



## OPEN

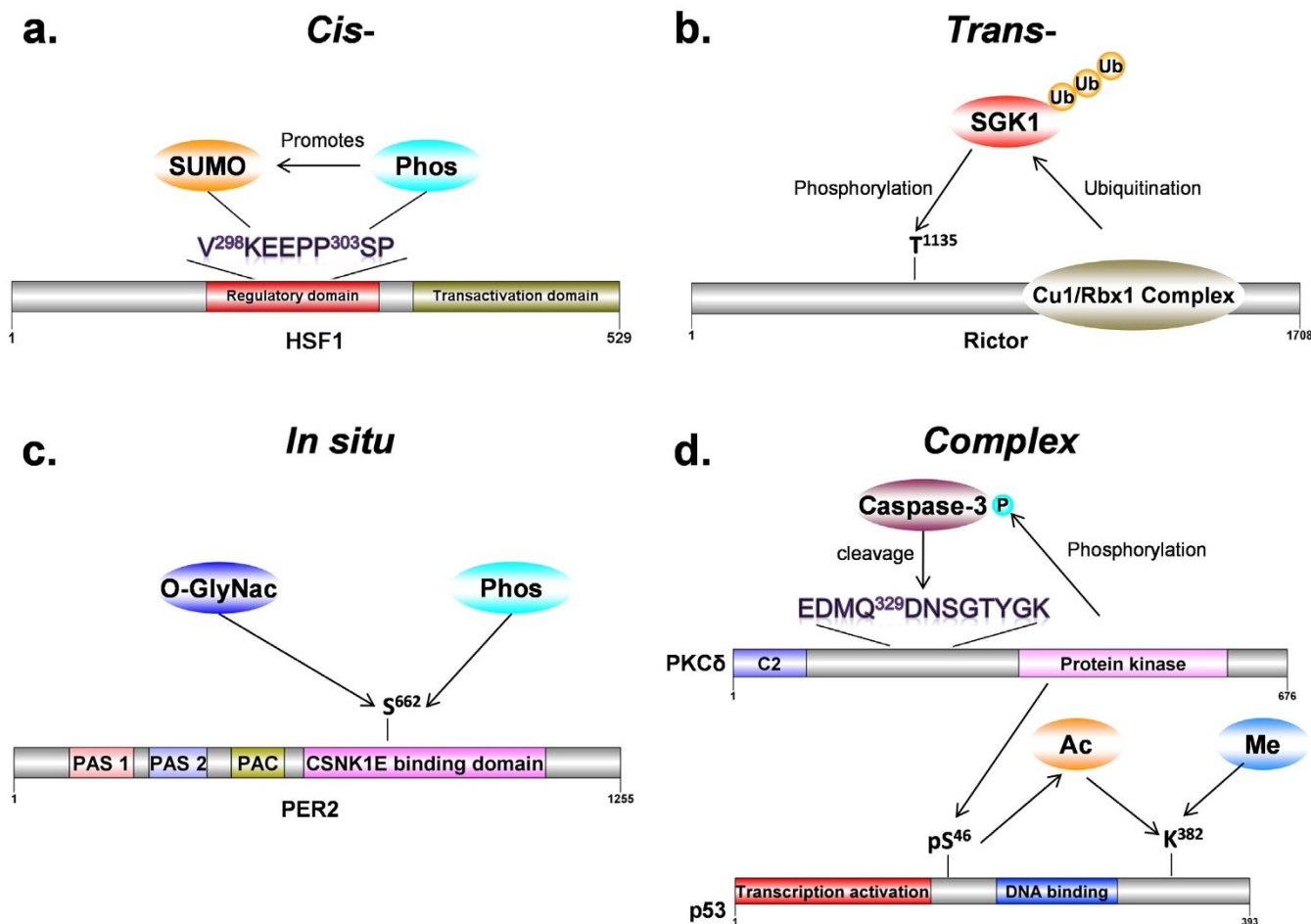
SUBJECT AREAS:  
PHOSPHORYLATION  
PROTEOME INFORMATICSReceived  
20 August 2014Accepted  
11 November 2014Published  
5 December 2014Correspondence and  
requests for materials  
should be addressed to  
Y.X. (xueyu@hust.edu.  
cn)\* These authors  
contributed equally to  
this work.

