RTS method (**Fig. 1D**). Overall, our data demonstrates that InSeqAPI generates

447 comparable or better results than vendor software and therefore can be utilized in real-time

448 database search of mass spectrometry data.

Improving hyperplex pair quantification with PairQuant within inSeqAPI

Due to the stochastic nature of data-dependent MS acquisition methods, quantifying peptides or proteins across multiple TMT plexes can lead to missing data. For instance, we observed a loss of ~40% of replicate peptide quantification data when comparing the number of Cys-DTB (desthiobiotin-labeled cysteine) peptide pairs quantified between two TMT runs (**Supp Fig. 3**). In order to improve the number of replicate quantification between TMT plexes, we employed two alternative workflows utilizing protein-level isotopic labeling. In the first workflow, SILAC is used for labeling of cells in replicate samples, followed by lysis, labeling with DBIA probe, mixing of SILAC replicates and processing them as a single plex for TMT-based quantification (**Fig. 2A**). In the second workflow, replicate samples are labeled with isotopically "heavy" DBIA (HDBIA) and "light" DBIA (LDBIA), mixed and also processed as a single plex for TMT-based quantification (**Fig. 2A, DBIA structures in Supp. Fig. 1**). Therefore, in each workflow, a hyperplexed sample is formed containing protein-level isotopic tags for replicate analysis and peptide-level TMT tags for comparisons between conditions.

First, we tested quantification of paired peptides in a hyperplexed samples using non-RTS and Thermo RTS acquisition methods with Targeted Mass Difference (TMD) trigger. The TMD acquisition parameter triggers MS2 analysis of both isotopically labeled peptide pairs in a hyperplexed sample. As expected, RTS quantification produced an 18% boost in the number of quantified Cys-DTB peptides in SILAC/TMT and HLDBIA/TMT hyperplex analyses over non-RTS method (**Fig. 2B-C**). Interestingly, only a modest 10% improvement or no improvement was observed in the recovery of paired Cys-DTB peptides in SILAC/TMT or HLDBIA/TMT hyperplex, respectively (**Fig. 2B and 2C, top panels**).

473   Upon closer examination, we observed that some peptides had an isotopic pair that passed

474   online FDR filtering, but not offline FDR filtering. We reasoned that if one PSM from the isotopic

475   pair passed offline FDR filtering, the paired PSM is also a true match to that peptide sequence

476   and should be included in the final peptide identification list. As a result, we observed an

477   additional gain in the number of peptide pairs in the hyperplexed samples (**Figs. 2B and 2C,**

478   **bottom panels**). However, ~20-30% of peptides identified in the hyperplexed sample were still

479   missing quantification from their paired isotopic partner.

480

481   To address this issue, we developed the PairQuant method within InSeqAPI (**Fig. 2D**). Briefly,

482   using the vendor instrument method, the TMD filter was used to trigger MS/MS analysis of both

483   peptides within a hyperplexed pair. Within inSeqAPI, PairQuant waited for both peptide spectra

484   to be identified before sending instructions to perform separate SPS-MS3 scans on the

485   hyperplexed peptides (**Fig. 2D, 1**). For cases where only one of the hyperplexed peptides was

486   identified due to the correct peptide selection but failure to identify both peptides (**Figure 2D, 2**),

487   TMD selection of incorrect precursor pairs (**Fig 2D, 3**), or MS/MS only performed on one of the

488   hyperplexed peptides within the pair (**Figure 2D, 4**) PairQuant instructed the MS to perform an

489   ion trap MS scan to identify the missing hyperplexed partner. If this scan resulted in the

490   identification of the missing partner peptide then PairQuant instructed the MS to perform

491   separate SPS-MS3 scans on the hyperplexed partners.

492

493   As a result, we observed that 98%-100% of peptide pairs had SPS-MS3 scans acquired on both

494   partners when using the PairQuant method (**Figs. 2E and 2F**). While the number of unique cys-

495   DBIA peptides with complete quantitation (i.e., quantified in both hyperplexed pairs) was higher

496   than utilizing two separate TMT plexes, the overall number of unique cys-DBIA peptides was

497   reduced due to the increased sample complexity. To improve depth of cys-DBIA quantification

498   we reduced the complexity of the hyperplexed samples by creating two high-pH fractions

499   through stagetip fractionation (**Figs. 2E and 2F**). These two fractions were analyzed by

500   PairQuant using the same amount of instrument acquisition time as two separate TMT plexes.

501   This resulted in a similar number of quantified unique cys-DBIA peptides when compared to two

502   TMT plexes, but nearly all (98% - 100%) of these were quantified in both of the hyperplexed

503   partner peptides (**Figs. 2E and 2F**). A similar analysis of fractionated hyperplexed samples with

504   Thermo RTS demonstrated a decrease in the total number of quantified unique cys-DBIA

505   peptides as well as the percentage of these peptides that were completely quantified (72% -

506   78%) (**Figs. 2E and 2F**). These data demonstrate that both the SILAC/TMT and HLDBIA/TMT

507   hyperplexing workflows combined with PairQuant acquisition within InSeqAPI allow for similar

508   depth of peptide identification as a regular TMT workflow, but with a substantial improvement in

509   complete quantification of unique cys-DBIA peptides.

510   **Hyperplexed cysteine profiling in the nuclear fraction**

511   ABPP with cysteine-targeting covalent fragment compounds has been shown to aid in

512   identification of ligandable pockets across the cellular proteome.[5,10,11] We wanted to test

513   whether we can improve the throughput and coverage of ligandable pocket profiling in the

514   nucleus by applying an ABPP workflow that included nuclear isolation, hyperplexing, and

515   PairQuant acquisition.

516   Nuclei were isolated from MCF7 cells using the REAP protocol.[17] By Western blotting, we

517   observed an enrichment of the nuclear marker histone-H3 and depletion of the cytosolic marker

518   α-tubulin in the nuclear fraction compared to the whole-cell lysate and the cytosolic fraction

519   (**Supp. Fig. 4**). Nuclear lysates in PBS were treated with DMSO or eight cysteine-reactive

520   covalent fragments in four replicates, 36 samples total (**Figs. 3A and 3B**). Upon labeling of

521   unreacted cysteines with LDBIA or HDBIA, "light" and "heavy" sample pairs were combined and

522   processed as a single hyperplex for TMTPro-based quantification using 18 channels. As a

523   result, we were able to quantify 9953 sites from 4081 proteins, with the majority of sites

524   quantified across 4 replicate experiments in the final dataset (**Figs. 3C and 3D, Supp. Table 1**).

525   For each treatment, measured compound occupancy strongly correlated between HDBIA and

526   LDBIA replicate experiments, except for compound G59, which also showed no significant cys

527   site engagement (**Figs. 3E and 3F, Supp. Table 1**). Although all compounds showed high

528   reactivity in cysteine trapping assay (**Fig. 3F and Supp. Fig. 5**), their promiscuity greatly varied

529   as shown by the heatmap of sites with ≥75% compound occupancy (**Fig. 3F**). Among tested

530   compounds, three, KB02, KB03 and KB05, were previously utilized in other activity-based

531   proteome profiling studies.[5,11,23] These compounds were shown to have a high coverage of

532   ligandable sites when treatments are done at >100μM concentrations in cell lysates.[5] Similarly,

533   we observed ~1-4% site labeling by these compounds in the nuclear fraction. Two compounds

534   with even greater promiscuity of 23% and 28% were a chloroacetamide compound G84 and a

535   quinazolinone-based compound G51, respectively (**Fig. 4F**). These results demonstrate that the

536   hyperplexing approach allows for fast and accurate screening of compounds with various

537   reactivity across the nuclear proteome.

538   **Ligandable nuclear targets**

539   To compare the depth of nuclear cysteine peptide enrichment to a whole-proteome profiling

540   experiment, we performed a standard ABPP experiment in whole cell lysates from MCF7 cells.

541   InSeqAPI RTS acquisition of biotinylated cysteine peptides from whole-cell lysates resulted in

542   quantification of 5851 proteins with 16706 cysteine sites from three TMT plexes (**Supp. Table**

543   **2**). We observed that 2805 sites from 1784 proteins were uniquely identified in the nuclear

544   fraction compared to the whole-cell lysate; this included 560 proteins that were quantified only in

545   the nuclear experiment (**Fig. 4A**). The cellular component annotation analysis of the proteins

546   detected in the nuclear fraction revealed significant enrichment of the nucleus-localized proteins

547 (**Fig. 4B, Supp. Table 1**), supporting our hypothesis that we can improve the depth of activity-

548 based proteome profiling studies through subcellular fractionation.

549 For further evaluation of ligandable nuclear cysteine sites, we focused on the compounds that

550 showed medium promiscuity in ABPP experiment, KB02, KB03, KB05, G35 and G92 (**Fig. 3F**).

