

# SCIENTIFIC REPORTS



OPEN

## SWATH-based proteomics identified carbonic anhydrase 2 as a potential diagnosis biomarker for nasopharyngeal carcinoma

Received: 14 October 2016  
Accepted: 15 December 2016  
Published: 24 January 2017

Yanzhang Luo<sup>1,\*</sup>, Tin Seak Mok<sup>2,\*†</sup>, Xiuxian Lin<sup>2</sup>, Wanling Zhang<sup>1</sup>, Yizhi Cui<sup>1</sup>, Jiahui Guo<sup>1</sup>, Xing Chen<sup>1</sup>, Tao Zhang<sup>2</sup> & Tong Wang<sup>1</sup>

Nasopharyngeal carcinoma (NPC) is a serious threat to public health, and the biomarker discovery is of urgent needs. The data-independent mode (DIA) based sequential window acquisition of all theoretical fragment-ion spectra (SWATH) mass spectrometry (MS) has been proved to be precise in protein quantitation and efficient for cancer biomarker researches. In this study, we performed the first SWATH-MS analysis comparing the NPC and normal tissues. Spike-in stable isotope labeling by amino acids in cell culture (super-SILAC) MS was used as a shotgun reference. We identified and quantified 1414 proteins across all SWATH-MS analyses. We found that SWATH-MS had a unique feature to preferentially detect proteins with smaller molecular weights than either super-SILAC MS or human proteome background. With SWATH-MS, 29 significant differentially express proteins (DEPs) were identified. Among them, carbonic anhydrase 2 (CA2) was selected for further validation per novelty, MS quality and other supporting rationale. With the tissue microarray analysis, we found that CA2 had an AUC of 0.94 in differentiating NPC from normal tissue samples. In conclusion, SWATH-MS has unique features in proteome analysis, and it leads to the identification of CA2 as a potentially new diagnostic biomarker for NPC.

Although nasopharyngeal carcinoma (NPC) is a rare cancer type in most parts of the world, it is a common malignancy in a few well-defined populations, including natives of Southern China, Southeast Asia, the Middle East and Southern Africa<sup>1</sup>. In Guangdong province, China, the approximate incidence rate is 10~30 cases per 100000 people-years<sup>2</sup>, which positions NPC one of the most serious local health treats. Similar to other cancer types, the five-year overall survival (OS) of NPC decreases along with the increased disease stages. For example, Lee *et al.* have found that the OS according to stage is 90% for stage I, 84% for stage II, 75% for stage III, and 58% for stage IVA-B<sup>3</sup>. To be noted, a very promising recent clinical trial has reached a 3-year failure-free survival of 80% in the late stage NPC treatment<sup>4</sup>. Despite this, it has been generally accepted that early diagnosis has critical significance on containing the NPC disease progression and increasing survival, which emphasizes the field of NPC biomarker discovery.

As NPC is highly correlated to the Epstein-Barr virus (EBV) infection, EBV viral RNAs (miRNAs<sup>5,6</sup>, long non-coding RNAs<sup>7</sup>) and proteins are of diagnostic powers. In clinical practice, EBV capsid antigen IgA (VCA-IgA), early antigen Ig A (EA-IgA), nuclear antigen (EBNA1) IgA, and circulating EBV DNA are widely used for diagnosis and screening of NPC<sup>8-10</sup>. It is known that EBV infection is very common in general population; thus, the viral factor in peripheral blood is not a necessary indicator of NPC, rather an efficient parameter to monitor the NPC progression or therapeutic effects<sup>11</sup>.

As such, the host protein factors are emphasized for their biomarker potentials. Individual proteins, such as Bmi-1<sup>12</sup>, HMGB1<sup>13</sup> and KLHDC4<sup>14</sup>, have been associated with diagnosis and/or prognosis of NPC. Furthermore,

<sup>1</sup>Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, Guangzhou 510632, P. R. China.

<sup>2</sup>Department of Otorhinolaryngology, The First Affiliated Hospital, Jinan University, Guangzhou 510632, P. R. China.

<sup>†</sup>Present address: Department of Otorhinolaryngology, Centro Hospitalar Conde de São Januário, Macao SAR 999078, P. R. China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.Z. (email: tzhangt@jnu.edu.cn) or T.W. (email: tongwang@jnu.edu.cn)

high-throughput methods, including proteomics and next generation sequencing, are rapidly moving the field forward<sup>15</sup>. For example, Chen *et al.* employed isobaric tags for relative and absolute quantitation (iTRAQ) -based mass spectrometry (MS) to compare the motile and non-motile NPC cells, in which they found and validated the biomarker potentials of RAN, SQSTM1 and TRIM29<sup>16</sup>. Besides iTRAQ-MS, shotgun MS approaches are primarily based on the data-dependent acquisition (DDA) mode, known with the advantage of deep proteome coverage. Theoretically, shotgun MS is a probability based identification; and controlling the false discovery is an inevitable challenge, especially when dealing with the database search against large reference databases<sup>17</sup>.

As a major advance in the field of proteomics, the data-independent acquisition (DIA) based approach of sequential window acquisition of all theoretical fragment-ion spectra (SWATH) -MS has brought general attentions<sup>18</sup>. The basic theory of SWATH is to acquire all theoretical fragment ion spectra, and assemble them back to the parent ions to achieve high-throughput identification and quantification<sup>19</sup>. This is an extraordinary and complementary expansion of the relatively low-throughput selected/multiple reaction monitoring (SRM/MRM)<sup>20,21</sup>.

SWATH has been justified to be valuable in a few cancer types for biomarker discovery, such as prostate cancer<sup>22</sup>, colorectal cancer<sup>23</sup>, gastric cancer<sup>24</sup>, and lung cancer<sup>25</sup>. However, its application in the NPC field is very limited. In this study, we performed the first SWATH-MS analysis comparing the NPC and normal tissues, in which we successfully identified and verified carbonic anhydrase 2 (CA2) as a potentially new diagnosis biomarker of NPC.

## Results

**SWATH and super-SILAC MS analyses.** To quantify the proteome change of NPC tissues, we took advantage of the DIA feature of SWATH, and the super-SILAC MS was used as a general shotgun MS verification. Tissue lysates from 9 normal and 9 NPC subjects were respectively pooled and analyzed by super-SILAC MS analysis. In addition, 6 normal and 5 NPC subjects from the same batch of donors used for the super-SILAC MS were individually analyzed with SWATH-MS. Six more clinical samples were obtained for the subsequent verification steps. No statistical difference was observed regarding donor gender or age ( $P > 0.05$ , Fisher's exact test, Supplementary Table S1).