# Systematic analysis of the *in situ* crosstalk of tyrosine modifications reveals no additional natural selection on multiply modified residues

Zhicheng Pan<sup>1\*</sup>, Zexian Liu<sup>1\*</sup>, Han Cheng<sup>1</sup>, Yongbo Wang<sup>1</sup>, Tianshun Gao<sup>1</sup>, Shahid Ullah<sup>1</sup>, Jian Ren<sup>2</sup> & Yu Xue<sup>1</sup><sup>1</sup>Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China, <sup>2</sup>State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, China.

Recent studies have indicated that different post-translational modifications (PTMs) synergistically orchestrate specific biological processes by crosstalks. However, the preference of the crosstalk among different PTMs and the evolutionary constraint on the PTM crosstalk need further dissections. In this study, the *in situ* crosstalk at the same positions among three tyrosine PTMs including sulfation, nitration and phosphorylation were systematically analyzed. The experimentally identified sulfation, nitration and phosphorylation sites were collected and integrated with reliable predictions to perform large-scale analyses of *in situ* crosstalks. From the results, we observed that the *in situ* crosstalk between sulfation and nitration is significantly under-represented, whereas both sulfation and nitration prefer to co-occupy with phosphorylation at same tyrosines. Further analyses suggested that sulfation and nitration preferentially co-occur with phosphorylation at specific positions in proteins, and participate in distinct biological processes and functions. More interestingly, the long-term evolutionary analysis indicated that multi-PTM targeting tyrosines didn't show any higher conservation than singly modified ones. Also, the analysis of human genetic variations demonstrated that there is no additional functional constraint on inherited disease, cancer or rare mutations of multiply modified tyrosines. Taken together, our systematic analyses provided a better understanding of the *in situ* crosstalk among PTMs.

Through modification of proteins by covalent attachment of other functional groups or by proteolytic cleavage, post-translational modifications (PTMs) temporally and spatially affect protein activity, stability and trafficking, regulate most of biological and physiological functions, and determine the cellular dynamics and plasticity<sup>1–3</sup>. In particular, one PTM can crosstalk with other PTMs to synergistically orchestrate specific processes through three distinct mechanisms<sup>4–13</sup>. First, different PTMs can co-occur in the same proteins and crosstalk in a *cis*-regulatory manner<sup>4,5</sup>. For example, the phosphorylation of the motif ΨKXEXXpSP (Ψ is a hydrophobic residue, X is any amino acid, pS is a phosphorylatable serine) at S303 of HSF1, a heat shock transcription factor, enhances the adjacent lysine sumoylation at K298<sup>4</sup> (Fig. 1a). Also, a recently computational analysis suggested that a considerable proportion of acetylated lysines might influence the PTMs such as phosphorylation, methylation and ubiquitination of adjacent sites<sup>6</sup>. Second, one PTM can regulate another PTM by modifying its cognate enzymes and *vice versa*, in a *trans*-regulatory mode<sup>7–9</sup>. For example, the E3 ubiquitin ligase complex of Rictor/Cullin-1/Rbx1 ubiquitinates an AGC kinase of SGK1 and promote its degradation, whereas the T1135 of Rictor can be phosphorylated by multiple AGC kinases including SGK1, and such a phosphorylation disrupts the interaction of Rictor and Cullin-1 to inhibit the ubiquitination of SGK1<sup>8</sup> (Fig. 1b). Third, multiple PTMs can “*in situ*” interplay with each other by competitively modifying same residues<sup>10–13</sup>. For example, a circadian clock protein PER2 is competitively O-GlcNAcylated and phosphorylated at S662, while the *in situ* crosstalk precisely regulates the PER2 repressor activity<sup>13</sup> (Fig. 1c). In addition, different types of PTM crosstalks can simultaneously occur and regulate biological functions in a complicated manner. For example, PKCδ phosphorylates Caspase-3<sup>14</sup>, which reciprocally cleaves PKC-δ as a *trans*-crosstalk<sup>15</sup> (Fig. 1d). Also, p53 can be activated through the PKCδ-mediated phosphorylation of S46<sup>16</sup>, and such a phosphorylation promotes its