551 Among liganded proteins in the nuclear fraction (i.e. occupancy ≥50%), we observed 21 E3

552 ubiquitin-protein ligases and core components of their complexes (such as Skp, Cullin, F-box

553 containing complex, SCF), as well as two deubiquitinases (DUBs) (**Fig. 4C, Supp. Table 1**).

554 These proteins localize and function primarily or partially in the nucleus, participating in

555 processes such as DNA damage response (BRCA1, HERC2, RNF169, TRIP12, HUWE1,

556 UBR5) and transcription (TRIM33, IRF2BPL, UBR2, USP16, USP22, DCAF1, HUWE1, RNF40).

557 DUB enzymes ubiquitin carboxyl-terminal hydrolase 16 and 22 (USP16 and USP22) function in

558 the transcription regulation in the nucleus through deubiquitination of histone substrates

559 (reviewed in[24]). USP22 Cys 494 (87.5% occupancy by KB03) is localized in the USP domain,[25]

560 and USP16/Ubp-M Cys 191 (78% occupancy by G92) is adjacent to the USP domain.[26] Both of

561 these sites were also liganded in the whole-cell lysate analysis, further supporting the finding of

562 the nuclear fractionation experiment (**Fig. 4C**). We also observed ligandable sites on three E3

563 ubiquitin ligases, UBR2, UBR4 and UBR5 that function in the N-end rule proteolytic pathway.

564 Their targets are usually short-lived proteins that are recognized by n-degrons that consist of

565 exposed or modified n-terminal residues.[27] UBR2 has an essential function in regulation of

566 chromatin-wide transcriptional silencing through ubiquitination of H2A histone.[28] UBR2 Cys1717

567 was liganded by G92 in the nuclear and whole-cell lysates, and is localized in the UBR

568 autoinhibitory domain (UAIN) at the protein c-terminus.[29] A few sites were uniquely quantified in

569 the nuclear fraction, such as BRCA1 C360. This site is present in the BRCA1 region interacting

570 with DNA recombinase RAD50.[30] Cys360 was liganded by G92 and G35 electrophiles at 73%

571 and 53% occupancy, respectively. Therefore, nuclear activity-based proteome profiling allows

572   for identification of new sites in the protein-protein interaction domains of nuclear ubiquitin

573   ligases and deubiquitinases.

574   Among 577 proteins unique to the nuclear fraction, we identified 27 transcription regulators

575   (**Supp. Table 1**). Transcription factors represent challenging targets for drug development, but

576   their function can potentially be modulated through protein-protein and protein-DNA interactions.

577   In the nuclear ABPP analysis, we identified ligandable cysteine sites on 12 transcription

578   regulators (**Fig. 4D**). ATF2 Cys79 is localized in a zinc-coordinated transactivation domain and

579   was liganded by G92 at 50.8%. Interestingly, another activation transcription factor family

580   member (ATF7) was also liganded by G92 at a homologous site. These cysteine sites are

581   localized near a phosphoswitch motif regulated by MAPK-mediated phosphorylation.[31]

582   Transcription activator PBX4 was occupied at 59-77% by three fragments, G35, G92 and KB05

583   on Cys 58, which is present in protein-protein interaction domain PBC-A.[32] Transcriptional

584   regulating protein of 132 kDa (TReP-132/TRERF1) was liganded by four compounds at 50-79%

585   on Cys 876, which is localized in SANT domain that may mediate protein-protein interactions

586   with HDAC1/2 within MiDAC complex.[33] Transcription factor HES1 was liganded by all

587   compounds with medium promiscuity at 48-87% occupancy on Cys 128. HES1 is a basic helix-

588   loop-helix (bHLH) transcription repressor that contains a unique Orange domain with unknown

589   function, which is present exclusively in transcription repressors (**Fig. 4E**).[34] Quantification of

590   Cys 128 abundance was based on detection of the light and heavy Cys-DTB pair of ions (**Fig.**

591   **4F**), both of which were selected for MS2 sequencing. Eight TMT-containing b ion fragments

592   from each MS3 spectrum were assigned by InSeqAPI for MS3 fragmentation (**Fig. 4G**). Relative

593   TMT reporter ion signal was reproducible between heavy and light-labeled peptide species and

594   within each TMT plex (**Fig. 4H**). Therefore, our hyperplexing approach enabled by PairQuant

595   acquisition method can provide accurate quantification of cysteine engagement in activity-based

596   proteome profiling studies by reducing the number of missing values between replicate TMT

597   plexes.

Discussion

599   In this work, we described the analysis of ABPP samples using intelligent data acquisition (IDA)

600   enabled by InSeqAPI. We focused specifically on improving sample throughput by employing

601   hyperplexed analysis, an approach where samples are isotopically labeled on protein level for

602   replicate analysis before TMT labeling at the peptide level to enable comparisons between

603   conditions. To achieve complete coverage of hyperplexed peptide pairs we introduced

604   PairQuant, a method that utilizes a real-time database search to trigger quantitative scans only

605   when both hyperplexed peptide pairs have been identified. If only one of the peptides within a

606   hyperplexed pair is identified, PairQuant sends an identification scan for the missing

607   hyperplexed partner. This approach improved the quantification of hyperplexed pairs to ~98-

608   100% after rescuing partners that were matched to the correct peptide, but did not pass offline

609   FDR filters. A challenge of this approach is that increased sample complexity led to a decrease

610   in the total number of unique cysteine containing peptides quantified in a single run. To reduce

611   sample complexity, hyperplexed samples were fractionated into two fractions and analyzed in

612   the same amount of instrument time as two-plex TMT analysis. We demonstrated that this

613   approach allows for a significant reduction in missing values between plexes (<2%), while

614   maintaining the same number of total unique peptide sequences quantified.

615   While our analyses featured two hyperplexed TMT plexes, the level of hyperplexing can be

616   increased and potentially incorporated into specific workflows through the use of alternative

617   metabolic labeling strategies. For instance, triple SILAC would allow for triplicate analysis of 18

618   conditions (54 samples total) in an ABPP workflow. Such labeling approach was previously

619   demonstrated in combination with a 6-plex TMT to monitor yeast proteome changes in response

620   to rapamycin stimulation. [15] Additionally, the use of metabolic labeling by isotopic tyrosine was

621   previously demonstrated in phosphoprofiling studies.[35] Therefore, incorporation of tyrosine

622   SILAC into a hyperplexed method could extend the number of timepoints used in

623   phosphosignaling studies.

624   We also tested ABPP with a nuclear enrichment strategy, where nuclear extracts were labeled

625   with scout probes, promiscuous electrophilic fragments with highly reactive cysteine-targeting

626   warheads. A single hyperplexed nuclear sample combined 36 conditions which included eight

627   compounds and a DMSO control in four replicates, where two replicates were labeled with

628   isotopically "light" and "heavy" cysteine-reactive tags and two replicates were isobarically tagged

629   by an 18-plex TMT reagent. Although all of the scout fragments tested were intrinsically highly

630   reactive and stable in PBS over the labeling period, we observed a broad range of hit rates

631   across compounds. This result underlies the complexity of electrophilic compound interactions

632   with their targets that is not driven purely by the warhead reactivity. One of the advantages of

633   nuclear enrichment strategy in the covalent fragment screening is the removal of highly reactive

634   cysteine sites present in the cytosolic and mitochondrial proteome, which can outcompete less

635   reactive sites in the nuclear proteome. As a result, nuclear enrichment allowed for identification

636   of ligandable cysteine sites on nuclear-localized proteins that were missed in a whole-proteome

637   analysis, including nuclear ubiquitin ligase machinery and transcription regulators.

638   Recent studies have shown that covalent ligands can be used as molecular glues[36] or can be

639   incorporated into heterobifunctional electrophilic proteolysis targeting chimeras (PROTACs) for

640   targeted degradation.[23,37–39] As such, covalent small molecules also provide opportunities for

641   targeting functions of transcription factors, which are historically challenging to drug. For

642   instance, identification of a ligandable cysteine within the Orange domain of HES1 transcription

643   repressor highlights a possibility for developing a covalent small molecule that modulates its

644   functions through protein-protein interactions and/or a covalent PROTAC.