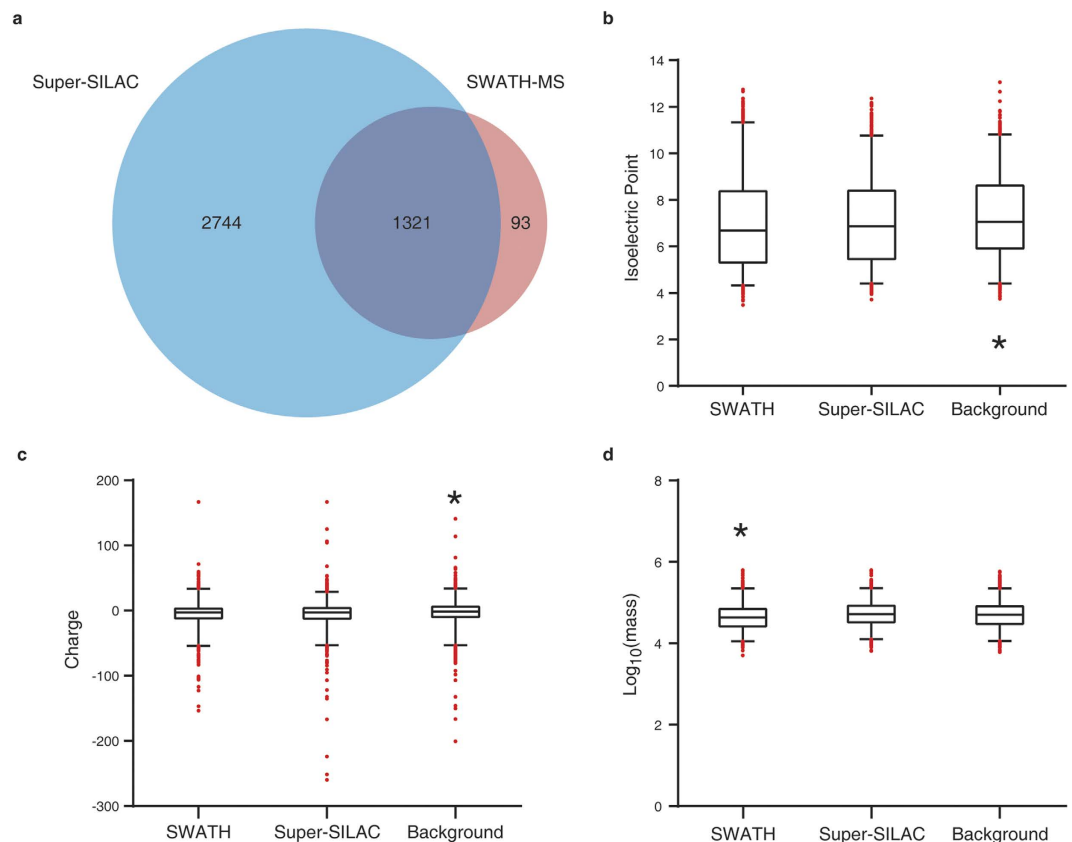
We identified and quantified 1414 proteins across all 11 samples in the SWATH-MS analysis (Supplementary Table S3). While in the super-SILAC MS analysis on the pooled samples, 4065 proteins were quantified from both the normal and the NPC groups with protein FDR < 1% (Supplementary Table S4). Comparing the two methods, 1321 proteins were quantified in overlap (Fig. 1a). All MS raw data are available in iProX (accession number: IPX00080100).

We next focused on the physical-chemical characteristics of the SWATH-MS identified 1414 proteins, in comparison of the same amount (1414) of proteins randomly selected from super-SILAC identifications and the neXtProt PE1 proteins. Both MS methods preferentially detected proteins with lower isoelectric points (pI) (Fig. 1b) and higher charges (Fig. 1c) than the human proteome background. It is known that MS is keen on identifying more acidic proteins. In addition, more charges will lead to easier ionization of peptides, which is favorable for MS analyses. Hence, the above two features can be expected. Nonetheless, we found that SWATH-MS tended to identify smaller proteins with the median molecular weight (MW) of 43.7 kDa, significantly less than the super-SILAC MS identifications (median MW = 52.5 kDa) and the background PE1 proteins (median MW = 51.3 kDa) (Fig. 1d).

**Differentially expressed proteins in SWATH-MS.** We next used the power law global error model (PLGEM) algorithm<sup>26</sup> to determine differentially expressed proteins (DEPs). We and others have found that this method is of advantages to consider the global error while fitting the proteome abundance data for statistical tests<sup>27,28</sup>. We found that our SWATH-MS data could be well fitted by PLGEM, with the slope of 0.886 and adjusted  $r^2$  of 0.996 (Pearson  $r = 0.961$ ) (Fig. 2a). The residuals between the modeled and the measured standard deviation (SD) generally followed normal distribution (Fig. 2b). These results suggested that PLGEM worked properly in dealing with the SWATH-MS data.

It has been shown that PLGEM provides a more powerful signal-to-noise (STN) ratio, especially in low-abundance protein evaluations, by incorporating the PLGEM-derived SD<sup>26,27</sup>. In this study, we observed that the PLGEM-STN was significantly correlated to the relative protein fold changes, with a Spearman  $r$  of 0.99 (Fig. 2c). Based on PLGEM-STN, we computed out 29 proteins as DEPs ( $P < 0.01$ ), which took over ~2% of total proteins (Fig. 2d, Supplementary Table S2). These DEPs have certain reproducibility as quantified by super-SILAC and SWATH-MS (Fig. 2e). Such 29 DEPs could not completely separate the NPC tissues from the normal tissues per cluster analyses, potentially due to the small sample size (Fig. 2f). Despite this, we noted that 4 subgroups could be differentiated, and no NPC and normal tissues were clustered together (Fig. 2f).

**Common upstream regulator detected for SWATH and super-SILAC DEPs.** The upstream analysis module of Ingenuity Pathway Analysis (IPA) suggested that interferon gamma (IFNG) was the significantly unbiased upstream activator of SWATH-MS detected 10 DEPs ( $z$ -score = 2.282,  $P = 1.44 \times 10^{-6}$ ; Fig. 3a). In the super-SILAC detected DEPs, consistent prediction of IFNG was computed out (Fig. 3b); in addition, other known viral infection relevant pathways, such as IFNA2, IFNB1 and TLR, were also deemed up-stream activators by IPA (Fig. 3b). Specific to the mechanistic network of IFNG, 5 common DEPs, such as STAT1 and IFIT1, were shared by both SWATH-MS (Fig. 3c) and super-SILAC MS (Fig. 3d). While other proteins were unique to methods; for example, PRDX2 was only detected in SWATH-MS (Fig. 3c). These results favored the complementary feature of different MS methods. In addition, such a similar upstream regulator prediction suggested that even we used different acquisition mode-based MS approaches (DDA or DIA), the bioinformatics implicated reproducibility, which favorably argued that our general research strategy was valid.