**Figure 1** | The examples for different types of PTM crosstalks. (a) *Cis*-crosstalk, phosphorylation at S303 of HSF1 promotes its K298 sumoylation<sup>4</sup>. (b) *Trans*-crosstalk between ubiquitination of SGK1 by Rictor/Cullin-1/Rbx1 and phosphorylation of Rictor by SGK1<sup>8</sup>. (c) *In situ* crosstalk, protein PER2 is competitively O-GlcNAcylated and phosphorylated at S662<sup>13</sup>. (d) A complex crosstalk among PKCδ, Caspase-3, and p53 that different types of PTM crosstalks can simultaneously occur and regulate biological functions.

acetylation at K382 through a *cis*-crosstalk<sup>17</sup> (Fig. 1d). Again, lysine acetylation and methylation compete at K382 for modulating the p53 transcription activity<sup>18</sup> (Fig. 1d).

Recently, the *in situ* crosstalk of PTMs has emerged to be an intriguing topic and attracted much attention. The first question is whether PTMs recognizing the same type of amino acid residues are significantly co-occurred. For this issue, Wang *et al.* totally characterized 141 O-GlcNAcylation and >350 phosphorylation sites from the mitotic spindle and midbody samples of human HeLa cells, and observed that both PTMs extensively co-occupy at same serine/threonine (S/T) residues or adjacent regions<sup>19</sup>. However, in a following study, Trinidad *et al.* systematically identified ~1750 O-GlcNAcylation and ~16,500 phosphorylation sites from mouse synaptosomes, and revealed that only 8% (135) of O-GlcNAcylation sites are also phosphorylated<sup>20</sup>. With a computational simulation, they observed that the overlap of two PTMs is not significant and nearly equal to be random<sup>20</sup>. The result was supported by an analysis of lysine succinylation and acetylation, in which Park *et al.* detected that only ~22% of succinylation sites were also acetylated at the same lysines from mouse embryonic fibroblasts (MEFs)<sup>21</sup>. However, in a different analysis, Weinert *et al.* demonstrated that succinylation and acetylation extensively co-occupy at same residues from prokaryotes and eukaryotes, whereas 66% of *E. coli*, 56% of *S. cerevisiae*, and 57% of mouse liver succinylation sites are overlapped with acetylation<sup>22</sup>. Interestingly, they also detected that only 27% of succinylation sites are acetylated in human HeLa

cells<sup>22</sup>. In this regard, the preference of the *in situ* crosstalk may not only be dependent on the PTM types, but also exhibit a tissue-specific manner. The second and more important question is whether amino acids targeted by multiple PTMs undergo additional evolutionary pressure against singly modified sites. Because the co-occurrence of O-GlcNAcylation and phosphorylation are similar to be random, Trinidad *et al.* suggested that there is little or no evolutionary pressure for the *in situ* crosstalk between two PTMs<sup>20</sup>. In contrast, by analyzing the potentially *in situ* crosstalk of protein lysine modifications (PLMs), although the long-term phylogenetic analysis among multi-species revealed multi-PTM targeting sites are only slightly more conserved than singly modified lysines, the short-term evolutionary results demonstrated that multiply modified lysines dramatically enriched more human disease-associated and rare variations<sup>23</sup>. Thus, Gray *et al.* concluded that mutations of lysines with multi-PTMs undergo much greater purifying selection against singly modified lysine residues<sup>23</sup>.

None of two questions mentioned above have been fully addressed, while different viewpoints were raised for both problems. To further clarify the controversial issues, more efforts should be taken. Recent progresses in the identification of substrates and sites for tyrosine modifications, such as sulfation<sup>24–26</sup>, nitration<sup>27,28</sup> and tyrosine phosphorylation<sup>29</sup> provided a great opportunity to address the above two problems, as an independent, justified and unbiased framework. Also, previous studies observed that the *in situ* crosstalk truly occurs among the three PTMs. For example, both phosphor-



ylation and nitration can co-occur at Y125, Y133 and Y136 of  $\alpha$ -synuclein (UniProt ID: P37840)<sup>30,31</sup>. Moreover, human Gastrin (P01350) was identified to be phosphorylated by *v*-Src at Y87<sup>32</sup>, the site which is also modulated by sulfation<sup>33</sup>.