645   In summary, this study describes an application of intelligent data acquisition in hyperplexed

646   activity-based proteome profiling studies, which increases experimental throughput and

647   completeness of the site occupancy data. Importantly, the PairQuant data acquisition method

648   can be used on any hyperplexed sample, potentially unlocking this approach for other

649   proteomics applications. For example, hyperplexing was recently described for single cell

650   proteomics in a technique termed "hyperSCP", here multiplexing was increased to 28 samples

651   per run making it possible to analyze up to ~280 cells per day.[40] Lastly, this work represents an

652   introduction inSeqAPI, an iAPI framework that enables rapid development of custom IDA

653   methods. While this study focused on the PairQuant method, inSeqAPI will continue to evolve

654   with methods tailored specifically for other proteomics approaches (e.g., single cell proteomics,

655   HLA peptidomics, etc.).

Acknowledgements

Conflict of Interest

661   The authors declare no competing interests. All authors are employees and shareholders of

662   Genentech Inc., a wholly owned subsidiary of Roche.

Author contributions

664   H. G. B., and C. M. R. conceptualization; H. G. B and D. S. K. methodology; H. G. B., T. P. M.,

665   S. W., M. C., C. M. R. formal analysis; H. G. B., T. P. M., S. W., M. C., C. M. R. investigation; H.

666   G. B., and C. M. R. writing–original draft; H. G. B., T. P. M., S. W., M. C., C. M. R. writing–

667   review and editing; C. M. R. supervision.

668   **Figure Legends**

669   **Fig 1. Development and evaluation of InSeqAPI intelligent data acquisition**

670   **platform in activity-based proteome profiling workflow**

671   **A**. Schematic of activity-based proteome profiling workflow with real-time database searching

672   and MS3 acquisition of biotinylated cysteine peptides. **B**. User interface of InSeqAPI, an

673   intelligent data acquisition application. **C**. C2C12 mouse myoblast lysates were treated with 500

674   µM LDBIA probe for 1 hr. Equal amounts of enriched cys-containing biotinylated peptides were

675   injected for analysis by SPS-MS3 without RTS, with vendor RTS and InSeqAPI RTS methods in

676   two biological replicates. Number of unique cys-DTB peptide-spectrum matches and their

677   relative values compared to InSeqAPI RTS are compared (top and bottom barplot, respectively,

678   n=2). **D**. MCF7 cell lysates were labeled with iodoTMT reagent for 1hr and TMT-labeled

679   peptides were enriched with anti-TMT resin. Enriched peptide mixture was analyzed with RTS

680   or without RTS, as indicated (n=2).

681   **Fig 2. Sample hyperplexing in activity-based proteome profiling**

682   **A.** Workflow schematic for activity-based proteome profiling with sample hyperplexing. In

683   SILAC/TMT hyperplex, isotopic labeling is done on protein level by "heavy" or "light" SILAC,

684   followed by isobaric TMT labeling on peptide level. In HLDBIA/TMT hyperplex, isotopic labeling

685   on protein level is done by "heavy" or "light" DBIA, followed by TMT labeling on peptide level.

686   Duing MS analysis, Targeted Mass Difference node is enabled for detection of isotopic peptide

687   pairs in MS1. **B and C.** C2C12 SILAC lysates were labeled with 500 µM LDBIA probe for 1 hr

688   (B) or MCF7 lysates were labeled with 500 μM HDBIA ("heavy") or LDBIA ("light") probe for 1 hr

689   (C). Number of isotopically labeled biotinylated PSM pairs (both replicates acquired) and PSMs

690   without pairs (missing replicate measurement) is plotted for each experiment. Top panel shows

691   PSMs present after offline FDR filtering. Bottom panel includes additional PSMs rescued from

692   unfiltered data due to the presence of their isotopic pairs in the filtered data. Data from two

693   individual experiments is shown. **D**. PairQuant acquisition of hyperplexed peptide pairs within

694   InSeqAPI for SPS-MS3 analysis. (1) Both hyperplexed peptides within pair are fragmented and

695   identified triggering SPS-MS3. (2) Both hyperplexed peptides within pair are fragmented but

696   only one is identified, triggering PairQuant MS2 to identify missing peptide and trigger SPS-

697   MS3. (3) Incorrect selection of hyperplexed peptides leading to only one peptide identified

698   prompts PairQuant MS2 to identify missing peptide and trigger SPS-MS3. (4) Only one peptide

699   within pair is fragmented prompting PairQuant MS2 to identify missing peptide and trigger SPS-

700   MS3. **E and F**. Experiments performed as described in panels B and C, including a separate

701   analysis of isotopically labeled peptides as single TMT plexes (data labeled as "2XTMT

702   plexes"). For the fractionation experiment, hyperplexed samples were fractionated on stagetip in

703   high pH into two fractions. Average from two individual experiments is plotted.

704

Fig 3. Activity-based proteome profiling in the nuclear fraction

706   **A**. Cysteine ligandability profiling workflow with HLDBIA/TMT in MCF7 nuclear lysates with

707   electrophilic scout probes (8 treatments, 4 replicates). Nuclear lysates were treated with each

708   compound at 100 μM for 1 hr, followed by HDBIA or LDBIA labeling at 500 μM for 1 hr. **B**.

709   Structures of electrophilic fragments used in the nuclear ABPP experiment. **C**. Number of

710   cysteine sites quantified across two or four replicates in the hyperplexed nuclear ABPP

711   experiment. **D**. Fraction of proteins occupied by at least one electrophilic compound at ≥75% on

712   one or more cysteine sites. **E**. Correlation of replicate measurements derived from "heavy" and

713   "light" biotinylated site abundances as log2 ratios of compound treatment over DMSO control. **F**.

714   Heatmap of site occupancy measured as log2(Treatment/DMSO) across treatments. Summary

715   table includes compound hit rate (fraction of sites occupied by the compound at ≥75%,

716   pvalue≤0.05) and intrinsic compound reactivity measured as the amount of time it takes for 50%

717   of 1 µM compound to be consumed during incubation with 5 mM cysteine ($t_{1/2}$).

718

Fig 4. Characterization of ligandable cysteine sites in the nuclear proteome

720   **A**. For the whole-cell lysate ABPP experiment, MCF7 lysates were incubated with 500µM of

721   each electrophilic compound for 1 hr, followed by labeling with 500 µM LDBIA probe for 1 hr.

722   Nuclear lysates were treated with each compound at 100 µM for 1 hr as described in Fig. 3.

723   Venn diagrams showing the number of proteins (left panel) and unique cysteine sites (right

724   panel) identified in each experiment. **B**. Gene Ontology Cellular Component analysis of proteins

725   identified in the nuclear fractions. **C**. Heatmaps of site occupancies of proteins in ubiquitination

726   pathway observed in the nuclear and whole-cell ABPP experiment with indicated electrophilic

727   fragments. **D**. Site occupancy of selected transcription regulators uniquely identified in the

728   nuclear ABPP experiment. **E**. Localization of Cys128 in the "Orange" domain of HES1

729   transcription repressor. **F** and **G**. MS1 and MS2 spectra of HES1 Cys128 peptide modified with

730   HDBIA or LDBIA. **H**. Relative abundance of HES1 Cys128 peptide modified with HDBIA or

731   LDBIA between conditions.

Supplemental Fig.1

733   Structures of LDBIA and HDBIA reactive probes.

734   **Supplemental Fig.2**

735   **A**. and **B**. Precursor ion intensity of biotinylated cys peptides (A) and TMT S/N distribution from

736   all MS3 spectra (B) between runs acquired with SPS-MS3 without RTS and with indicated RTS

737   method (two individual experiments), data from experiment in Fig. 1C. **C**. Number of biotinylated

738   cysteine PSMs identified using indicated acquisition method before peptide FDR filter ("All

739   scans"), after RTS peptide FDR filter ("Passed online filtering"), after offline search and peptide

740   FDR filter ("Passed offline filtering"), and after RTS and offline FDR filter ("Passed online and

741   offline filtering"). Data from experiment in Fig. 1C.

742   **Supplemental Fig.3**

743   Lysates from C2C12 "heavy" and "light" SILAC cell culture were labeled with 500 µM LDBIA

744   probe for 1 hr. After digestion and streptavidin enrichment, biotinylated peptides were analyzed

745   using indicated acquisition methods. Number of biotinylated cysteine PSMs identified between

746   two TMT runs ("heavy" SILAC and "light" SILAC) is shown from two individual experiments.

747   **Supplemental Fig.4**

748   MCF7 cells were fractionated into nuclear and cytosolic fraction using REAP protocol[17].

749   Fractions were analyzed by Western Blotting with anti-tubulin and anti-histone H3 antibodies as

750   cytosolic and nuclear markers, respectively.

751   **Supplemental Fig.5**

752   **A**. Test compounds (1 µM) and cysteine (5 mM) were mixed sequentially to start the reaction.

753   Aliquots were taken at indicated timepoints and quantified via LC-MS/MS. Percentage

754   remaining was calculated using peak area ratios normalized to the 0 minutes time-point sample

755 (% $remaining = 100 * e^{-kreact * 5\,mM * time}$). **B**. Incubation without cysteine was also carried out

756 to check compound stability in the buffer.

