PulseDIA: in-depth data independent acquisition mass spectrometry using enhanced gas phase fractionation

Xue Cai [1,2], Weigang Ge [1,2], Xiao Yi [1,2], Rui Sun[1,2], Jiang Zhu [3], Cong Lu[3], Ping Sun[4], Tiansheng Zhu[1,2], Guan Ruan[1,2], Chunhui Yuan [1,2], Shuang Liang [1,2], Mengge Lv [1,2], Shiang Huang [3], Yi Zhu[1,2*], Tiannan Guo[1,2*]

1. Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, China, Hangzhou, Zhejiang, China
2. Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, China, Hangzhou, Zhejiang, China
3. Center for Stem Cell Research and Application, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China
4. Department of Hepatobiliary Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

***** co-correspondence: zhuyi@westlake.edu.cn; guotiannan@westlake.edu.cn

ABSTRACT

An inherent bottleneck of data independent acquisition (DIA) analysis by Orbitrap-based mass spectrometers is the relatively large window width due to the relatively slow scanning rate compared to TOF. Here we present a novel gas phase separation and MS acquisition method called PulseDIA-MS, which improves the specificity and sensitivity of Orbitrap-based DIA analysis. This is achieved by dividing the ordinary DIA-MS analysis covering the entire mass range into multiple injections for DIA-MS analyses with complementary windows. Using standard HeLa digests, the PulseDIA method identified 61,078 peptides from 7,308 protein groups with ten MS injections of 30 min LC gradient. The PulseDIA scheme containing two complementary windows led to the highest gain of peptide and protein identifications per time unit compared to the conventional 30 min DIA method. We further applied the method to profile the proteome of 18 cholangiocarcinoma (CCA) tissue samples (benign and malignant) from nine patients. PulseDIA identified 6,063 protein groups in these CCA samples, with 23% increase of protein identifications, compared to the classical DIA method. The missing value for protein matrix dropped by 25% with PulseDIA acquisition. 618 proteins were significantly dysregulated in tumorous CCA samples. Together, we presented and benchmarked an alternative DIA method with higher sensitivity and lower missing rate.

INTRODUCTION

Mass spectrometry (MS)-based quantitative proteomics is increasingly applied to identify dysregulated proteins in clinical specimens, facilitating tumor diagnosis and prognosis.[1-3] DIA emerges recently as a significant discovery proteomics method enabling high-throughput and reproducible single-shot analysis of complex proteomes including those from clinical specimens.[3-5] The DIA method fragments a pile of precursor ions after the MS1 scan in an incremental and recursive fashion (called DIA windows), then records the fragment signals of all of these peptide precursors in the respective MS2 scans for peptide identification and quantification.[4]

Several DIA-MS methods have been reported, including all-ion fragmentation (AIF)[6] and MS^{E} [7] . In AIF and MS^{E}, all precursor ions are analyzed in one window, collecting MS1 and MS2 spectra alternatively. The thus obtained MS2 spectra are highly convoluted and could not be effectively analyzed with conventional algorithms for data-dependent acquisition (DDA) proteomics. The emerging DIA-MS reduces the proteome complexity by making use of gas-phase isolation of flying peptide precursors in multiple windows with certain mass-charge (*m/z*) width. However, the sensitivity and specificity of DIA-MS are therefore limited by the window width.

When the number of isolation windows reaches the limit, gas-phase separation of peptide precursor ions could improve the sensitivity. A DIA-MS method called Precursor Acquisition Independent From Ion Count (PAcIFIC)[8,9] acquires tandem mass spectra with every 1.5 *m/z* segment, leading to increased peptide and protein identifications. However, it takes several days to cycle through an entire *m/z* range and lacks quantitative information due to absence of precursor ion acquisitions. Another DIA method called multiplexing strategy (MSX) divides the peptide precursors in gas-phase into 100 windows.[10] To demultiplex the spectra, five separate 4-*m/z* isolation windows are analyzed in each scan of MSX DIA, and 20 scans are performed to cover the *m/z* range of 500-900. Both precursor ion selectivity (5-fold higher than the classical DIA) and fragment-ion spectra quality are improved under the narrower window width. However, MSX DIA doesn't increase the mean ion injection time. Therefore, the peptide and protein identification are not significantly improved.

Here, we present an alternative gas-phase separation PulseDIA-MS method, in which multiple DIA runs are performed to increase the number of windows and reduce the window width to improve specificity and sensitivity. We demonstrated its applicability in identifying dysregulated proteins from cholangiocarcinoma (CCA), a rare

malignant tumor composed of cells that resemble those of the biliary tract.[11] CCA usually eludes from early diagnosis due to hardly discernable symptoms, leading to a poor prognosis and high morbidity. The serum marker carbohydrate antigen 19-9 (CA19-9) and carcinoembryogenic antigen (CEA) are sometimes used to detect CCA, however they are neither sensitive nor specific.[12] The need for discovering proteomic biomarkers for early diagnosis of CCA is pressing. Several proteomic studies of cell lines, bile fluid and sera have been reported.[13-15] Juliet P et al. used DDA-MS to identify several promising biomarkers (such as ANXA1, ANXA10, ANXA13) of CCA by proteomic analysis of microdissected cells extracted from 11 tissue samples collected from 11 CCA patients and verification by immunohistochemistry (IHC) of other 83 samples.[16]

In this study, we performed proteomic analysis of 18 CCA tissue samples by the thus developed PulseDIA-MS method. 6,063 protein groups were quantified across all CCA samples and 618 proteins were found to be dysregulated significantly. Ingenuity pathway analysis (IPA) revealed that they were highly enriched in pathways such as RXR activation, xenobiotic metabolism signaling, AMPK signaling and PI3K/AKT signaling pathway.