From the scientific literature, public databases and our previous studies<sup>27</sup>, we totally collected 273 sulfation sites, 1,050 nitration sites and 24,242 phosphorylation sites in 171, 539 and 11,034 proteins, respectively. By integrating the datasets together, we observed 2 sulfation-nitration, 3 sulfation-phosphorylation and 183 nitration-phosphorylation site-specific crosstalks (Table S1). Using known sulfation sites as the training data, we developed a novel predictor of GPS-TSP (Tyrosine sulfation predictor, available at <http://tsp.biocuckoo.org>), which exhibited superior performance than other existed tools by comparison. Together with a nitration sites predictor of GPS-YNO2<sup>27</sup>, we predicted potential nitration sites in sulfated substrates, and *vice versa*. Our results demonstrated the *in situ* crosstalk between sulfation and nitration was significantly under-represented, and two PTMs prefer to regulate different functions. Moreover, by predicting potential sulfation and nitration sites in known phosphorylated substrates, we detected that a considerable proportion of tyrosine phosphorylation sites (24.5%) might be modified by either sulfation or nitration, and both sulfation and nitration preferentially target phosphorylated tyrosines rather than non-phosphorylated sites by 1.71- and 1.45-fold, respectively. Again, statistical results suggested sulfation and nitration prefer to crosstalk with phosphorylation in regulating potentially distinct biological processes and functions. Further analyses of the sequence and structure preferences revealed that different types of tyrosine modifications prefer to co-occur at distinct structural positions in proteins. Interestingly, whether PTMs prefer to *in situ* crosstalk or not has no correlation with the evolutionary constraint on multiply modified tyrosines. In our results, although three types of tyrosine modifications have distinct preferences for the *in situ* crosstalk, the long-term evolutionary analysis across eight vertebrate species revealed that multiply modified tyrosines are not more conserved than unmodified ones. In addition, by mapping all known or predicted tyrosine modification sites to human genetic variations, we observed that multiply modified tyrosines didn't significantly enrich inherited disease, cancer or rare mutations. Taken together, our results suggested that there is no functional constraint on multiply modified tyrosines, and the *in situ* crosstalk of tyrosine modification does not need additional natural selection.

## Results

**Development of GPS-TSP for the prediction of sulfation sites.** For systematically analyzing the tyrosine sulfation and its crosstalk with other tyrosine modifications, a prerequisite is to establish a comprehensive and reliable dataset. However, the experimental identification of sulfation substrates is still labor-intensive and time-consuming, while only 273 known sulfation sites were collected. Thus, computational prediction of sulfation sites from protein primary sequences can serve as an alternative solution. Before developing the tool, we first analyzed the sequence profile of 202 non-redundant sulfation sites by WebLogo<sup>34</sup> (Fig. 2a). Although a previous analysis proposed that the +2 position didn't contain any information with a limited data set<sup>26</sup>, our analysis clearly exhibited that negatively charged residues including aspartic acids and glutamic acids were enriched around the sulfated tyrosine especially the position -1, while tyrosine was founded to be over-presented in position +1, +2, and upstream positions. In addition, glycine and glutamine were observed to frequently occur near the sulfated tyrosine. These observations suggested that there were considerable sequence preferences around sulfation sites (Fig. 2a). Then we designed a software package of GPS-TSP for predicting tyrosine sulfation sites, with a previously developed algorithm of

Group-based Prediction System (GPS)<sup>27</sup> (Fig. 2b). More details on the algorithm were shown in Supplemental experimental procedures.

To evaluate the prediction performance and robustness, the LOO validation and 4-, 6-, 8-, 10-fold cross-validations were performed. The corresponding ROC curves were presented, while the AROC values were calculated as 0.9424 (LOO), 0.9527 (4-fold), 0.9595 (6-fold), 0.9547 (8-fold) and 0.9563 (10-fold), respectively (Fig. 2c). Since the results of the 4-, 6-, 8- and 10-fold cross-validations were closely similar to the LOO validation, the prediction is evidently stable and robust. The performance of the LOO validation was used for the cut-off setting and further comparison, and the three thresholds of high, medium and low were selected with the *Sp* values of 85%, 90% and 95%, respectively (Table 1).