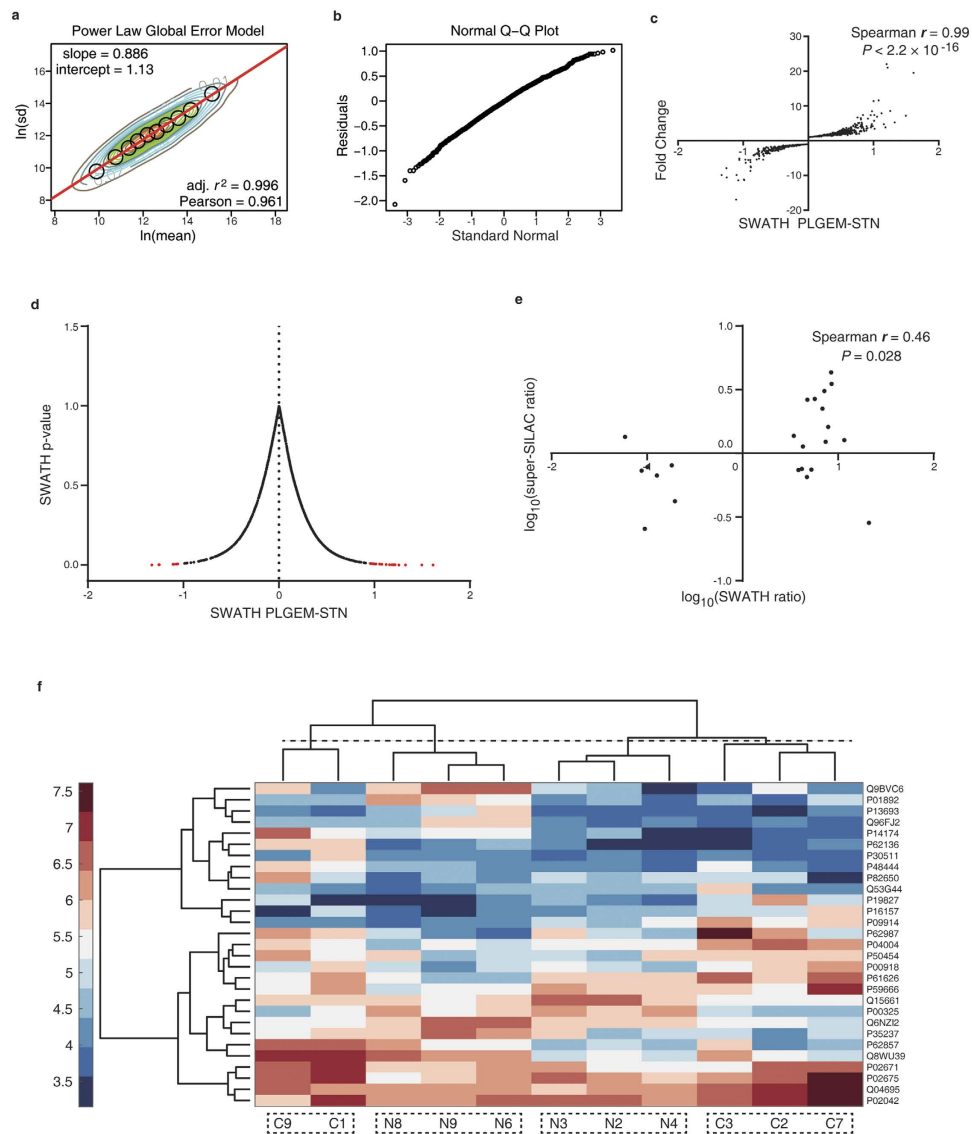


**Figure 1. SWATH and super-SILAC-based MS identifications.** (a) Venn diagram comparison of identification numbers. (b–d) Protein physical-chemical feature distributions, regarding isoelectric points (b) and charges at physiological conditions (c) and molecular weights (d). Data are presented with box-and-whisker plot, and 5% outliers are shown with red dots. A total of 1414 proteins were randomly sampled from the super-SILAC identifications and background human proteome, respectively, for the comparison with the SWATH MS identifications. \* $P < 0.01$ , as compared with any of the other groups, two-tailed Kolmogorov–Smirnov test.

**SWATH spectra of CA2.** Among the 29 DEPs detected by SWATH-MS, thirteen of them were also detected to be up-regulated in the super-SILAC MS analysis (Table 1). When applying the 1.5-fold threshold as referenced by the super-SILAC quantitation, seven proteins were left for further evaluation. We prioritized CA2 for further evaluation based on the following reasons: 1) CA2 had 4 unique peptides identified in the shotgun mode (Fig. 4a, and Supplementary Fig. S2), and the SWATH results could pass the manual spectra inspection; 2) Wu *et al.* previously reported that CA2 was up-regulated in the serum of NPC xenograft mice<sup>29</sup>; and 3) the CA2's diagnostic power had not been reported in the NPC field.

As illustrated in Fig. 4, we showed an exclusively unique peptide spectrum of CA2 identified in the shotgun mode; it had 8 consecutive  $y$ -ions labeled with mass error less than 0.5 Da (Fig. 4a). When analyzed in the SWATH mode, 6  $y$ -ions were found to have the retention time at ~97 min (black arrow, Fig. 4b), which was consistent with their parent-ion retention time as recorded in the shotgun mode. These 6  $y$ -ions could be found in the MS2 spectrum of the SWATH analysis (Fig. 4c). These results proved that the SWATH identification of CA2 was of high confidence.

**CA2 has diagnostic value in NPC.** We next used more clinical samples to perform immunoblotting (IB) verification of CA2 changes. We found that although the NPC group tended to have more CA2 expression than the normal group, no significant difference was observed (Fig. 5a). The raw images of IB could be found in Supplementary Fig. S3. We reasoned that the sample size was not sufficient for the statistics of IB results. Thus, we employed a commercialized tissue microarray (NPC donor  $n = 52$  and normal donor  $n = 13$ ) to evaluate the diagnostic power of CA2. We could observe remarkably more CA2 expression in some NPC tissues, such as the example shown in Fig. 5b. But, we could still visually distinguish the within group variations; the immunohistochemistry (IHC) results of all tissue points could be found in Supplementary Fig. S1 and Table S5. Despite the variation, the KS-test results showed that the CA2 Histologic Scores of the NPC tissues were significantly higher than the normal tissues (Fig. 5c). We noted that the sample size was considerably different between the two groups. As such, we used bootstrap analysis to further verify the conclusion. We found that post 10000 times bootstrap resampling from each group, the 95% confidence intervals of the two groups were not overlapped, suggesting that the mean values of the two groups were statistically different (Fig. 5d). With receiver operating characteristic



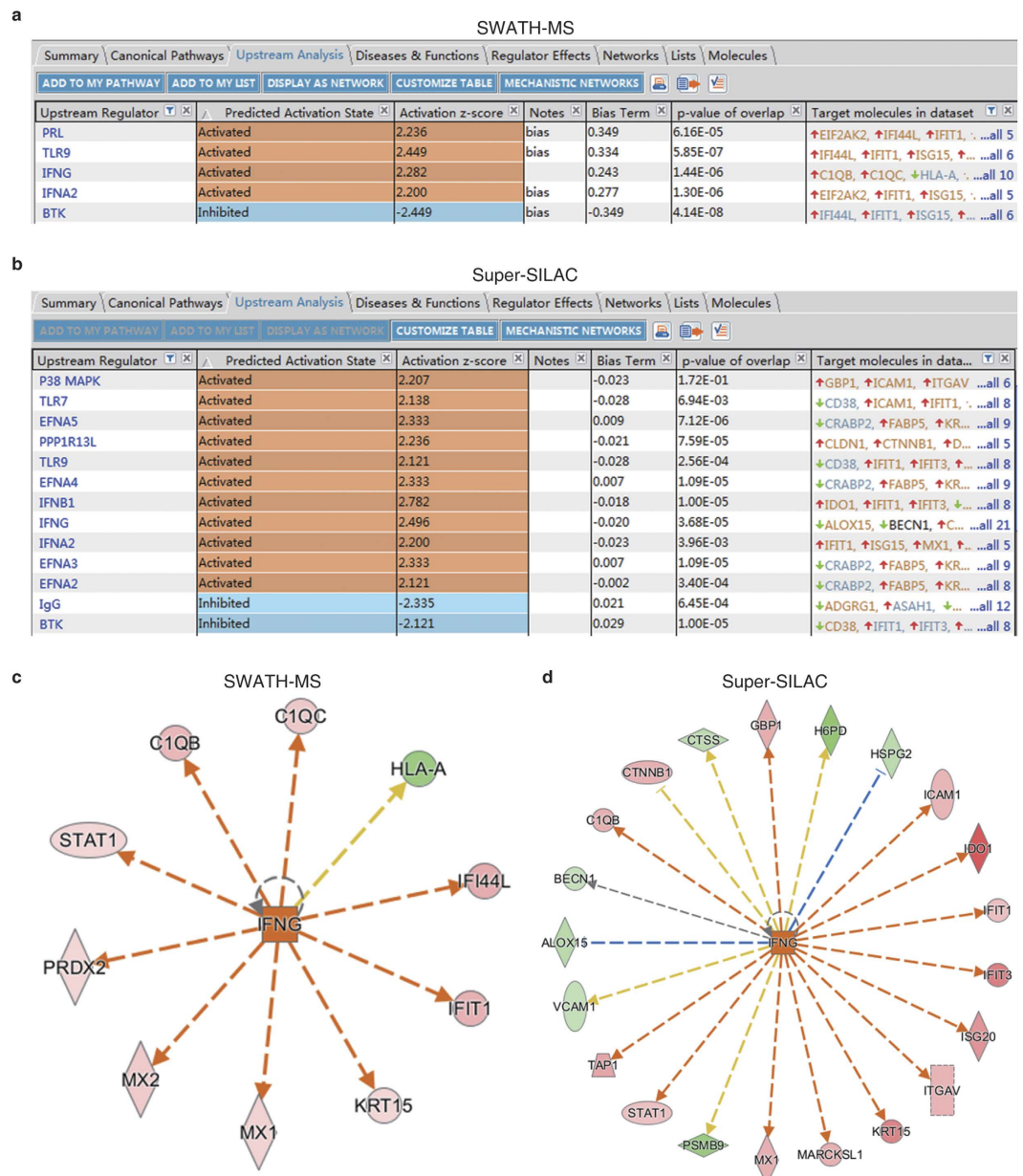
**Figure 2. Determination of differentially expressed proteins in SWATH-MS.** (a) Contour plots of  $\ln(\text{rowCV})$  versus  $\ln(\text{rowMean})$  scatter plots and the regression fitting analysis with PLGEM model. Black circles indicate the modeling points used to fit a PLGEM model. (b) Quantile-quantile (Q-Q) plot. (c) Correlation of signal-to-noise ratios (STN) and protein abundance fold changes. (d) A volcano plot showing the relationship between the PLGEM-STN and the p-value calculated by PLGEM. The DEPs are shown in red ( $P < 0.01$ ). (e) DEP fold change correlation of SWATH-MS and super-SILAC MS analyses. (f) Cluster analysis using DEPs. C represents subjects with NPC, and N stands for normal subjects.