EXPERIMENTAL SECTION

Samples

HeLa Protein Digest Standard peptides were purchased from the Thermo Fisher Scientific™ (Product number 88329, Rockford, USA), and stored at -20□ until analysis. All cell lines were provided by Dr Chenhuan Yu from Zhejiang Academy of Medical Sciences, Hangzhou, China. BT549, Hs578T, ZR75-1 and MDA-MB-231 cells were cultured in DMEM/F-12 medium (Cat No. 01-172-1ACS, Biological industries, Cromwell, CT, USA). MDA-MB-468 was maintained in L15 medium (Cat No. 01-115-1A, Biological industries, Cromwell, CT, USA). T47D was adapted in the DMEM medium (Cat No. 06-1055-57-1ACS, HyClone, Biological industries, Cromwell, CT, USA). MCF7, MX-1 and SK-BR-3 cells were cultured in 1640 medium (Cat No. 01-100-1ACS, Biological industries, Cromwell, CT, USA) supplemented with 10% fetal bovine serum (Cat No. 04-001AUS-1A, Biological industries, Cromwell, CT, USA) and penicillin-streptomycin (Cat No. 03-031-5B, HyClone, General Electric, USA) at 37 □ with 5% CO_2.

18 tissue samples from nine CCA patients were collected within one hour after hepatectomy, then snap frozen and stored at -80□. A pair of tumorous tissue and the non-tumorous tissue from an adjacent region around the tumor as determined by histomorphological examination were collected from the same patient. Ethical permission was approved by Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. This study was also approved by Ethics Committee of Westlake University.

Peptides extraction

HeLa Protein Digest Standard peptides were redissolved with HPLC-grade water containing 0.1% formic acid (FA) and 2% acetonitrile (ACN) at the final concentration of 0.25 μg/ul. Cell pellets and tissue samples were lysed and digested using PCT technology as described previously with some modifications.[17] Briefly, the cells were lysed with 50 μL lysis buffer containing 6 M urea (Sigma) and 2 M thiourea (Sigma) in 100 mM ammonium bicarbonate in a PCT -MicroTube. The PCT-assisted lysis was performed under the scheme with 60 cycles with each cycle consisting of 30 s at 45,000 p.s.i. and 10 s at ambient pressure at 30 □. PCT-assisted digestion was performed using lysC (1:40) and trypsin (1:50) in a barocycler under the PCT scheme for 45 cycles and 60 cycles, respectively. Each cycle contains 50 s at 20,000 p.s.i. and 10 s at ambient pressure at 30 □. The thus generated peptides were acidified with trifluoroacetic acid (TFA) to pH 2-3, and cleaned with the Nest Group C18 columns (17-170 μg capacity, Part No. HEM S18V, MA, USA) prior to MS analysis.

Data Acquisition with PulseDIA-MS

With the PulseDIA method, the MS1 was performed over a *m/z* range of 390-1210 with the resolution at 60,000, AGC target of 3e6, and the maximum ion injection time of 80 ms. The MS2 was performed with the resolution at 30,000, AGC target of 1e6, and the maximum ion injection time of 50 ms.

The PulseDIA acquisition of HeLa peptides was performed on a nano-flow HPLC system (EASY-nLC™ 1200 System, Thermo Fisher Scientific™, San Jose, USA) coupled to a QE HF-X mass spectrometer (Q Exactive HF-X hybrid Quadrupole-Orbitrap, Thermo Fisher Scientific™, San Jose, USA). For each PulseDIA acquisition, 0.5 μg of peptides was injected and separated across a 30 min LC gradient (from 8% to 40% buffer B) at a flowrate of 300 nl/min (precolumn, 3μm, 100 Å, 20mm*75μm i.d.; analytical column, 1.9um, 120 Å, 150mm*75um i.d.). Buffer A was HPLC-grade water containing 0.1% FA, and buffer B was 80%ACN, 20%H_2O containing 0.1%FA.

The PulseDIA acquisition of peptides from other cell lines and CCA tissues was performed on a nano-flow HPLC system (DIONEX UltiMate 3000 RSLCnano System, Thermo Fisher Scientific™, San Jose, USA) coupled

to a QE HF mass spectrometer (Q Exactive Hybrid Quadrupole-Orbitrap, Thermo Fisher Scientific™, San Jose, USA). For each PulseDIA acquisition, 0.5 μg of peptides was injected and separated across a 30 min LC gradient with the same settings as described above.

PulseDIA data analysis

PulseDIA raw files were converted into mzXML format using msconvert[18] and analyzed using OpenSWATH (2.4)[19] against suitable spectral library[20]. A library named DIA Pan Human Library (DPHL) containing 396,245 peptide precursors and 14,786 protein groups (accepted by GPB at Aug 22, 2019, manuscript No. GPB-D-19-00131R1) was used for HeLa peptides and CCA tissue samples analysis. The second library named BRP DIA library (provided in **Data deposition**) containing 58,015 peptide precursors and 6,331 protein groups was built based on the results of pFind[21] combined with the information of the synthetic iRT peptides (SiRT)[22] for the breast cancer cell line samples analysis.