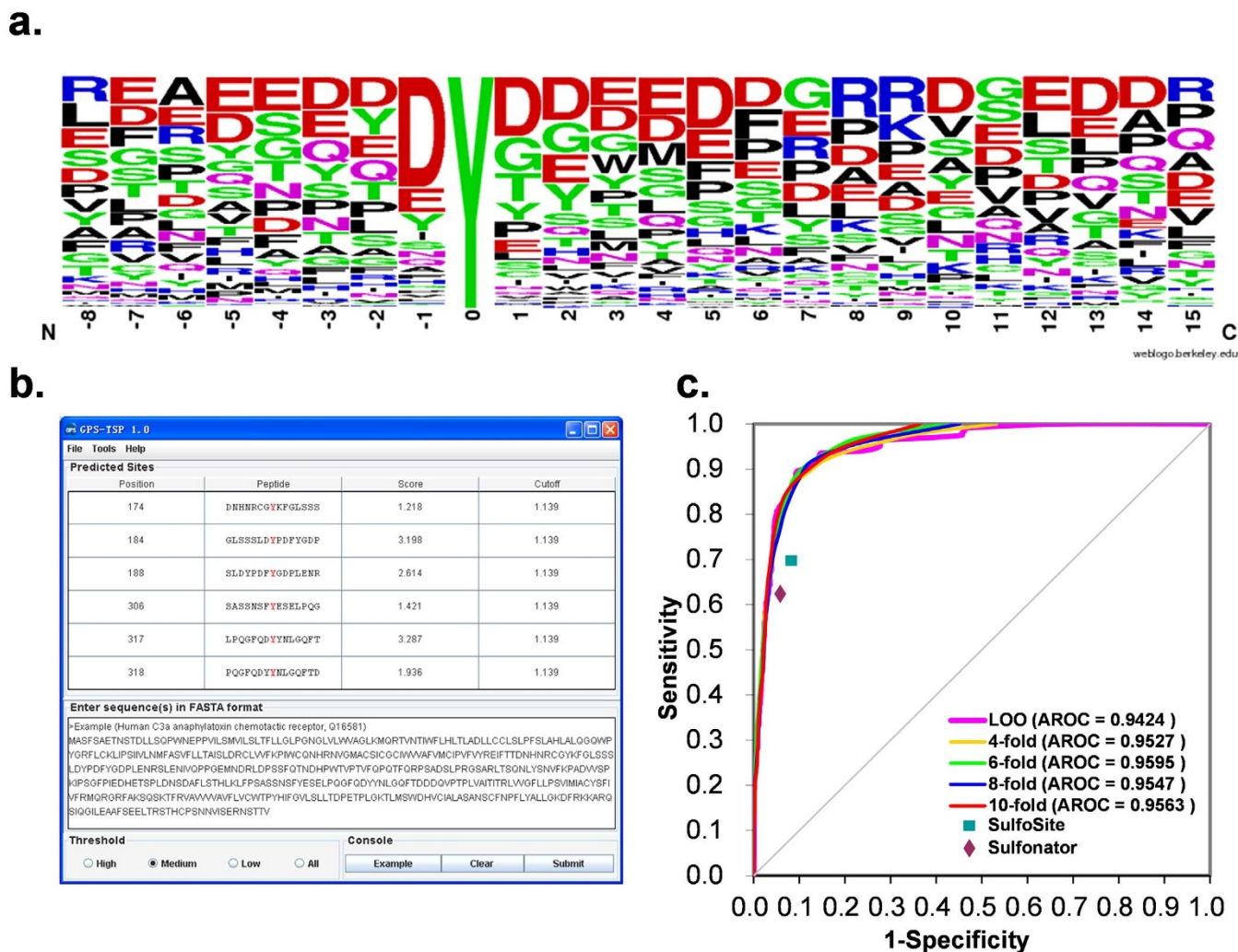
Previously, a number of computational studies were performed for predicting protein sulfation sites, however, only Sulfinator<sup>35</sup> and SulfoSite<sup>36</sup> were implemented into online services. To demonstrate the superiority of GPS-TSP, here we used its training dataset of 202 sulfation sites and 1027 negative sites to evaluate the performances of two predictors. To avoid any bias, we compared the *Sn* values for GPS-TSP, Sulfinator and SulfoSite at the equal level of *Sp* values (Table 1, Fig. 2c). When the *Sp* value was ~93%, the *Sn* value of GPS-TSP and Sulfinator were 83.17% and 61.79%, respectively (Table 1). Also, when the *Sp* value was ~91%, the *Sn* of GPS-TSP (87.12%) was much greater than SulfoSite (69.73%) (Table 1). In this regard, the prediction performance of GPS-TSP 1.0 is much better than other existed tools.