(ROC) analysis on the Histologic Score, we found that CA2 showed high diagnostic power with the Area Under Curve (AUC) of 0.94 (Fig. 5e).

## Discussion

In this study, we demonstrated that SWATH-MS in the DIA mode worked properly in the NPC biomarker discovery with a unique feature to preferentially detect proteins with low MWs. With such an aide, we have identified and justified that CA2 is a potentially new diagnostic biomarker that has high statistical power to differentiate NPC from normal tissues.

In our past work, we have found that human cells are preferentially translating shorter mRNAs, and when considering the mRNA length, translating mRNAs and proteins are highly correlated in their abundances<sup>30</sup>. We have further proposed a computational model to justify that such length-dependence translation preference is a survival strategy of human cells to maintain a functional proteome by avoiding erroneous protein products<sup>31</sup>. These findings suggest that the stoichiometry of translating mRNA into protein is correlated to the mRNA length<sup>30,31</sup>. A recent study from Huang *et al.* have evidenced that SWATH-MS signal intensities have precisely linear correlation to the sample loading abundances in a label-free analysis<sup>32</sup>. It is known that SWATH scanning is generally referenced by the shotgun library, which is usually based on single-injection MS analysis. In such a scenario,



**Figure 3. Upstream regulator analysis with IPA. (a,b)** Upstream analysis of DEPs from SWATH-MS (a) and super-SILAC MS (b). Activated ( $z\text{-score} \geq 2$ ) and inhibited ( $z\text{-score} \leq -2$ ) upstream regulators are highlighted in orange and blue, respectively. (c,d) IFNG regulated proteins in SWATH-MS (c) and super-SILAC MS (d) analyses. Up-regulated and down-regulated proteins are highlighted in red and green, respectively, and the color depth is correlated to the fold change. Orange and blue dashed lines with arrows indicate indirect activation and inhibition, respectively. Yellow and gray dashed lines with arrows depict inconsistent effects and no prediction, respectively.

proteins with higher molar concentrations will have a higher probability to be identified by the shotgun MS. In sum, SWATH-MS tends to focus on the fragment ions derived from protein products with less amino acid length, while it has outstanding quantitation precision. Such a unique feature of SWATH-MS should partially explain its complementary feature to other MS approaches, such as super-SILAC results demonstrated in this study. It should not be a surprise to see its capacity in identifying new biomarkers.

In human, the carbonic anhydrase family has 16 enzymes that catalyze the reversible reaction from carbon dioxide and water to bicarbonate and protons. They have diversified associations with cancer, autoimmune disease and viral infection<sup>33,34</sup>. Viikila *et al.* have found that CA2 and CA12 have prognostic power in colorectal carcinomas<sup>35</sup>; and Kurono *et al.* have reported that CA2 expression in breast cancer is significantly higher than normal tissues<sup>36</sup>. It is known that solid tumors are featured by the tumor microenvironment, with aberrant activation of numerous immune cells<sup>28</sup>. Interestingly, CA2 and other carbonic anhydrases can cause autoimmune

Swiss-Prot ID	Protein name	NPC/normal ratio in super-SILAC	NPC/normal average ratio in SWATH	SWATH P-value	HGNC Gene name
P16157	Ankyrin-1	4.32	8.48	$8.64 \times 10^{-3}$	<i>ANK1</i>
P04004	Vitronectin	3.51	8.56	$9.53 \times 10^{-4}$	<i>VTN</i>
Q04695	Keratin, type I cytoskeletal 17	3.09	7.21	$1.10 \times 10^{-4}$	<i>KRT17</i>
P00918	Carbonic anhydrase 2	2.67	5.71	$7.05 \times 10^{-3}$	<i>CA2</i>
P50454	Serpin H1	2.64	4.77	$8.84 \times 10^{-3}$	<i>SERPINH1</i>
P09914	Interferon-induced protein with tetratricopeptide repeats 1	2.24	6.88	$5.39 \times 10^{-3}$	<i>IFIT1</i>
P14174	Macrophage migration inhibitory factor	1.61	7.88	$3.24 \times 10^{-3}$	<i>MIF</i>
Q8WU39	Marginal zone B- and B1-cell-specific protein	1.37	3.44	$6.74 \times 10^{-3}$	<i>MZB1</i>
P30511	HLA class I histocompatibility antigen, alpha chain F	1.3	11.4	$7.40 \times 10^{-3}$	<i>HLA-F</i>
P48444	Coatamer subunit delta	1.27	11.61	$4.53 \times 10^{-3}$	<i>ARCNI</i>
P82650	28 S ribosomal protein S22, mitochondrial	1.23	7.4	$6.93 \times 10^{-3}$	<i>MRPS22</i>
P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	1.19	22.04	$1.61 \times 10^{-3}$	<i>PPP1CA</i>
P62857	40 S ribosomal protein S28	1.13	4.29	$9.40 \times 10^{-3}$	<i>RPS28</i>

**Table 1.** Differentially expressed proteins of SWATH-MS referenced by super-SILAC MS.

reaction via activating mast cells<sup>37</sup> and plasma cells<sup>38,39</sup>. Thus, the diagnostic power of CA2 on NPC has comparable evidence from other solid tumor types, and it potentially reflects the inflaming tumor microenvironment.

Furthermore, we demonstrated that the DEPs detected by SWATH-MS had significantly common regulator of IFNG, which implicated the anti-viral response in NPC subjects. IFNG is an essential cytokine produced by both innate and adaptive immune cells, which is highly active to fight against viral, certain bacteria and protozoal infections. The IFNG level is to rise post primary EBV infection in humans<sup>40</sup>; while along with the increased EBV viral load, the serum IFNG level increases in NPC patients<sup>41</sup>. Such rationales help to argue the biological relevance of the SWATH-MS DEPs to NPC or EBV infection. Favorably, among these DEPs, other groups have proven VTN<sup>42,43</sup> and MIF<sup>44,45</sup> as potential NPC biomarkers. Therefore, as the viral response contributes to the NPC-associated inflammation, these findings suggest that proteins participated such a bioprocess constitute a resource of potential NPC diagnosis biomarkers.