The retention time extraction window was set at 600 seconds, and *m/z* extraction for MS2 was performed with 30 ppm tolerance while *m/z* extraction for MS1 was performed with 20 ppm tolerance. Retention time was then calibrated using common internal Retention Time standards (CiRT) peptides. The *m/z* extraction for CiRT peptides was performed with 50 ppm tolerance. Peptide precursors from the same sample that were identified by OpenSWATH and pyprophet with FDR<0.01 were combined using a R program named pulseDIA_pep_combine (https://github.com/Allen188/PulseDIA) to quantify the proteins in each sample.

RESULTS AND DISCUSSION

Design of the PulseDIA Acquisition

The 400-1200 *m/z* mass range is divided into 24 windows in a conventional DIA method (**Figure 1**). In PulseDIA, the peptide precursors were isolated in gas phase into more windows with n times smaller range of *m/z*. These windows were then assigned to n times MS injections in a pulse manner. For instance, if n is four, the PulseDIA evenly divides each window from the classical setting to four portions, and then allocate them into four MS injections sequentially (**Figure 1**). As a result, the PulseDIA acquires data for 96 windows with 1/n *m/z* width of the original ones (**Figure 1**). We also evaluated a duplicate PulseDIA method with 50% overlap between adjacent windows to generate duplicated signals with potential to improve the quantitative accuracy and reproducibility.

Customized PulseDIA window scheme can be generated using a R program named PulseDIA_calcu_wins (https://github.com/Allen188/PulseDIA). Parameters including MS1 acquisition range, number of windows, number of injection fractions, whether overlap of windows is allowed, and fixed or variable window scheme can be defined in the script for PulseDIA window scheme generation. As to variable windows, the precursor ions density information is required. With the four-pulse scheme, one would get four PulseDIA raw data for one sample. All the isolation window tables in this experiment were provided in the supplementary **Table S1-4**.

To analyze the PulseDIA data (**Figure S1**), all the PulseDIA raw data were converted to mzXML format using msconvert software for analysis. To analyze the pulseDIA-MS data generated from discontinuous *m/z* windows, we developed scripts to identify the conserved high-abundance peptides with CiRT for retention time calibration. With the four-pulse scheme, four TraML formats of CiRT were required for the analysis using OpenSWATH, and the four matrices obtained from OpenSWATH were then combined for the final peptide and protein identification.

Optimization of PulseDIA

A mixed peptide digest sample from nine breast cancer cell lines was firstly used to evaluate the technical reproducibility of the method, and then to optimize PulseDIA parameters systematically. Results showed that proteins were identified at a high degree of reproducibility (**Figure S2**). The median coefficients of variation (CV) values between the two technical replicates are around 0.5%-1.0%. Four parameters were tested and optimized to maximize the performance of PulseDIA: i) number of injections; ii) length of LC gradient; iii) fixed or variable window; iv) the width of overlaps between adjacent windows (**Figure S3**). A total of 39,139 peptides and 5,033 protein groups were identified by the 5-injection PulseDIA scheme with 60 min LC gradient, fixed windows without half window width using the BRP DIA spectral library containing 58,015 peptide precursors and 6,331 protein groups (**Table S5**). The number of increased proteins identified per increased LC gradient time unit (min) further showed that two PulseDIA runs of 30 min LC gradient using fixed windows without half window overlaps resulted in the maximum increased protein identifications per time unit compared to the conventional 30 min DIA analysis (**Figure S4**). We further compared the PulseDIA with two runs of 30 min LC gradient with the conventional DIA run of 60 min LC gradient (**Figure S5**), and found that the peptide and protein identifications in-

creased by 13.2% and 7.3% in average, respectively.

Next, we used the standard HeLa Protein Digest to verify the optimal parameters of PulseDIA which were obtained from the breast cancer cell line samples above. In order to test the PulseDIA capacity for maximum peptide and protein identification, a 10-injection PulseDIA runs of 30 min LC gradient, fixed windows without half window overlap were applied. As shown in **Figure 2a**, 61,078 peptides and 7,308 protein groups were identified under this 10-injection PulseDIA scheme (**Table S6**) against the DPHL spectral library. Compared to the 30 min LC gradient of conventional DIA, the peptide and protein identifications increased 112.5 % and 37.4% in average, respectively. Besides, with the increase of injection number, the number of peptide and protein identifications both increased too (**Figure 2a**). Similarly, two PulseDIA runs of 30 min LC gradient lead to the maximum increase identification number of peptide and protein per time unit compare to the conventional DIA with 30 min LC and 24 standard DIA windows (**Figure 2b**). The Pearson correlation coefficient of determination (r) between two technical replicates for proteins under these conditions were all greater than 0.9, indicating high repeatability and stability (**Figure 2c**).