Supplemental Fig.6

758 Detailed schematic PairQuant method within inSeqAPI.

759

References

761 (1) Li, J.; Cai, Z.; Bomgarden, R. D.; Pike, I.; Kuhn, K.; Rogers, J. C.; Roberts, T. M.; Gygi, S.

762 P.; Paulo, J. A. TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for

763 Sample Multiplexing. *J Proteome Res* **2021**, *20* (5), 2964–2972.

764 https://doi.org/10.1021/acs.jproteome.1c00168.

765 (2) McAlister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B.

766 K.; Rad, R.; Haas, W.; Gygi, S. P. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed

767 Detection of Differential Expression across Cancer *Cell* Line Proteomes. *Anal Chem* **2014**, *86*

768 (14), 7150–7158. https://doi.org/10.1021/ac502040v.

769 (3) Erickson, B. K.; Mintseris, J.; Schweppe, D. K.; Navarrete-Perea, J.; Erickson, A. R.;

770 Nusinow, D. P.; Paulo, J. A.; Gygi, S. P. Active Instrument Engagement Combined with a Real-

771 Time Database Search for Improved Performance of Sample Multiplexing Workflows. J

772 Proteome Res **2019**, *18* (3), 1299–1306. https://doi.org/10.1021/acs.jproteome.8b00899.

773 (4) Schweppe, D. K.; Eng, J. K.; Yu, Q.; Bailey, D.; Rad, R.; Navarrete-Perea, J.; Huttlin, E. L.;

774 Erickson, B. K.; Paulo, J. A.; Gygi, S. P. Full-Featured, Real-Time Database Searching Platform

775 Enables Fast and Accurate Multiplexed Quantitative Proteomics. J Proteome Res **2020**, *19* (5),

776 2026–2034. https://doi.org/10.1021/acs.jproteome.9b00860.

777 (5) Kuljanin, M.; Mitchell, D. C.; Schweppe, D. K.; Gikandi, A. S.; Nusinow, D. P.; Bulloch, N. J.;

Vinogradova, E. V.; Wilson, D. L.; Kool, E. T.; Mancias, J. D.; Cravatt, B. F.; Gygi, S. P.

Reimagining High-Throughput Profiling of Reactive Cysteines for Cell-Based Screening of Large

Electrophile Libraries. *Nat Biotechnol* **2021**, *39* (5), 630–641. https://doi.org/10.1038/s41587-

020-00778-3.

(6) Bender, A. T.; Gardberg, A.; Pereira, A.; Johenson, T.; Wu, Y.; Grenningloh, R.; Head, J.;

Morandi, F.; Haselmayer, P.; Liu-Bujalski, L. The Ability of Btk Inhibitors to Sequester Y551 and

Prevent Phosphorylation Determines Potency for Inhibition of FcR but Not B Cell Receptor

Signaling. *Mol. Pharmacol.* **2017**, *91* (3), mol.116.107037.

https://doi.org/10.1124/mol.116.107037.

(7) Wissner, A.; Mansour, T. S. The Development of HKI-272 and Related Compounds for the

Treatment of Cancer. *Arch. Pharm.* **2008**, *341* (8), 465–477.

https://doi.org/10.1002/ardp.200800009.

(8) Janes, M. R.; Zhang, J.; Li, L.-S.; Hansen, R.; Peters, U.; Guo, X.; Chen, Y.; Babbar, A.;

Firdaus, S. J.; Darjania, L.; Feng, J.; Chen, J. H.; Li, S.; Li, S.; Long, Y. O.; Thach, C.; Liu, Y.;

Zarieh, A.; Ely, T.; Kucharski, J. M.; Kessler, L. V.; Wu, T.; Yu, K.; Wang, Y.; Yao, Y.; Deng, X.;

Zarrinkar, P. P.; Brehmer, D.; Dhanak, D.; Lorenzi, M. V.; Hu-Lowe, D.; Patricelli, M. P.; Ren, P.;

Liu, Y. Targeting KRAS Mutant *Cancers* with a Covalent G12C-Specific Inhibitor. Cell **2018**, *172*

(3), 578-589.e17. https://doi.org/10.1016/j.cell.2018.01.006.

(9) Boike, L.; Henning, N. J.; Nomura, D. K. Advances in Covalent Drug Discovery. *Nat Rev

Drug Discov* **2022**, 1–18. https://doi.org/10.1038/s41573-022-00542-z.

(10) Vinogradova, E. V.; Zhang, X.; Remillard, D.; Lazar, D. C.; Suciu, R. M.; Wang, Y.; Bianco,

G.; Yamashita, Y.; Crowley, V. M.; Schafroth, M. A.; Yokoyama, M.; Konrad, D. B.; Lum, K. M.;

Simon, G. M.; Kemper, E. K.; Lazear, M. R.; Yin, S.; Blewett, M. M.; Dix, M. M.; Nguyen, N.;

Shokhirev, M. N.; Chin, E. N.; Lairson, L. L.; Melillo, B.; Schreiber, S. L.; Forli, S.; Teijaro, J. R.;

Cravatt, B. F. An Activity-Guided Map of Electrophile-Cysteine Interactions in Primary Human T

Cells. Cell 2020, *182* (4), 1009-1026.e29. https://doi.org/10.1016/j.cell.2020.07.001.

804   (11) Bar-Peled, L.; Kemper, E. K.; Suciu, R. M.; Vinogradova, E. V.; Backus, K. M.; Horning, B.

805   D.; Paul, T. A.; Ichu, T.-A.; Svensson, R. U.; Olucha, J.; Chang, M. W.; Kok, B. P.; Zhu, Z.; Ihle,

806   N. T.; Dix, M. M.; Jiang, P.; Hayward, M. M.; Saez, E.; Shaw, R. J.; Cravatt, B. F. Chemical

807   Proteomics Identifies Druggable Vulnerabilities in a Genetically Defined Cancer. Cell **2017**, *171*

808   (3), 696-709.e23. https://doi.org/10.1016/j.cell.2017.08.051.

809   (12) Nusinow, D. P.; Szpyt, J.; Ghandi, M.; Rose, C. M.; McDonald, E. R.; Kalocsay, M.; Jané-

810   Valbuena, J.; Gelfand, E.; Schweppe, D. K.; Jedrychowski, M.; Golji, J.; Porter, D. A.; Rejtar, T.;

811   Wang, Y. K.; Kryukov, G. V.; Stegmeier, F.; Erickson, B. K.; Garraway, L. A.; Sellers, W. R.;

812   Gygi, S. P. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. Cell 2020, *180* (2),

813   387-402.e16. https://doi.org/10.1016/j.cell.2019.12.023.

814   (13) Brenes, A.; Hukelmann, J.; Bensaddek, D.; Lamond, A. I. Multibatch TMT Reveals False

815   Positives, Batch Effects and Missing Values*. *Mol. Cell. Proteom.* **2019**, *18* (10), 1967–1980.

816   https://doi.org/10.1074/mcp.ra119.001472.

817   (14) Rose, C. M.; Venkateshwaran, M.; Volkening, J. D.; Grimsrud, P. A.; Maeda, J.; Bailey, D.

818   J.; Park, K.; Howes-Podoll, M.; Os, D. den; Yeun, L. H.; Westphall, M. S.; Sussman, M. R.; Ané,

819   J.-M.; Coon, J. J. Rapid Phosphoproteomic and Transcriptomic Changes in the Rhizobia-

820   Legume Symbiosis*. Mol. Cell. Proteom. **2012**, *11* (9), 724–744.

821   https://doi.org/10.1074/mcp.m112.019208.

822   (15) Dephoure, N.; Gygi, S. P. Hyperplexing: A Method for Higher-Order Multiplexed

823   Quantitative Proteomics Provides a Map of the Dynamic Response to Rapamycin in Yeast. *Sci*

824   *Signal* **2012**, *5* (217), rs2. https://doi.org/10.1126/scisignal.2002548.

825   (16) Zecha, J.; Gabriel, W.; Spallek, R.; Chang, Y.-C.; Mergner, J.; Wilhelm, M.; Bassermann,

826   F.; Kuster, B. Linking Post-Translational Modifications and Protein Turnover by Site-Resolved

827   Protein Turnover Profiling. *Nat Commun* **2022**, *13* (1), 165. https://doi.org/10.1038/s41467-021-

828   27639-0.

829   (17) Suzuki, K.; Bose, P.; Leong-Quong, R. Y.; Fujita, D. J.; Riabowol, K. REAP: A Two Minute

830   Cell Fractionation Method. *Bmc Res Notes* **2010**, *3* (1), 294. https://doi.org/10.1186/1756-0500-

831   3-294.