## Methods

**Human nasopharyngeal tissue samples.** The nasopharyngeal tissue samples were acquired from The First Affiliated Hospital of Jinan University via biopsy. The scientific and ethics review committees of Jinan University approved this study, and written informed consents were obtained from all of the study participants. All methods were performed in accordance with the relevant guidelines and regulations.

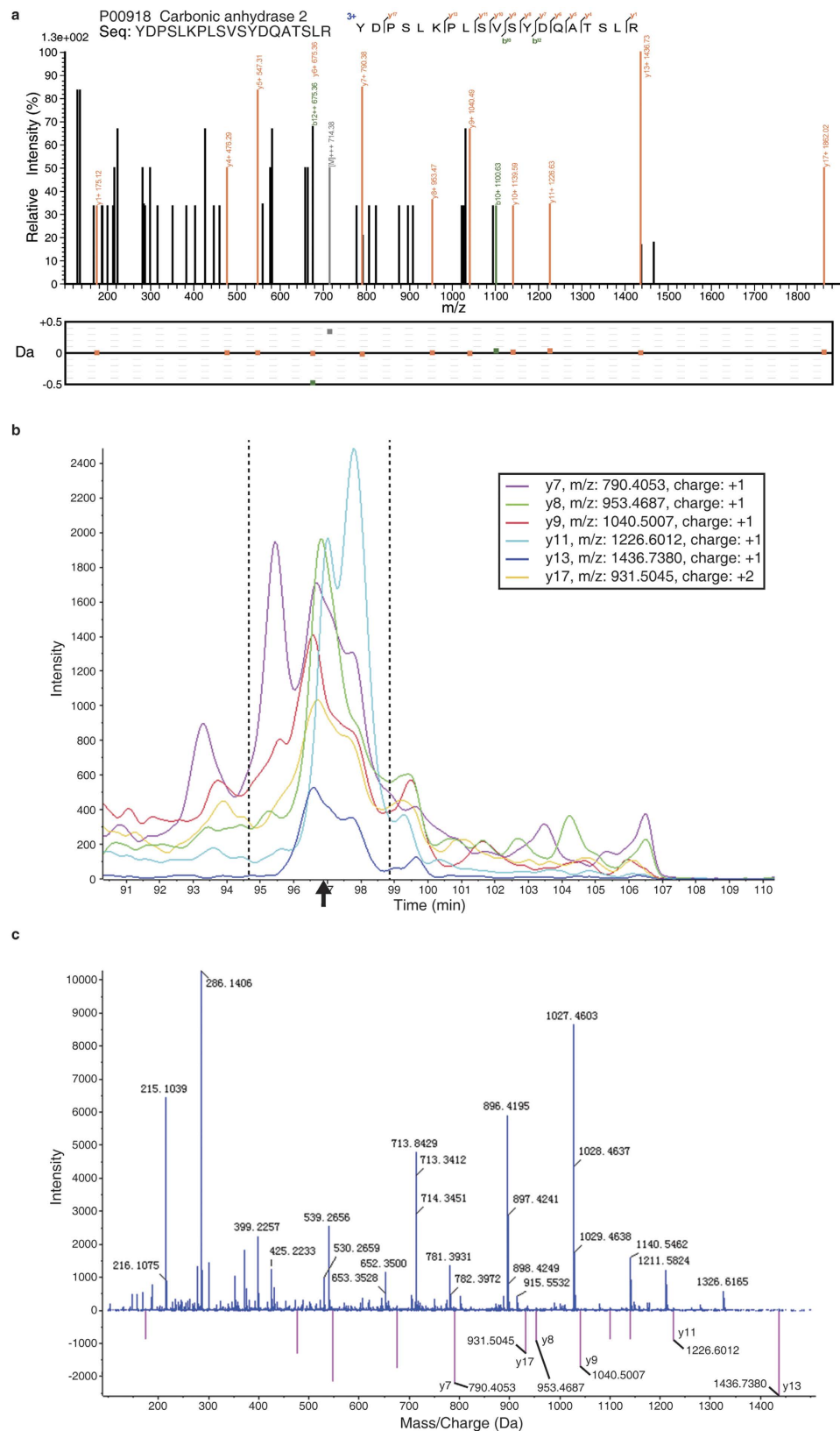
**Cell culture.** Human cell line CNE-1 and CNE-2 were acquired from American Type Culture Collections (ATCC, Rockville, MD). All cells were maintained in the complete Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, 10 µg/mL ciprofloxacin.

**SILAC labeling.** CNE-1 and CNE-2 were subjected to SILAC labeling as we previously described<sup>28,46</sup>. In brief, cells were cultured in the heavy SILAC medium, DMEM containing 73 mg/L <sup>13</sup>C<sub>6</sub>-<sup>15</sup>N<sub>2</sub>-L-lysine (Lys8) and 42 mg/L <sup>13</sup>C<sub>6</sub>-<sup>15</sup>N<sub>4</sub>-L-arginine (Arg10) (Cambridge Isotope Laboratories, Andover, MA, USA), supplemented with 10% dialytic FBS (Life Technologies), 1% pen/strep and various forms of essential amino acids (Cambridge). After at least 8 passages, cells were lysed, and a pooled cell lysate was used for the spike-in standard for the subsequent super-SILAC-based shotgun MS analysis as developed by Mann's group<sup>47</sup>.