Application of PulseDIA to proteotyping of CCA

After benchmarking the PulseDIA method with cell line samples, we next evaluated its applicability to clinical tissue samples. Both PulseDIA-MS and the conventional DIA-MS were applied to the proteomics profiling of 18 tissue samples (benign and tumor pairs) from nine CCA patients. The relevant clinical information was listed in supplementary **Table S7a**. These CCA peptide samples were analyzed with two MS strategies, i.e., the PulseDIA scheme containing two MS runs of 30 min LC gradient, and the conventional DIA-MS of 60 min LC gradient.

Results showed that PulseDIA identified more peptides and proteins than conventional DIA in each sample using the same LC gradient (**Figure S6**). The increased percentage of peptide and protein identification reach up to 34% and 42% respectively (**Table S7c**). In detail, for 18 tissue samples, PulseDIA identified 67,210 peptides and 6,063 protein groups from 36 PulseDIA data, improved by 10% and 23% compared to DIA from 18 DIA data, respectively.

We further checked the peptide and protein identification in different tissues (bebign vs. tumor). In peptides from tumorous tissues, PulseDIA identified a total of 59,505 peptides and 5,821 protein groups, while conventional DIA identified 52,968 peptides and 4,718 protein groups, both against the DPHL library **(Figure 3a**). As to tissue from the adjacent benign area, a total of 48,141 peptides and 4,782 protein groups were identified by PulseDIA, while a total of 44,848 peptides and 4,088 protein groups were identified using conventional DIA.

Moreover, the missing value for protein matrix generated by PulseDIA is lower than the one generated by conventional DIA, with average missing rate for 18 samples dropped from 60% to 45% (**Figure 3b**). The r values between two technical replicates for proteins were 0.99 for PulseDIA and 0.93 for DIA, respectively, indicating that PulseDIA achieved better reproducibility than DIA probably due to increased specificity (**Figure 3c**). The high correlation coefficient of determination (r =0.93) between the quantification of the same proteins identified by these two methods demonstrated the high accuracy of PulseDIA method (**Figure 3d**). Moreover, 618 significantly dysregulated proteins (p-value <0.05, log2 (FC, fold change) >0.5) were identified by PulseDIA while 470 proteins were identified by DIA, respectively, between tumorous and benign tissues (**Figure 3e**). These comparison shows the superiority of PulseDIA-MS compared to DIA-MS in clinical applications.

We then analyzed enriched pathways based on the 618 significantly regulated proteins by Ingenuity Pathway Analysis (IPA) (**Table S8a**).[23] Our data showed that the PPP2R1B protein was enriched in many cancer-associated pathways (**Table S8b**). This protein had been reported in a variety of cancers, such as lung cancer, colon cancer[24], cervical cancer[25] and hepatocellular carcinoma,[26] but there have been no report related to CCA so far. Our data suggest a potential novel role of PPP2R1B in CCA.

IPA also enriched 25 core networks (**Table S8c**). Hepatocyte nuclear factor 4-alpha (HNF4A) was found to be decreased in the 3rd network (show the name of this network please) (**Figure S7a**). Our data also uncovered downregulation of transcription factor HNF4A as well as its downstream proteins including RBM15, PRODH2, SAT2 and CCDC90B in the CCA tumorous samples (**Table S8d**), in agreement with previous report showing inhibition or loss of HNF4A promotes tumorigenesis.[27] The 2nd network involving PI3K pathway (**Figure S7b**), a regulator of CAA[28] was also highlighted in our data. This network contains two upregulated genes including endophilin-B1 (SH3GLB1) and lymphocyte-specific protein 1 (LSP1). SH3GLB1 activates lipid kinase activity of PIK3C3 during autophagy probably by associating with the PI3K complex II (PI3KC3-C2).[29] LSP1 has been reported as an independent factor for predicting the prognosis of hepatocellular carcinoma patients.[30] Together, with the PulseDIA method, we identified promising protein biomarker candidates for CCA.

CONCLUSION

In this study, we developed and benchmarked an alternative DIA-MS acquisition method called PulseDIA. Compared to the conventional DIA method, the PulseDIA method demonstrated higher sensitivity and specificity in peptide and protein identification. A total of 61,078 peptides and 7,308 protein groups were identified using ten PulseDIA runs of HeLa digest with 30 min LC gradient, which achieved an increase of 112.5% and 37.4% at peptide and protein group level, respectively, compared to the conventional 30-min gradient DIA. The reproducibility of PulseDIA is high (r>0.9). We further applied PulseDIA to analyze biopsy-level CCA tissue samples and identified 67,210 peptides and 6,063 protein groups. Compared to the conventional DIA method, the PulseDIA identified and 23% more proteins and reduced the missing value rate by 25%. We identified 618 significantly regulated proteins in CCA samples, uncovering novel protein biomarker candidates for CCA. Our case study showed that the PulseDIA method can be practically applied to protein biomarker research using clinical specimens with higher specificity, sensitivity and reproducibility.

ASSOCIATED CONTENT

Supporting Information

Workflow for PulseDIA data analysis, optimization result for breast cancer lines, number of peptide and protein identification for each CCA tissue sample, and two typical networks obtained by IPA analysis (pdf). All isolation windows for MS method in this study (xlsx). Protein quantitative results of breast cancer lines (xlsx), HeLa digest (xlsx). Sample information, protein quantitative result, IPA analysis result of all CCA tissue samples(xlsx).