### Sulfation and nitration prefer not to co-occur at same tyrosines.

Besides sulfation, protein tyrosine nitration (PTN) is also an important PTM, and predominantly implicated in a variety of fundamental processes such as RNA splicing, mRNA processing and translation<sup>27,28</sup>, whereas sulfation was proposed to mostly occur in the secretory pathway<sup>24-26</sup>. To further clarify whether two PTMs prefer to target and regulate distinct processes and functions, the experimentally identified nitrated substrates were taken from our previous study<sup>27</sup>. In order to analyze and compare the functional abundance and diversity of sulfation and PTN, we downloaded the gene ontology (GO) (March 31<sup>th</sup>, 2012) association files from the GOA database at the EBI (<http://www.ebi.ac.uk/goa>)<sup>37</sup>. There were 44,741 human proteins, 65 sulfated proteins and 326 nitrated proteins annotated with at least one GO term.

With the hypergeometric distribution<sup>27</sup>, we statistically analyzed the enriched biological processes, molecular functions and cellular components with GO annotations for sulfated (Fig. 3a, *p*-value < 10<sup>-11</sup>) and nitrated (Fig. 3b, *p*-value < 10<sup>-14</sup>) proteins. For the sulfated substrates, the top five most enriched biological processes are chemokine-mediated signaling pathway (GO:0070098), chemotaxis (GO:0006935), inflammatory response (GO:0006954), elevation of cytosolic calcium ion concentration (GO:0007204), and cell adhesion (GO:0007155), which are consistent with previously experimental observations<sup>24-26</sup> (Fig. 3a). However, the top five most significant processes of nitration are gene expression (GO:0010467), cellular protein metabolic process (GO:0044267), mRNA metabolic process (GO:0016071), RNA metabolic process (GO:0016070), and translational initiation (GO:0006413) (Fig. 3b). In this regard, we proposed that sulfation and nitration are preferentially involved in distinct processes.

To confirm this analysis, we compared the functional diversity of sulfated and nitrated proteins using the Yates' Chi-square ( $\chi^2$ ) test<sup>27</sup> (Fig. 3c, *p*-value < 10<sup>-7</sup>). Indeed, sulfation was found to be preferentially involved in modifying membrane and extracellular proteins, while nitration prefers to attack substrates in cytosol (Fig. 3c). In addition, we used the high thresholds of GPS-TSP and GPS-YNO2<sup>27</sup>, and directly predicted potential nitration sites from sulfated substrates and *vice versa* (Table 2). With the hypergeometric distribution, the results clearly demonstrated that sulfation and nitration prefer not to *in situ* interplay at the same positions (Table 3).



**Figure 2** | The development of GPS-TSP 1.0. (a) The sequence logo of sulfation sites. (b) The snapshot of GPS-TSP with the example of human C3a complement anaphylatoxin chemotactic receptor (C3aR, Q16581). (c) The ROC curves and AROC values for the LOO validation and 4-, 6-, 8-, 10-fold cross-validations.