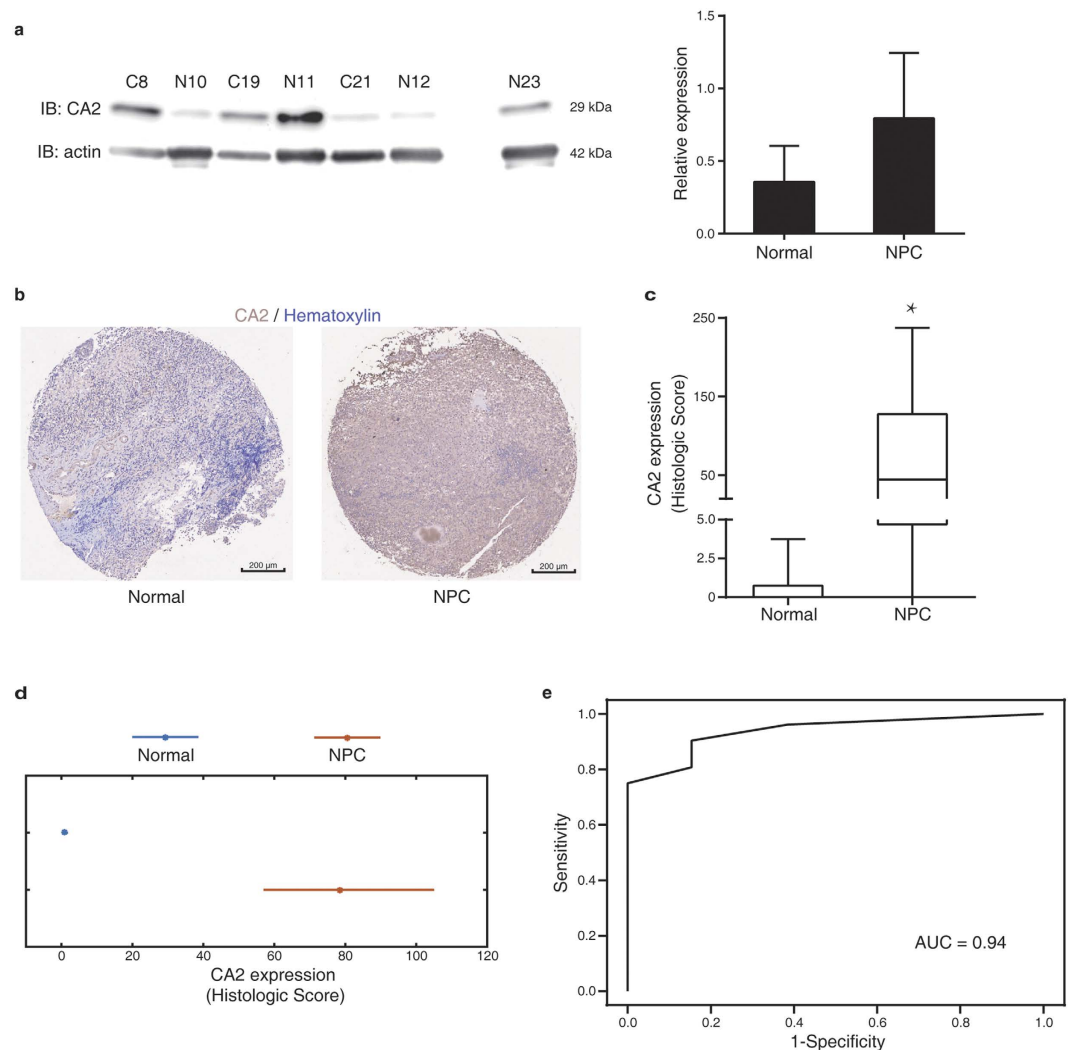
**Protein extraction and digestion.** Post-biopsy, tissue samples were immediately treated with 1% SDS lysis buffer (Beyotime, Nanjing, China), supplemented with 1 mM phenylmethanesulfonyl fluoride (PMSF), and 2% (v:v) protease inhibitor (Roche, Shanghai, China), followed by grounding extraction in liquid nitrogen. The tissue lysate was sonicated and centrifuged at 17,000 × g for 30 min. Supernatants were collected, and the protein concentration was determined by a BCA kit (ThermoFisher Scientific, Shanghai, China). Regarding cell sample lysis, we followed our reported procedure<sup>48</sup>.

We employed in-solution protein digestion with a filter-aided sample preparation (FASP) method<sup>49</sup>, as we described previously<sup>48</sup>. Briefly, samples were subjected to reduction (8 M urea and 50 mM DTT at 37 °C, 1 h) and alkylation (100 mM IAA, at room temperature, 30 min) in the 30 kDa ultracentrifugal filters (Sartorius Stedim Biotech, Shanghai, China). After 15 min centrifugation (12,000 × g, 4 °C), two sequential buffer changes were performed using 8 M urea and 50 mM NH<sub>4</sub>HCO<sub>3</sub>, respectively. Trypsin was then added into the filter at a mass ratio of 1:30 for 8 h, at 37 °C. Peptides were collected by centrifugation at 12,000 × g, 4 °C, 15 min.

**Super-SILAC-based shotgun mass spectrometry.** Tissue protein extracts were mixed with the SILAC labeled spike-in standard at a 1:1 mass ratio prior to the in-solution digestion. Peptides were fractionated with an approach using SAX StageTip and C18 StageTips, and 6 peptide fractions were collected by using serial



**Figure 4. SWATH-MS spectra of CA2.** (a) A shotgun spectrum of a CA2 unique peptide. The mass tolerance is shown for each detected fragment ion. (b) Retention time for the SWATH detected fragment ions. The black arrow indicates the retention time referenced by the parent ion detected in the shotgun mode. (c) The  $m/z$  distribution of SWATH detected fragment ions. The theoretical peaks of the CA2 ions were indicated by the vertical red lines, while the experimentally detected CA2 y-ions were specified.



**Figure 5. Diagnostic power verification of CA2.** (a) Immunoblotting analysis on the CA2 expression in NPC and normal tissues. Data are shown as mean  $\pm$  s.e.m. (b) Representative immunohistochemical images of CA2 staining in a tissue array of nasopharyngeal tissue from normal and NPC donors. (c) Statistical analysis of CA2 expressions. The Histologic Scores of each group are shown in a box-and-whisker plot.  $*P = 5.28 \times 10^{-6}$ , donor of normal  $n = 13$ , donor of NPC  $n = 52$ , two-tailed Kolmogorov–Smirnov test. (d) Bootstrap comparison of the CA2 expression. The star indicates the mean Histologic Score of each group, while the horizontal line depicts 95% confidence interval of the mean value acquired from 10000 resampled Histologic Scores. (e) The ROC curve generated from the Histologic Scores of all donors.

elution buffers with pH values of 11, 8, 6, 5, 4 and 3<sup>46,47</sup>. Peptides were then analyzed with a TripleTOF<sup>®</sup> 5600 MS (5600 MS; AB SCIEX, Framingham, CA, USA) as we previously described<sup>28,46</sup>. MS parameters: spray voltage, 2.3 kV; interface heater temperature, 120 °C; scan range, 350–1500  $m/z$ ; mass tolerance, 50 mDa; resolution, >30 k fwhm; information-dependent acquisition (IDA) MS/MS scans, applied; maximum number of candidate ions per cycle, 40; charge state, 2–4 and >200 cps; dynamic exclusion, applied; cooccurrence, 1; and duration, 20 s.

The MS raw data were searched with ProteinPilot<sup>®</sup> v4.5 (Revision: 1656; Paragon Algorithm: 4.5.0.0, 1654) against the Human Swiss-Prot database (Feb, 2015). Search parameters included, Sample Type: SILAC (Lys + 8, Arg + 10); Cys Alkylation: Iodoacetamide; Digestion: Trypsin; Instrument: TripleTOF 5600; Species: Homo sapiens; ID Focus: Biological modifications; Search Effort: Thorough ID; Detected Protein Threshold: >10% Conf.; Special Factors: none. The global FDR from Fit at protein level was set to 1%.

**SWATH mass spectrometry.** To generate the ion library for the SWATH-MS analysis, peptides from different nasopharyngeal tissue samples were pooled, and analyzed in a shotgun model with a TripleTOF<sup>®</sup> 5600 MS (AB). In detail, peptides were analyzed in the high sensitivity IDA mode; precursor ion selection range was 350–1500  $m/z$ , using 0.25 s accumulation. For each precursor ion with 50 ms minimum accumulation time in the range of 100–1500  $m/z$ , the maximum precursor number per cycle was set to 40. Dynamic exclusion was applied.



Next, peptides of each tissue samples were individually analyzed in the SWATH mode<sup>18</sup>. Specifically, the 14 *m/z* precursor isolation window was used for the consecutive data-independent acquisition across the range of 400–1150 *m/z*. The accumulation time was set to 50 ms, and the total cycle time was ~3.2 s.

The SWATH-MS raw data were searched by ProteinPilot with the identical parameters to those used for the SILAC searches, except that the Sample Type was set as Identification. The resulting group file was loaded into PeakView® v1.2.0.3 (SWATH™ Acquisition MicroApp: v1.0.0.653). The Maximum Number of Protein to Import was set as the number of proteins identified at 1% global FDR from Fit. Other parameters included: The Ion Library Mass Tolerance, 75 ppm; XIC Extraction Window, 20 min; and XIC width, 75 ppm. The ion library was filtered by 6 peptides per protein and 6 fragment ions per peptide with a Peptide Confidence of 85 per protein. Variably modified or shared peptides were excluded.

**Physical and Chemical Features of Proteins.** The amino acid sequence and mass were obtained from Swiss-Prot human protein database. The isoelectric point (pI) and the charge at physiological conditions (pH = 7.4) of each protein were calculated by MATLAB bioinformatics toolbox (MathWorks, Natick, MA, USA).