Data deposition

All the raw data and TraML files in this report are deposited in iProX. Project accession of the breast cancer lines data: IPX0001769001. Project accession of the HeLa data: IPX0001769002. Project accession of the CCA data: IPX0001769003. All the data will be publicly released upon publication.

ACKNOWLEDGEMENTS

This work was supported by the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars (Grant No. LR19C050001 to T.G.), Hangzhou Agriculture and Society Advancement Program (Grant No. 20190101A04 to T.G.), National Natural Science Foundation of China (General Program) (Grant No. 81972492 to T.G.), National Science Fund for Young Scholars (Grant No. 21904107). We thank Chenhuan Yu for providing breast cancer cell lines.

AUTHOR CONTRIBUTIONS

T.G., X.C., Y.Z. designed the project. J.Z., C.L., P.S. procured the CAA cohorts. X.Y., R.S. performed the PCT-based sample preparation. X.C performed the PulseDIA analysis. W.G., X.C. T.Z analyzed the data. W.G., X.C., G.R. drew the graphs. X.C., T.G., Y.Z., C.Y., S.L., S.H., M.L. wrote the manuscript with inputs from all co-authors. T.G. supported and supervised the project.

CONFLICT OF INTEREST

The research group of T.G. is supported by Pressure Biosciences Inc, which provides access to advanced sample preparation instrumentation.

REFERENCES

(1) Aebersold, R.; Mann, M. *Nature* **2016**, *537*, 347-355.

(2) Gillet, L. C.; Leitner, A.; Aebersold, R. *Annual review of analytical chemistry* **2016**, *9*, 449-472.

(3) Guo, T.; Kouvonen, P.; Koh, C. C.; Gillet, L. C.; Wolski, W. E.; Rost, H. L.; Rosenberger, G.; Collins, B. C.; Blum, L. C.; Gillessen, S.; Joerger, M.; Jochum, W.; Aebersold, R. *Nat Med* **2015**, *21*, 407-413.

(4) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. J. M.; Proteomics, C. **2012**, *11*, O111. 016717.

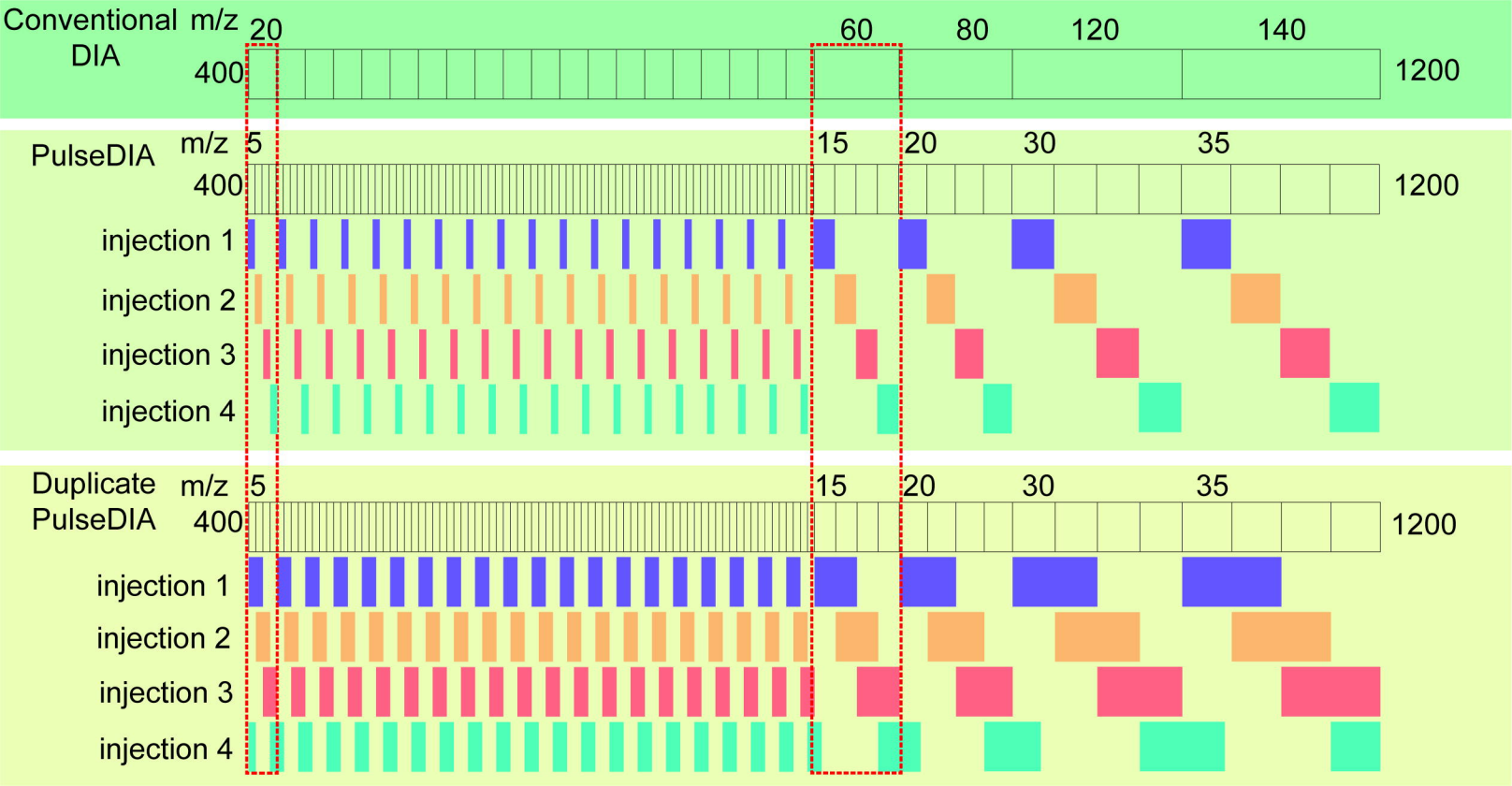
(5) Liu, Y.; Buil, A.; Collins, B. C.; Gillet, L. C.; Blum, L. C.; Cheng, L. Y.; Vitek, O.; Mouritsen, J.; Lachance, G.; Spector, T. D. *Molecular Systems Biology* **2015**, *11*, 786.