832   (18) Kirkpatrick, D. S.; Bustos, D. J.; Dogan, T.; Chan, J.; Phu, L.; Young, A.; Friedman, L. S.;

833   Belvin, M.; Song, Q.; Bakalarski, C. E.; Hoeflich, K. P. Phosphoproteomic Characterization of

834   DNA Damage Response in Melanoma Cells Following MEK/PI3K Dual Inhibition. *Proc National*

835   *Acad Sci* **2013**, *110* (48), 19426–19431. https://doi.org/10.1073/pnas.1309473110.

836   (19) Kohler, D.; Staniak, M.; Tsai, T.-H.; Huang, T.; Shulman, N.; Bernhardt, O. M.; MacLean, B.

837   X.; Nesvizhskii, A. I.; Reiter, L.; Sabido, E.; Choi, M.; Vitek, O. MSstats Version 4.0: Statistical

838   Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with

839   Chromatography-Based Quantification at Scale. *J. Proteome Res.* **2023**, *22* (5), 1466–1482.

840   https://doi.org/10.1021/acs.jproteome.2c00834.

841   (20) Huang, T.; Choi, M.; Tzouros, M.; Golling, S.; Pandya, N. J.; Banfai, B.; Dunkley, T.; Vitek,

842   O. MSstatsTMT: Statistical Detection of Differentially Abundant Proteins in Experiments with

843   Isobaric Labeling and Multiple Mixtures. *Mol Cell Proteomics* **2020**, *19* (10), 1706–1723.

844   https://doi.org/10.1074/mcp.ra*120.002105.

845   (21) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and Integrative Analysis of

846   Large Gene Lists Using DAVID Bioinformatics Resources. *Nat Protoc* **2009**, *4* (1), 44–57.

847   https://doi.org/10.1038/nprot.2008.211.

848   (22) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Bioinformatics Enrichment Tools: Paths

849   toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res* **2009**,

850   *37* (1), 1–13. https://doi.org/10.1093/nar/gkn923.

851   (23) Zhang, X.; Crowley, V. M.; Wucherpfennig, T. G.; Dix, M. M.; Cravatt, B. F. Electrophilic

852   PROTACs That Degrade Nuclear Proteins by Engaging DCAF16. *Nat Chem Biol* **2019**, *15* (7),

853   737–746. https://doi.org/10.1038/s41589-019-0279-5.

854   (24) Cao, J.; Yan, Q. Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage

855   Response, and Cancer. *Frontiers Oncol* **2012**, 2, 26. https://doi.org/10.3389/fonc.2012.00026.

856   (25) Jeusset, L. M.-P.; McManus, K. J. Ubiquitin Specific Peptidase 22 Regulates Histone H2B

857   Mono-Ubiquitination and Exhibits Both Oncogenic and Tumor Suppressor Roles in Cancer.

858   Cancers 2017, 9 (12), 167. https://doi.org/10.3390/cancers9120167.

859   (26) Zheng, J.; Chen, C.; Guo, C.; Caba, C.; Tong, Y.; Wang, H. The Pleiotropic Ubiquitin-

860   Specific Peptidase 16 and Its Many Substrates. Cells 2023, 12 (6), 886.

861   https://doi.org/10.3390/cells12060886.

862   (27) Tasaki, T.; Sriram, S. M.; Park, K. S.; Kwon, Y. T. The N-End Rule Pathway. *Annu Rev*

863   *Biochem* **2012**, *81* (1), 261–289. https://doi.org/10.1146/annurev-biochem-051710-093308.

864   (28) An, J. Y.; Kim, E.-A.; Jiang, Y.; Zakrzewska, A.; Kim, D. E.; Lee, M. J.; Mook-Jung, I.;

865   Zhang, Y.; Kwon, Y. T. UBR2 Mediates Transcriptional Silencing during Spermatogenesis via

866   Histone Ubiquitination. Proc National Acad Sci **2010**, *107* (5), 1912–1917.

867   https://doi.org/10.1073/pnas.0910267107.

868   (29) Lee, P. C. W.; Sowa, M. E.; Gygi, S. P.; Harper, J. W. Alternative Ubiquitin

869   Activation/Conjugation Cascades Interact with N-End Rule Ubiquitin Ligases to Control

870   Degradation of RGS Proteins. Mol Cell **2011**, *43* (3), 392–405.

871   https://doi.org/10.1016/j.molcel.2011.05.034.

872   (30) Zhong, Q.; Chen, C.-F.; Li, S.; Chen, Y.; Wang, C.-C.; Xiao, J.; Chen, P.-L.; Sharp, Z. D.;

873   Lee, W.-H. Association of BRCA1 with the HRad50-HMre11-P95 Complex and the DNA

874   Damage Response. *Science* **1999**, *285* (5428), 747–750.

875   https://doi.org/10.1126/science.285.5428.747.

876   (31) Kirsch, K.; Zeke, A.; Tőke, O.; Sok, P.; Sethi, A.; Sebő, A.; Kumar, G. S.; Egri, P.; Póti, Á.

877   L.; Gooley, P.; Peti, W.; Bento, I.; Alexa, A.; Reményi, A. Co-Regulation of the Transcription

878   Controlling ATF2 Phosphoswitch by JNK and P38. *Nat. Commun.* **2020**, *11* (1), 5769.

879   https://doi.org/10.1038/s41467-020-19582-3.

880   (32) Liu, Y.; Ao, X.; Zhou, X.; Du, C.; Kuang, S. The Regulation of PBXs and Their Emerging

881   Role in Cancer. *J Cell Mol Med* **2022**, *26* (5), 1363–1379. https://doi.org/10.1111/jcmm.17196.

882   (33) Bantscheff, M.; Hopf, C.; Savitski, M. M.; Dittmann, A.; Grandi, P.; Michon, A.-M.; Schlegl,

883   J.; Abraham, Y.; Becher, I.; Bergamini, G.; Boesche, M.; Delling, M.; Dümpelfeld, B.; Eberhard,

884   D.; Huthmacher, C.; Mathieson, T.; Poeckel, D.; Reader, V.; Strunk, K.; Sweetman, G.; Kruse,

885   U.; Neubauer, G.; Ramsden, N. G.; Drewes, G. Chemoproteomics Profiling of HDAC Inhibitors

886   Reveals Selective Targeting of HDAC Complexes. Nat Biotechnol 2011, *29* (3), 255–265.

887   https://doi.org/10.1038/nbt.1759.

888   (34) Sun, H.; Ghaffari, S.; Taneja, R. BHLH-Orange Transcription Factors in Development and

889   Cancer. *Transl Oncogenomics* **2007**, *2*, 107–120. https://doi.org/10.4137/tog.s436.

890   (35) Ibarrola, N.; Molina, H.; Iwahori, A.; Pandey, A. A Novel Proteomic Approach for Specific

891   Identification of Tyrosine Kinase Substrates Using [13C]Tyrosine*. *J Biol Chem* **2004**, *279* (16),

892   15805–15813. https://doi.org/10.1074/jbc.m311714200.

893   (36) King, E. A.; Cho, Y.; Hsu, N. S.; Dovala, D.; McKenna, J. M.; Tallarico, J. A.; Schirle, M.;

894   Nomura, D. K. Chemoproteomics-Enabled Discovery of a Covalent Molecular Glue Degrader

895   Targeting NF-KB. *Cell Chem Biol* **2023**. https://doi.org/10.1016/j.chembiol.2023.02.008.

896   (37) Luo, M.; Spradlin, J. N.; Boike, L.; Tong, B.; Brittain, S. M.; McKenna, J. M.; Tallarico, J. A.;

897   Schirle, M.; Maimone, T. J.; Nomura, D. K. Chemoproteomics-Enabled Discovery of Covalent

898   RNF114-Based Degraders That Mimic Natural Product Function. Cell Chem Biol **2021**, *28* (4),

899   559-566.e15. https://doi.org/10.1016/j.chembiol.2021.01.005.

900   (38) Henning, N. J.; Manford, A. G.; Spradlin, J. N.; Brittain, S. M.; Zhang, E.; McKenna, J. M.;

901   Tallarico, J. A.; Schirle, M.; Rape, M.; Nomura, D. K. Discovery of a Covalent FEM1B Recruiter

902   for Targeted Protein Degradation Applications. *J Am Chem Soc* **2022**, *144* (2), 701–708.

903   https://doi.org/10.1021/jacs.1c03980.

904   (39) Tao, Y.; Remillard, D.; Vinogradova, E. V.; Yokoyama, M.; Banchenko, S.; Schwefel, D.;

905   Melillo, B.; Schreiber, S. L.; Zhang, X.; Cravatt, B. F. Targeted Protein Degradation by

906   Electrophilic PROTACs That Stereoselectively and Site-Specifically Engage DCAF1. J Am

907   Chem Soc 2022, 144 (40), 18688–18699. https://doi.org/10.1021/jacs.2c08964.

908   (40) Liang, Y.; Truong, T.; Saxton, A. J.; Boekweg, H.; Payne, S. H.; Ry, P. M. V.; Kelly, R. T.