**Ingenuity Pathway Analysis.** We performed core analysis of IPA for fold changes of DEPs in each MS method as we describe previously<sup>28,50</sup>. The Upstream Analysis was used to find out upstream regulators of DEPs with activation/inhibition status predictions<sup>51</sup>.

**Immunoblotting analysis.** IB was performed as we previously described<sup>46,48,51</sup>. Antibodies used were rabbit anti-CA2 pAb (1:2000, PROTEINTECH GROUP, INC., China) and HRP-conjugated goat anti-rabbit mAb (1:2000, Bioworld).

**Immunohistochemistry.** A commercial tissue array (array number: NPC1503, Super Biotek, Shanghai, China) was used to analyze the diagnostic power of CA2. IHC was performed by using rabbit anti-CA2 pAb (1:100, Sino Biological Inc., Beijing, China) and the SuperPicture™ 3rd Gen IHC Detection Kit (Thermo). The tissue array was incubated with the primary antibody, followed by the visualization with DAB staining and hematoxylin counter-staining. The staining intensity was scored as 0 (negative), 1 (weak), 2 (moderate) and 3 (strong), independently evaluated by at least two professional pathologists. A Histologic Score (intensity score × percentage of stained cell) was used to evaluate the CA2 expression.

**Statistics.** DEPs were determined by using the R package PLGEM<sup>27,28</sup>, and the statistical significance was accepted when  $P < 0.01$ . All of the correlation analyses were shown with Spearman *r*. Histologic Score of IHC results were examined by the unpaired Kolmogorov–Smirnov test (KS-test), using GraphPad Prism version 6.02 (GraphPad Software, Inc., San Diego, CA, USA), which  $P < 0.05$  was deemed significantly different. In addition, we employed the bootstrap analysis to compensate the variation from the different sample sizes provided by the tissue microarray. The 10000 bootstrap resampling was performed for each group to statistically compare the Histologic Scores of the NPC and the normal groups<sup>46,48</sup>. Both bootstrap and cluster analyses were generated by MATLAB software version R2016a. The ROC curve and the AUC were generated and computed by GraphPad Prism.

## References

- Xu, C., Chen, Y. P. & Ma, J. Clinical trials in nasopharyngeal carcinoma—past, present and future. *Chin Clin Oncol* **5**, 20, doi: 10.21037/cco.2016.03.12 (2016).
- Zhang, L. F. *et al.* Incidence trend of nasopharyngeal carcinoma from 1987 to 2011 in Sihui County, Guangdong Province, South China: an age-period-cohort analysis. *Chin J Cancer* **34**, 350–357, doi: 10.1186/s40880-015-0018-6 (2015).
- Lee, A. W. *et al.* Treatment results for nasopharyngeal carcinoma in the modern era: the Hong Kong experience. *Int J Radiat Oncol Biol Phys* **61**, 1107–1116, doi: 10.1016/j.ijrobp.2004.07.702 (2005).
- Sun, Y. *et al.* Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial. *Lancet Oncol* **17**, 1509–1520, doi: 10.1016/S1470-2045(16)30410-7 (2016).
- Ye, S. B. *et al.* Exosomal miR-24-3p impedes T-cell function by targeting FGF11 and serves as a potential prognostic biomarker for nasopharyngeal carcinoma. *J Pathol* **240**, 329–340, doi: 10.1002/path.4781 (2016).
- Spence, T., Bruce, J., Yip, K. W. & Liu, F. F. MicroRNAs in nasopharyngeal carcinoma. *Chin Clin Oncol* **5**, 17, doi: 10.21037/cco.2016.03.09 (2016).
- Zou, Z. W. *et al.* LncRNA ANRIL is up-regulated in nasopharyngeal carcinoma and promotes the cancer progression via increasing proliferation, reprogramming cell glucose metabolism and inducing side-population stem-like cancer cells. *Oncotarget*, doi: 10.18632/oncotarget.11437 (2016).
- Chan, K. H. *et al.* EBV specific antibody-based and DNA-based assays in serologic diagnosis of nasopharyngeal carcinoma. *Int J Cancer* **105**, 706–709, doi: 10.1002/ijc.11130 (2003).
- Chen, Y. *et al.* Diagnostic Value of Serum Epstein-Barr Virus Capsid Antigen-IgA for Nasopharyngeal Carcinoma: a Meta-Analysis Based on 21 Studies. *Clin Lab* **62**, 1155–1166 (2016).
- Fung, S. Y., Lam, J. W. & Chan, K. C. Clinical utility of circulating Epstein-Barr virus DNA analysis for the management of nasopharyngeal carcinoma. *Chin Clin Oncol* **5**, 18, doi: 10.21037/cco.2016.03.07 (2016).
- Leung, S. F. *et al.* Plasma Epstein-Barr viral deoxyribonucleic acid quantitation complements tumor-node-metastasis staging prognostication in nasopharyngeal carcinoma. *J Clin Oncol* **24**, 5414–5418, doi: 10.1200/JCO.2006.07.7982 (2006).
- Song, L. B. *et al.* Bmi-1 is a novel molecular marker of nasopharyngeal carcinoma progression and immortalizes primary human nasopharyngeal epithelial cells. *Cancer Res* **66**, 6225–6232, doi: 10.1158/0008-5472.CAN-06-0094 (2006).
- Wu, T. *et al.* HMGB1 overexpression as a prognostic factor for survival in cancer: a meta-analysis and systematic review. *Oncotarget*, doi: 10.18632/oncotarget.10413 (2016).
- Lian, Y. F. *et al.* Upregulation of KLHD4 Predicts a Poor Prognosis in Human Nasopharyngeal Carcinoma. *PLoS One* **11**, e0152820, doi: 10.1371/journal.pone.0152820 (2016).