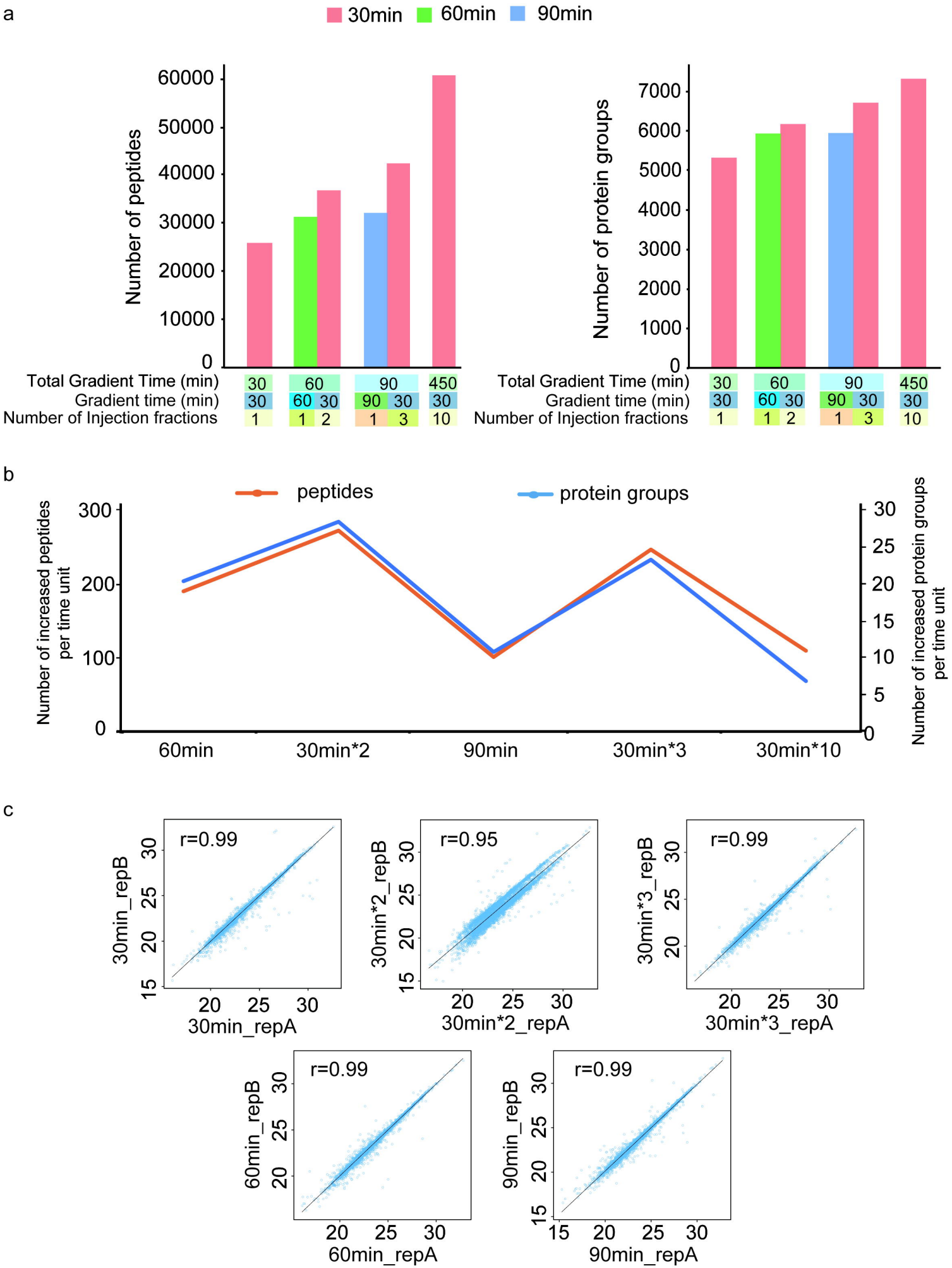
(6) Geiger, T.; Cox, J.; Mann, M. *Molecular cellular proteomics : MCP* **2010**, *9*, 2252-2261.