909   HyperSCP: Combining Isotopic and Isobaric Labeling for Higher Throughput Single-Cell

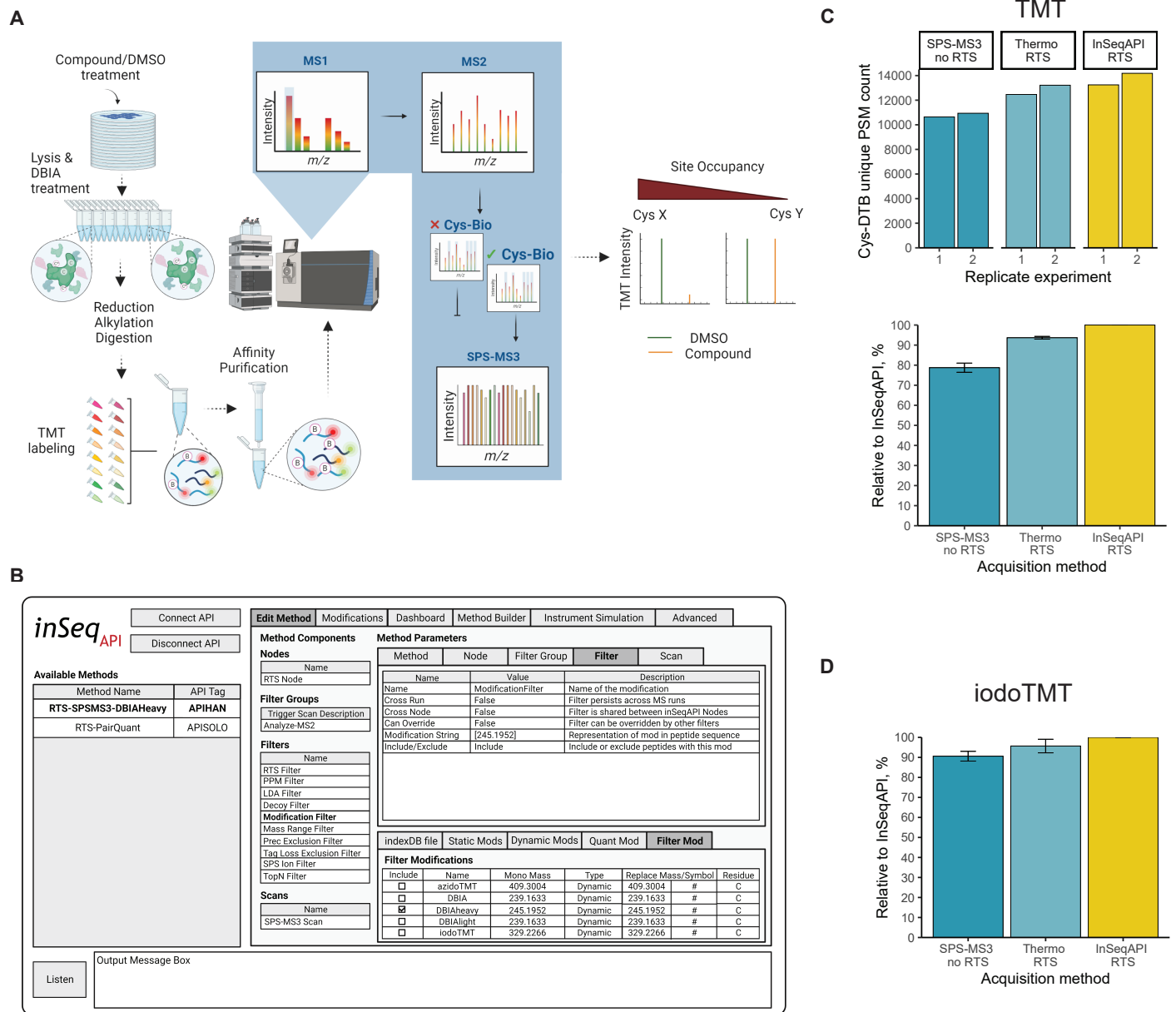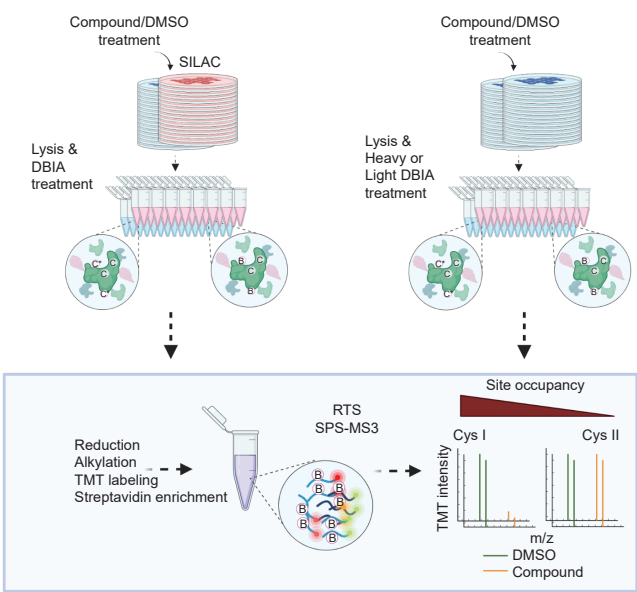910   Proteomics. *Anal. Chem.* **2023**, *95* (20), 8020–8027.

911   https://doi.org/10.1021/acs.analchem.3c00906.

912

Figure 1

Figure 2

Figure 3

A

0.1% NP40-PBS → 10sec → Crude nuclei (0.1% NP40 wash) → 10sec → Nuclei → Lysis in PBS → Nuclear ABPP with hyperplexing

B
KB05, G59, KB02, G51, G92, G35, KB03, G84

C Cys-DTB counts — Replicates 2, 3, 4; Observed PSMs 96.7%, Quantified PSMs 96.6%, Quantified Sites 96.7%

D Occupancy ≥ 75%, < 75%

E Treatment: G35, G51, G59, G84, G92, KB02, KB03, KB05
Compound/DMSO,log2 (LDBIA) vs Compound/DMSO,log2 (HDBIA)
r = 0.88, p < 0.05
r = 0.83, p < 0.05
r = 0.8, p < 0.05
r = 0.93, p < 0.05
r = 0.83, p < 0.05
r = 0.23, p < 0.05
r = 0.92, p < 0.05
r = 0.88, p < 0.05
>75% occupancy