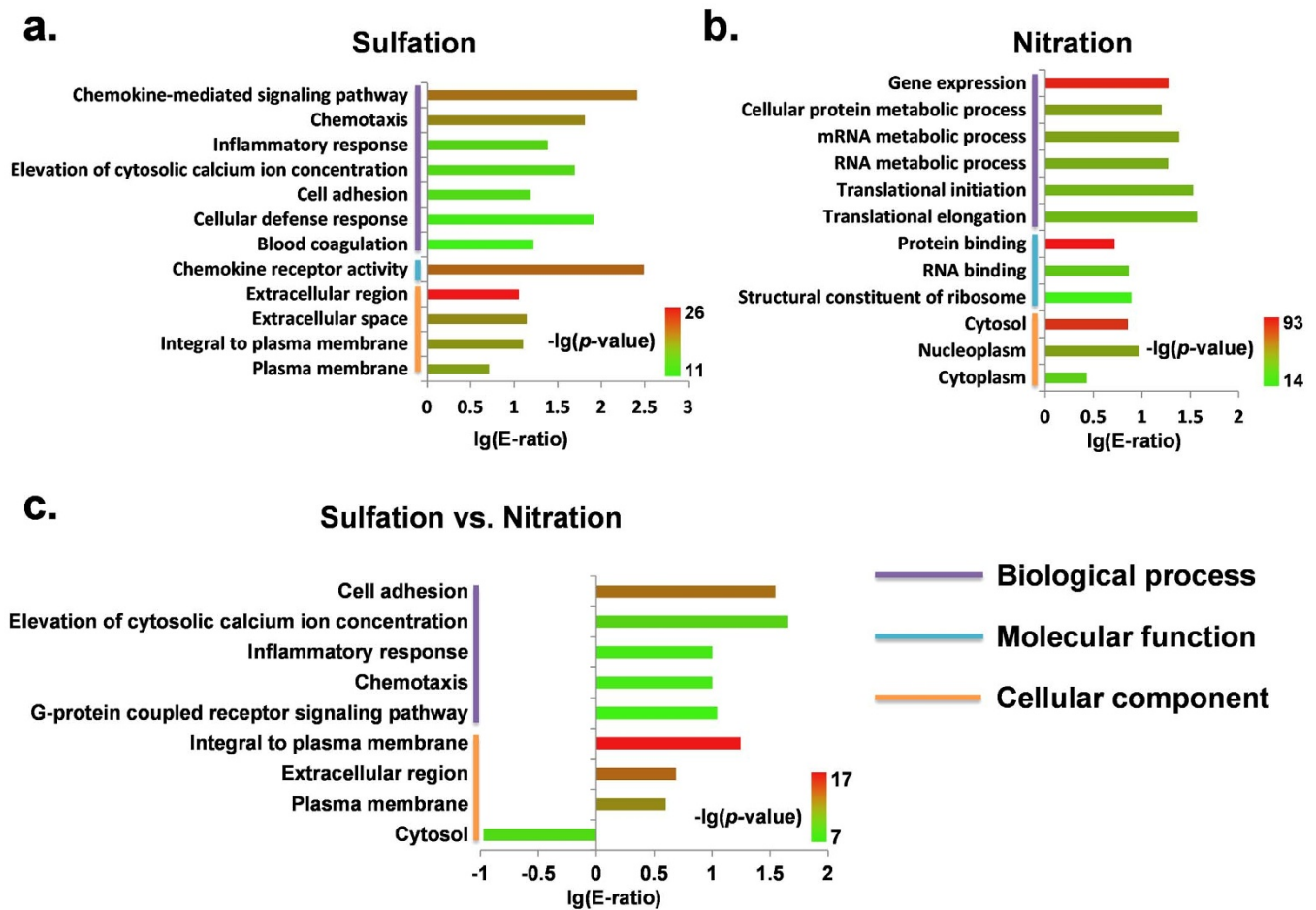
**Sulfation and nitration prefer to *in situ* crosstalk with tyrosine phosphorylation.** To further dissect the relations among tyrosine modifications, the *in situ* crosstalk between sulfation or nitration and phosphorylation were surveyed. We used GPS-TSP and GPS-YNO2 with the high thresholds to predict potential sulfation and nitration sites in tyrosine phosphorylated substrates (Table 3 & Table S4). Totally, it was observed that 5,939 (24.5%) known

phosphorylation sites might be modified by either sulfation (2,913, 12.0%) or nitration (3,689, 15.2%) (Table S2). Although the *p*-values in *D. melanogaster* and *C. elegans* were not much significant due to the data limitation, the statistical results suggested that sulfation and nitration prefer to occur at phosphorylated tyrosines rather than non-phosphorylated tyrosines with the enrichment ratios (E-ratios) of 1.71 and 1.45, respectively (Table 3). Because our dataset of tyrosine phosphorylation contained sites identified from large-scale studies which couldn't guarantee all data to be real phosphorylation sites, we further obtained 3,254 well curated tyrosine phosphorylation sites from Phospho.ELM (version 9.0, released in April 2010)<sup>38</sup>. Again, the results still suggested that sulfation and nitration prefer to *in situ* crosstalk with phosphorylation (Table S3).