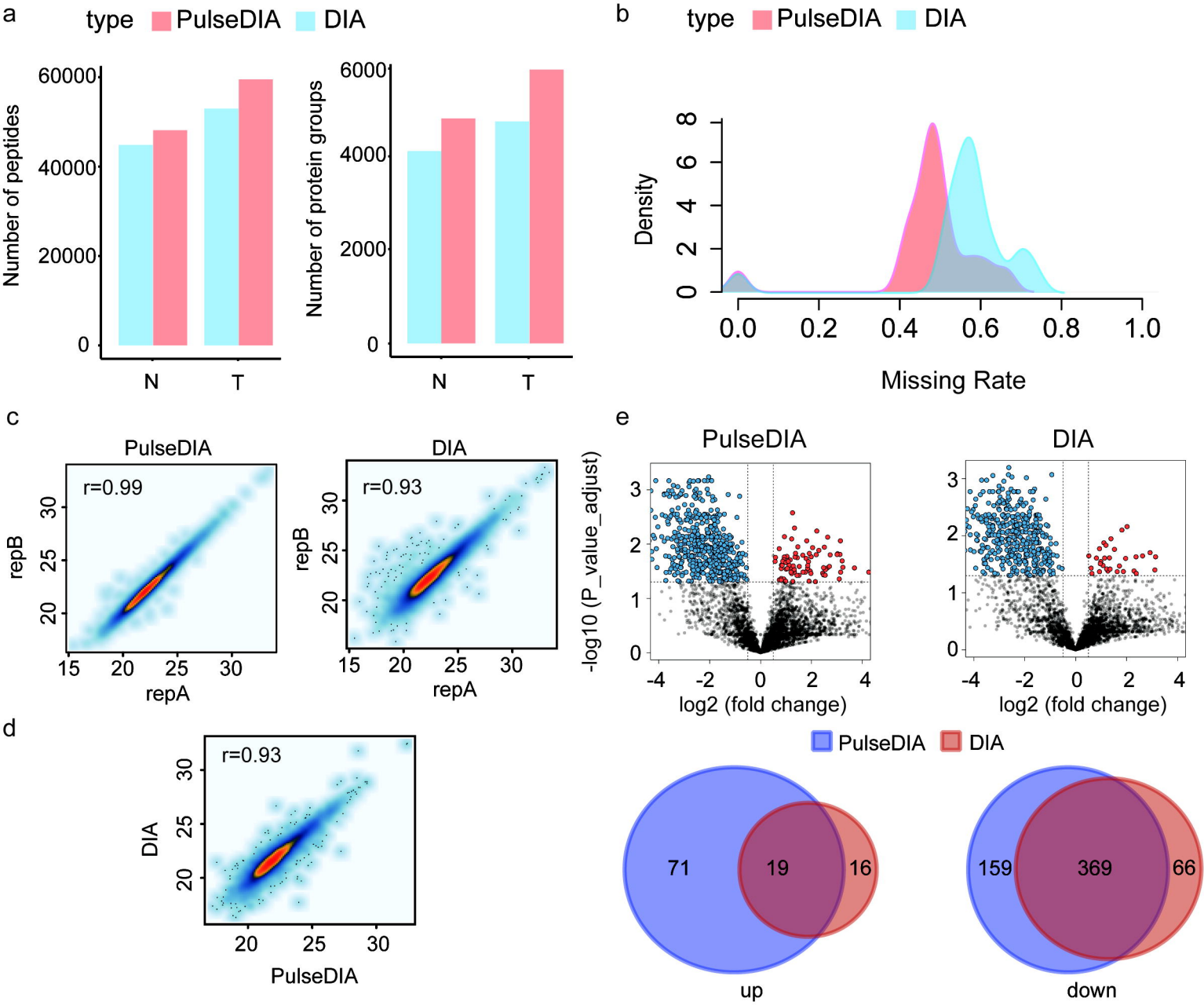
(7) Plumb, R. S.; Johnson, K. A.; Paul, R.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. *Rapid Commun Mass Spectrom* **2006**, *20*, 1989–1994.

(8) Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. *Anal Chem* **2009**, *81*, 6481-6488.