F Treatment/DMSO,log2FC

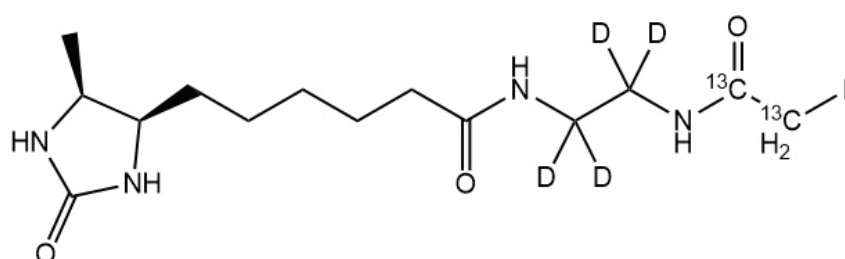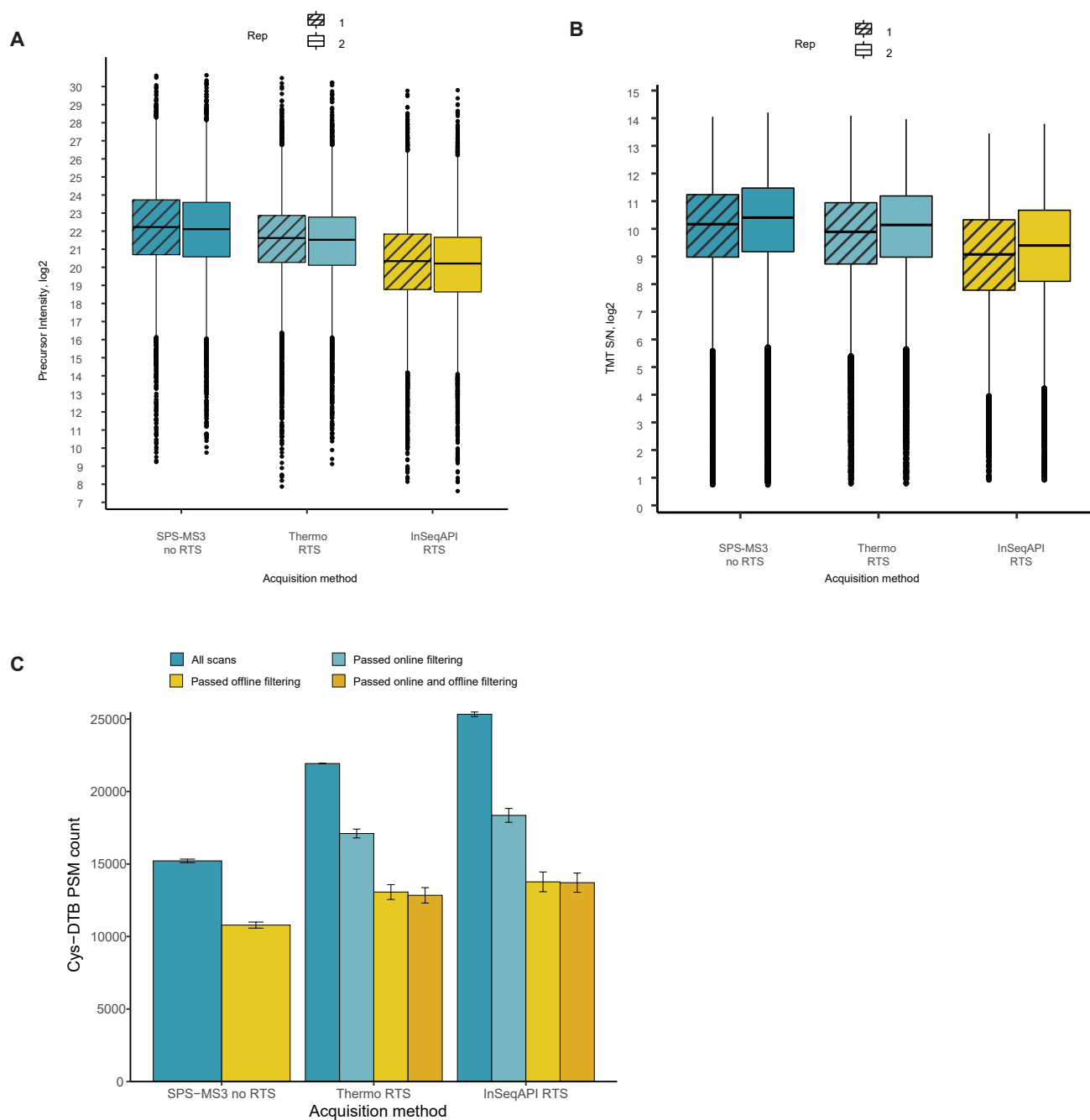| Compound	G51	G84	G35	G92	G59	KB05	KB02	KB03
Hit rate, %	28.42	23.30	1.38	6.99	0.08	4.24	0.94	2.00
Cys Reactivity (t1/2, min)	<5	<5	<5	<5	7.8	<5	5.3	12.4
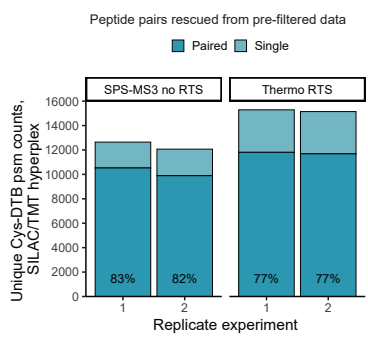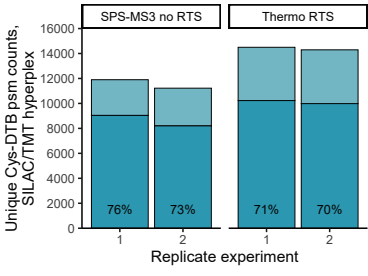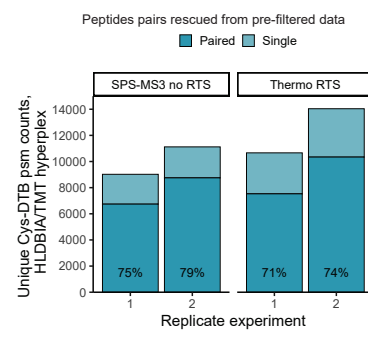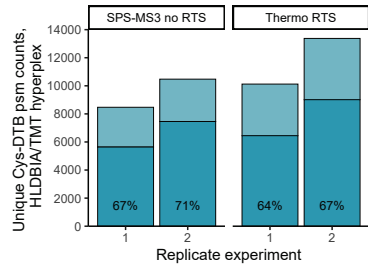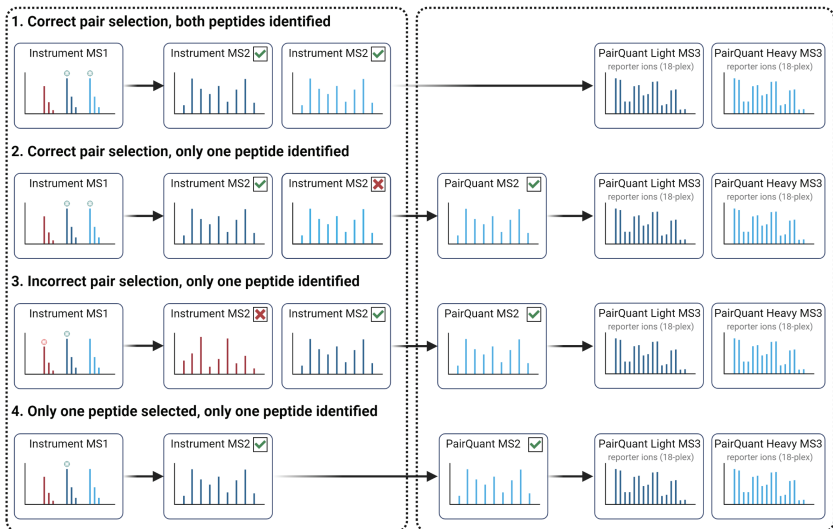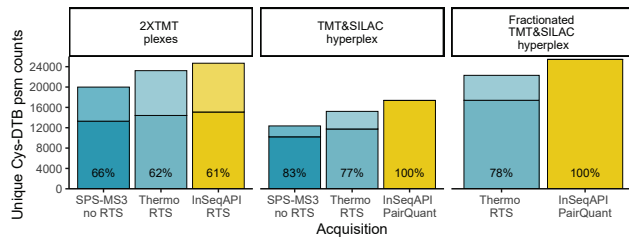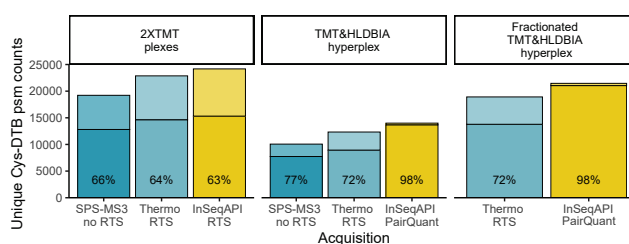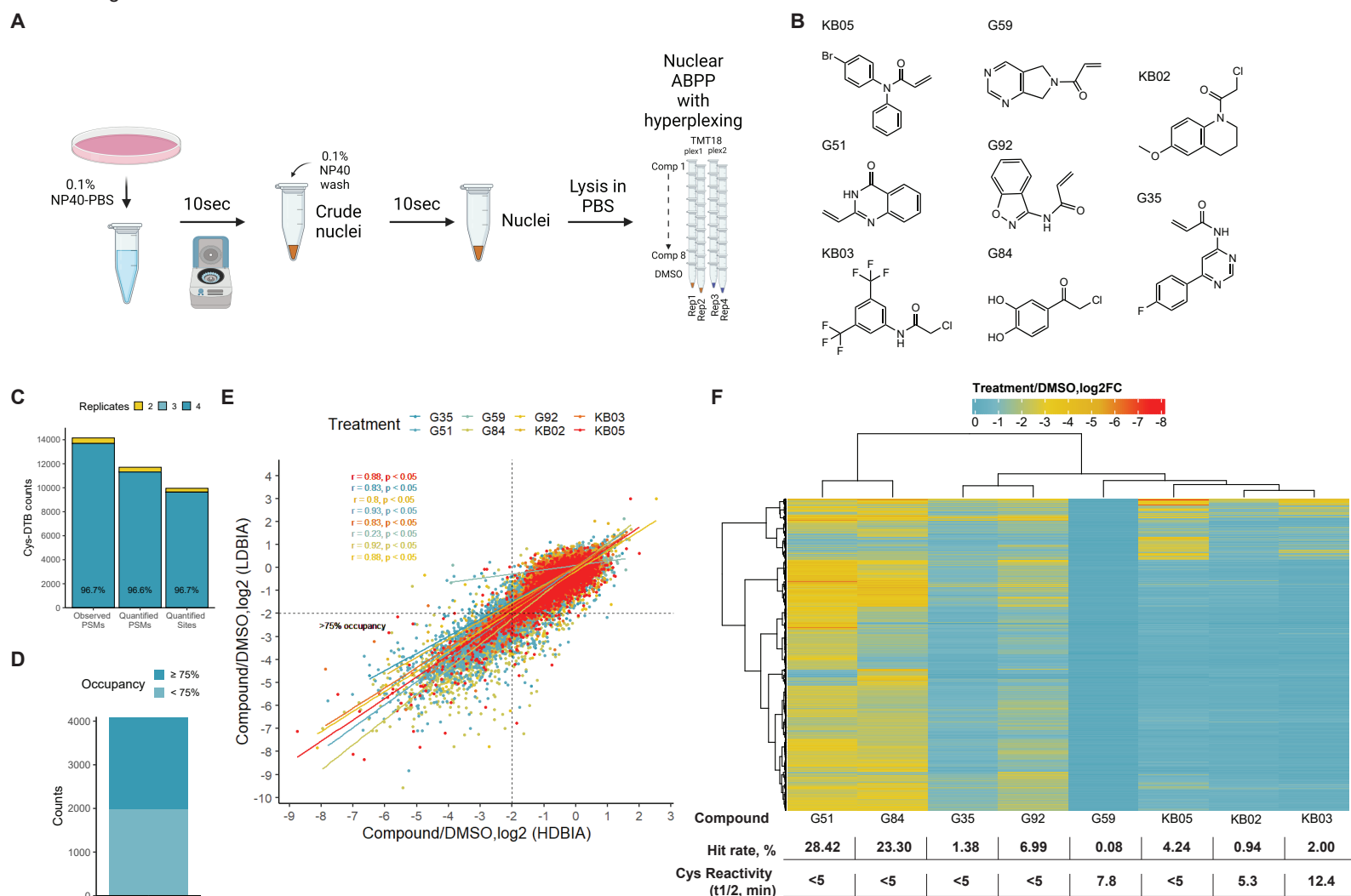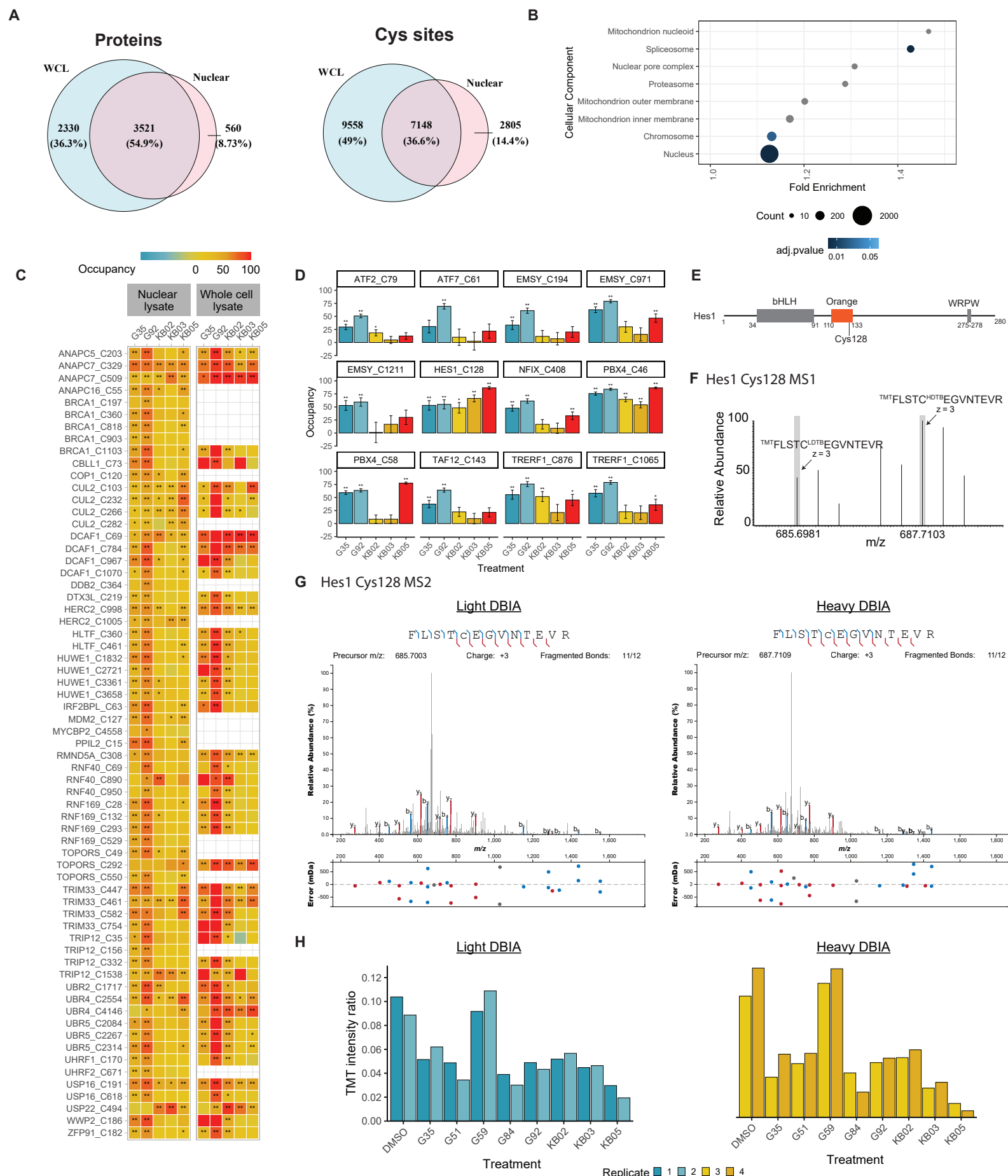
Figure 4