15. Xiao, L., Xiao, T., Wang, Z. M., Cho, W. C. & Xiao, Z. Q. Biomarker discovery of nasopharyngeal carcinoma by proteomics. *Expert Rev Proteomics* **11**, 215–225, doi: 10.1586/14789450.2014.897613 (2014).
16. Chen, Z. *et al.* Identification of nasopharyngeal carcinoma metastasis-related biomarkers by iTRAQ combined with 2D-LC-MS/MS. *Oncotarget* **7**, 34022–34037, doi: 10.18632/oncotarget.9067 (2016).
17. Omenn, G. S. *et al.* Metrics for the Human Proteome Project 2015: Progress on the Human Proteome and Guidelines for High-Confidence Protein Identification. *J Proteome Res* **14**, 3452–3460, doi: 10.1021/acs.jproteome.5b00499 (2015).
18. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111 016717, doi: 10.1074/mcp.O111.016717 (2012).
19. Aebersold, R., Bensimon, A., Collins, B. C., Ludwig, C. & Sabido, E. Applications and Developments in Targeted Proteomics: From SRM to DIA/SWATH. *Proteomics* **16**, 2065–2067, doi: 10.1002/pmic.201600203 (2016).
20. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **4**, 222, doi: 10.1038/msb.2008.61 (2008).
21. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **9**, 555–566, doi: 10.1038/nmeth.2015 (2012).
22. Liu, Y. *et al.* Glycoproteomic analysis of prostate cancer tissues by SWATH mass spectrometry discovers N-acylethanolamine acid amidase and protein tyrosine kinase 7 as signatures for tumor aggressiveness. *Mol Cell Proteomics* **13**, 1753–1768, doi: 10.1074/mcp.M114.038273 (2014).
23. Lin, Q. *et al.* Analysis of colorectal cancer glyco-secretome identifies laminin beta-1 (LAMB1) as a potential serological biomarker for colorectal cancer. *Proteomics* **15**, 3905–3920, doi: 10.1002/pmic.201500236 (2015).
24. Wu, W., Yong, W. W. & Chung, M. C. A simple biomarker scoring matrix for early gastric cancer detection. *Proteomics*, doi: 10.1002/pmic.201600194 (2016).
25. Ortea, I., Rodriguez-Ariza, A., Chicano-Galvez, E., Arenas Vacas, M. S. & Jurado Gamez, B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *J Proteomics* **138**, 106–114, doi: 10.1016/j.jpro.2016.02.010 (2016).
26. Pavelka, N. *et al.* A power law global error model for the identification of differentially expressed genes in microarray data. *BMC Bioinformatics* **5**, 203, doi: 10.1186/1471-2105-5-203 (2004).
27. Pavelka, N. *et al.* Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics* **7**, 631–644, doi: 10.1074/mcp.M700240-MCP200 (2008).
28. Chen, Z. *et al.* Cytoskeleton-centric protein transportation by exosomes transforms tumor-favorable macrophages. *Oncotarget*, doi: 10.18632/oncotarget.11794 (2016).
29. Wu, C. C. *et al.* Identification of potential serum markers for nasopharyngeal carcinoma from a xenografted mouse model using Cy-dye labeling combined with three-dimensional fractionation. *Proteomics* **8**, 3605–3620, doi: 10.1002/pmic.200701034 (2008).
30. Wang, T. *et al.* Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res* **41**, 4743–4754, doi: 10.1093/nar/gkt178 (2013).
31. Guo, J., Lian, X., Zhong, J., Wang, T. & Zhang, G. Length-dependent translation initiation benefits the functional proteome of human cells. *Mol Biosyst* **11**, 370–378, doi: 10.1039/c4mb00462k (2015).
32. Huang, Q. *et al.* SWATH enables precise label-free quantification on proteome scale. *Proteomics* **15**, 1215–1223, doi: 10.1002/pmic.201400270 (2015).
33. Wykoff, C. C. *et al.* Hypoxia-inducible expression of tumor-associated carbonic anhydrases. *Cancer Res* **60**, 7075–7083 (2000).
34. Supuran, C. T. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat Rev Drug Discov* **7**, 168–181, doi: 10.1038/nrd2467 (2008).
35. Viikila, P. *et al.* Carbonic anhydrase enzymes II, VII, IX and XII in colorectal carcinomas. *World J Gastroenterol* **22**, 8168–8177, doi: 10.3748/wjg.v22.i36.8168 (2016).
36. Kurono, S. *et al.* Identification of potential breast cancer markers in nipple discharge by protein profile analysis using two-dimensional nano-liquid chromatography/nanoelectrospray ionization-mass spectrometry. *Proteomics Clin Appl* **10**, 605–613, doi: 10.1002/prca.201500016 (2016).
37. Henry, E. K. *et al.* Carbonic anhydrase enzymes regulate mast cell-mediated inflammation. *J Exp Med* **213**, 1663–1673, doi: 10.1084/jem.20151739 (2016).
38. Gordon, S. C. *et al.* Antibodies to carbonic anhydrase in patients with immune cholangiopathies. *Gastroenterology* **108**, 1802–1809 (1995).
39. Alver, A. *et al.* Carbonic anhydrase II autoantibody and oxidative stress in rheumatoid arthritis. *Clin Biochem* **44**, 1385–1389, doi: 10.1016/j.clinbiochem.2011.09.014 (2011).
40. Hornef, M. W., Wagner, H. J., Kruse, A. & Kirchner, H. Cytokine production in a whole-blood assay after Epstein-Barr virus infection *in vivo*. *Clin Diagn Lab Immunol* **2**, 209–213 (1995).
41. Fang, W. *et al.* EBV-driven LMP1 and IFN- $\gamma$  up-regulate PD-L1 in nasopharyngeal carcinoma: Implications for oncotargeted therapy. *Oncotarget* **5**, 12189–12202, doi: 10.18632/oncotarget.2608 (2014).
42. Wu, M., Li, X., Li, X. & Li, G. Signaling Transduction Network Mediated by Tumor Suppressor/Susceptibility Genes in NPC. *Curr Genomics* **10**, 216–222, doi: 10.2174/138920209788488481 (2009).
43. Ma, J. *et al.* Role of a novel EGF-like domain-containing gene NGX6 in cell adhesion modulation in nasopharyngeal carcinoma cells. *Carcinogenesis* **26**, 281–291, doi: 10.1093/carcin/bgh312 (2005).
44. Li, Z. *et al.* Macrophage migration inhibitory factor enhances neoplastic cell invasion by inducing the expression of matrix metalloproteinase 9 and interleukin-8 in nasopharyngeal carcinoma cell lines. *Chin Med J (Engl)* **117**, 107–114 (2004).
45. Lo, M. C. *et al.* Role of MIF/CXCL8/CXCR2 signaling in the growth of nasopharyngeal carcinoma tumor spheres. *Cancer Lett* **335**, 81–92, doi: 10.1016/j.canlet.2013.01.052 (2013).
46. Chen, Y. *et al.* Identification of Missing Proteins Defined by Chromosome-Centric Proteome Project in the Cytoplasmic Detergent-Insoluble Proteins. *J Proteome Res* **14**, 3693–3709, doi: 10.1021/pr501103r (2015).
47. Geiger, T. *et al.* Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat Protoc* **6**, 147–157, doi: 10.1038/nprot.2010.192 (2011).
48. Yang, L. *et al.* Finding Missing Proteins from the Epigenetically Manipulated Human Cell with Stringent Quality Criteria. *J Proteome Res* **14**, 3645–3657, doi: 10.1021/acs.jproteome.5b00480 (2015).
49. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359–362, doi: 10.1038/nmeth.1322 (2009).
50. Guo, J. *et al.* Phosphoproteome Characterization of Human Colorectal Cancer SW620 Cell-Derived Exosomes and New Phosphosite Discovery for C-HPP. *J Proteome Res* **15**, 4060–4072, doi: 10.1021/acs.jproteome.6b00391 (2016).
51. Shen, S. *et al.* Functional proteomics revealed IL-1 $\beta$  amplifies TNF downstream protein signals in human synoviocytes in a TNF-independent manner. *Biochem Biophys Res Commun* **450**, 538–544, doi: 10.1016/j.bbrc.2014.06.008 (2014).

## Acknowledgements

This work was supported by the National Basic Research Program “973” of China (2014CBA02000), the National Natural and Science Foundation of China (81372135), Key Project for Research and Development of Guangdong

province (2016B020238002), and Key Special Project on the Integration of Industry, Education and Research of Guangzhou (201604020002) to T.W.; Fundo para o Desenvolvimento das Ciências e da Tecnologia de Macau (025/2013/A) to T.S.M., the Fundamental Research Funds for the Central Universities of China to T.Z. and the Jinan University's Scientific Research Creativeness Cultivation Project for Outstanding Undergraduates Recommended for Postgraduate Study to Y.L.

### Author Contributions

T.W. and T.Z. conceived the idea, designed the study. T.W. wrote the paper. Y.L. contributed to the proteomic experiments, most of the data analyses and paper writing. T.S.M. contributed to data analyses, paper writing, and clinical sample acquisition. X.L. assisted the clinical sample preparation. W.Z., Y.C., J.G. and X.C. contributed to the proteomics analyses.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Luo, Y. *et al.* SWATH-based proteomics identified carbonic anhydrase 2 as a potential diagnosis biomarker for nasopharyngeal carcinoma. *Sci. Rep.* 7, 41191; doi: 10.1038/srep41191 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017