(9) Panchaud, A.; Jung, S.; Shaffer, S. A.; Aitchison, J. D.; Goodlett, D. R. *Anal Chem* **2011**, *83*, 2250-2257.
(10) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; Wu, C. C.; MacCoss, M. J. *Nat Methods* **2013**, *10*, 744-746.
(11) Shimoda, M.; Kubota, K. *World J Gastroenterol* **2007**, *13*, 1500-1504.
(12) Silsirivanit, A.; Sawanyawisuth, K.; Riggins, G. J.; Wongkham, C. *J Hepatobiliary Pancreat Sci* **2014**, *21*, 388-396.
(13) Scarlett, C. J.; Saxby, A. J.; Nielsen, A.; Bell, C.; Samra, J. S.; Hugh, T.; Baxter, R. C.; Smith, R. C. *Hepatology* **2006**, *44*, 658-666.
(14) Lankisch, T. O.; Metzger, J.; Negm, A. A.; Voßkuhl, K.; Schiffer, E.; Siwy, J.; Weismüller, T. J.; Schneider, A. S.; Thedieck, K.; Baumeister, R.; Zürbig, P.; Weissinger, E. M.; Manns, M. P.; Mischak, H.; Wedemeyer, J. *Hepatology* **2011**, *53*, 875-884.
(15) Mustafa, M. Z.; Nguyen, V. H.; Le Naour, F.; De Martin, E.; Beleoken, E.; Guettier, C.; Johanet, C.; Samuel, D.; Duclos-Vallee, J.-C.; Ballot, E. *J Transl Med* **2016**, *14*, 17-17.
(16) Padden, J.; Ahrens, M.; Kälsch, J.; Bertram, S.; Megger, D.; Bracht, T.; Eisenacher, M.; Kocabayoglu, P.; Meyer, H.; Sipos, B.; Baba, H.; Sitek, B. *Molecular cellular proteomics : MCP* **2015**, *15*.
(17) Zhu, Y.; Guo, T. In *Tissue Proteomics: Methods and Protocols*, Sarwal, M. M.; Sigdel, T. K., Eds.; Springer New York: New York, NY, 2018, pp 279-287.
(18) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534-2536.
(19) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. *Nature Biotechnology* **2014**, *32*, 219.
(20) Schubert, O. T.; Gillet, L. C.; Collins, B. C.; Navarro, P.; Rosenberger, G.; Wolski, W. E.; Lam, H.; Amodei, D.; Mallick, P.; MacLean, B.; Aebersold, R. *Nature Protocols* **2015**, *10*, 426.
(21) Li, D.; Fu, Y., Rx; Ling, C.; Wei, Y.; Zhou, H.; Zeng, R.; Yang, Q.; He, S.; Gao, W. *Bioinformatics* **2005**, *21*, 3049-3050.
(22) Claudia, E.; Lukas, R.; Brendan, M. L.; Reto, O.; Franz, H.; John, C.; Maccoss, M. J.; Oliver, R. *Proteomics* **2012**, *12*, 1111-1121.
(23) Andreas, K. M.; Jeff, G.; Jack, P.; Stuart, T. *Bioinformatics* **2014**, *30*, 523.
(24) Wang, S. S.; Esplin, E. D.; Li, J. L.; Huang, L., .; Gazdar, A., .; Minna, J., .; Evans, G. A. *Science* **1998**, *282*, 284-287.
(25) Pulido, H. A.; Fakruddin, M. J.; Chatterjee, A.; Esplin, E. D.; Beleño, N.; Martínez, G.; Posso, H.; Evans, G. A.; Murty, V. V. V. S. *Cancer Research* **2000**, *60*, 6677.
(26) Chou, H. C.; Chen, C. H.; Lee, H. S.; Lee, C. Z.; Huang, G. T.; Yang, P. M.; Lee, P. H.; Sheu, J. C. *Cancer Letters* **2007**, *253*, 138-143.
(27) Bei-Fang, N.; Jin, D.; Jiao, L.; Chuan, Y.; Wen-Ping, X.; Wen-Ming, C.; Qing, Z.; Fei, C.; Tao, H.; Xing, D. *Hepatology* **2014**, *60*, 1607–1619.
(28) Chung, J.-Y.; Hong, S.-M.; Choi, B. Y.; Cho, H.; Yu, E.; Hewitt, S. M. *Clinical Cancer Research* **2009**, *15*, 660.
(29) Takahashi, Y.; Coppola, D.; Matsushita, N.; Cualing, H. D.; Sun, M.; Sato, Y.; Liang, C.; Jung, J. U.; Cheng, J. Q.; Mulé, J. J.; Pledger, W. J.; Wang, H.-G. *Nat Cell Biol* **2007**, *9*, 1142-1151.
(30) Zhang, H.; Wang, Y.; Liu, Z.; Yao, B.; Dou, C.; Meng, X.; Li, Q.; Jia, Y.; Wu, S.; Tu, K. *Febs Open Bio* **2016**, *6*.


Conventional DIA
m/z
20
60
80
120
140
400
1200
PulseDIA
m/z
5
15
20
30
35
400
1200
injection 1
injection 2
injection 3
injection 4
Duplicate PulseDIA
m/z
5
15
20
30
35
400
1200
injection 1
injection 2
injection 3
injection 4


a
30min
60min
90min
Number of peptides
60000
50000
40000
30000
20000
10000
0
Number of protein groups
7000
6000
5000
4000
3000
2000
1000
0
Total Gradient Time (min)
Gradient time (min)
Number of Injection fractions
30
60
90
450
30
60 30
90 30
30
1
1 2
1 3
10
b
peptides
protein groups
Number of increased peptides per time unit
Number of increased protein groups per time unit
300
200
100
0
30
25
20
15
10
5
0
60min
30min*2
90min
30min*3
30min*10
c
r=0.99
30min_repB
30min_repA
r=0.95
30min*2_repB
30min*2_repA
r=0.99
30min*3_repB
30min*3_repA
r=0.99
60min_repB
60min_repA
r=0.99
90min_repB
90min_repA
15
20
25
30


a
type
PulseDIA
DIA
Number of peptides
60000
40000
20000
0
N
T
Number of protein groups
6000
4000
2000
0
N
T
b
type
PulseDIA
DIA
Density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Missing Rate
c
PulseDIA
r=0.99
repB
repA
DIA
r=0.93
repB
repA
d
r=0.93
DIA
PulseDIA
e
PulseDIA
DIA
-log10 (P_value_adjust)
log2 (fold change)
PulseDIA
DIA
71
19
16
up
159
369
66
down