Supplemental Figure 1
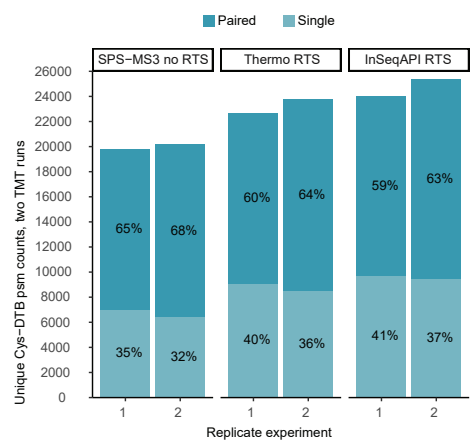
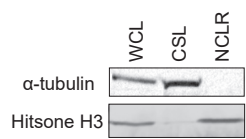Light Desthiobiotin Iodoacetamide
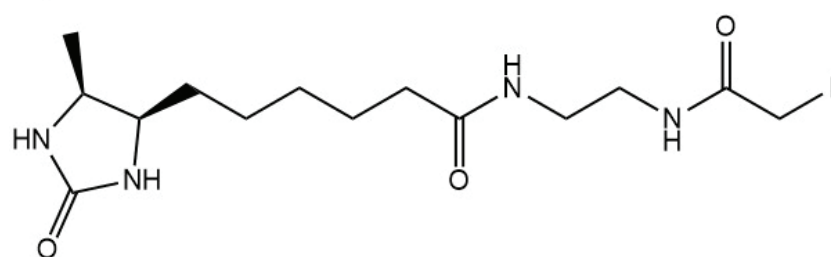
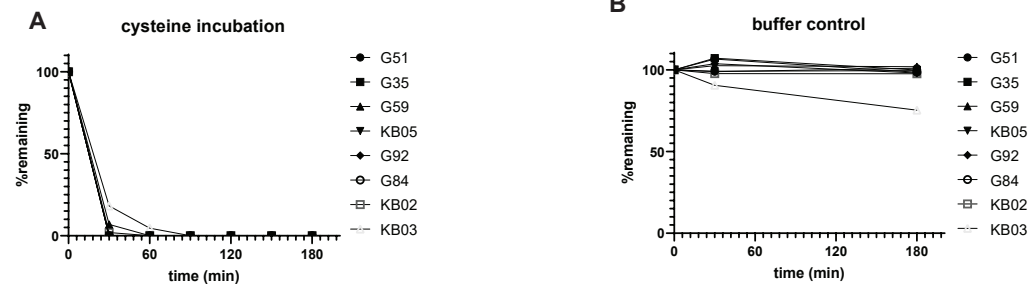
Heavy Desthiobiotin Iodoacetamide

Supplemental Figure 2

Supplemental Figure 3

Supplemental Figure 4

Supplemental Figure 5

Supplemental Figure 6