With the hypergeometric distribution, we statistically analyzed the over- or under-represented GO terms in phosphorylated substrates which might also be competitively regulated by sulfation (Fig. 4a, *p*-value < 10<sup>-5</sup>) or nitration (Fig. 4b, *p*-value < 10<sup>-5</sup>) at the same residues, separately. We only considered the predicted sulfated and nitrated proteins, whose sites were predicted from known phosphorylation sites. Clearly, sulfation prefers to *in situ* crosstalk with phosphorylation in a variety of biological processes, such as peptidyl-tyrosine phosphorylation (GO:0018108), positive regulation of phosphatidylinositol 3-kinase activity (GO:0043552), and

**Table 1** | Comparison of the GPS-TSP with other tools. For construction of the software package, the three thresholds of high, medium and low were chosen, while the medium threshold was selected as the default cut-off value. For comparison, we fixed the *Sp* values of GPS-TSP so as to be similar or identical to the other methods and compared the *Sn* values

Tool	Threshold	Ac	Sn	Sp	MCC
GPS-TSP	High	92.60%	79.70%	95.13%	0.7354
	Medium	90.23%	89.60%	90.36%	0.7066
	Low	86.57%	93.56%	85.20%	0.6519
Sulfonator		91.54%	83.17%	93.18%	0.7161
		90.48%	87.12%	91.13%	0.7034
SulfoSite		87.63%	61.79%	93.02%	0.5588
		88.45%	69.73%	91.09%	0.5403



**Figure 3** | Statistical analyses of GO annotations for sulfated and nitrated proteins. The enriched GO terms for sulfated proteins (a) or nitrated proteins (b) in comparison with proteome. (c) Comparison of GO terms between sulfated and nitrated proteins. E-ratio, enrichment ratio.

transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169) (Fig. 4a). In contrast, nitration prefers to *in situ* cross-talk with phosphorylation in blood coagulation (GO:0007596), peptidyl-tyrosine phosphorylation (GO:0018108), and response to unfolded protein (GO:0006986) (Fig. 4b). By comparison, it was observed that sulfation and nitration prefer to *in situ* crosstalk with phosphorylation in distinct biological processes and functions. Moreover, we mapped all phosphorylated proteins to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways<sup>39</sup>, and performed the statistical analyses for the *in situ* crosstalks. The results suggested that the sulfation-phosphorylation crosstalk is significantly enriched in pathways of leukocyte transendothelial migration (hsa04670), tight junction (hsa04530), and adherens junction

(hsa04520) (Fig. 4c), whereas the nitration-phosphorylation cross-talk prefers to target natural killer cell mediated cytotoxicity (hsa04650), viral carcinogenesis (hsa05203), and tight junction (hsa04530) (Fig. 4d). Taken together, although both sulfation and nitration prefer to co-occupy with phosphorylation, the sulfation-phosphorylation and nitration-phosphorylation crosstalks may preferentially occur in different pathways.

#### The sequence and structure preferences of the tyrosine modifications.

Since the relations among sulfation, nitration and phosphorylation are complicated, we further analyzed the sequence and structure preferences of the tyrosine modifications for dissecting the basic features of the *in situ* crosstalks. The experimentally identified PTM tyrosines including 273 sulfation, 1050 nitration and 24,242 phosphorylation sites were employed for the analysis. Furthermore, 17,015 and 25,306 predicted sulfation and nitration sites in phosphorylated substrates were also analyzed, while 2913 and 3689 phosphorylation sites were predicted to be sulfation- and nitration-phosphorylation crosstalk sites. With these datasets, we first analyzed the position distributions of modified tyrosines. In the result, known sulfation sites preferentially occur at N-terminal or C-terminal but not middle of proteins, whereas predicted sulfation sites slightly prefer to locate at C-terminal, as well as other types of tyrosine modifications and their crosstalks (Fig. 5a).

Also, various structural features were analyzed. From the results of secondary structures, it was observed that both known sulfation and nitration sites are enriched in Coil, whereas predicted nitration sites are deprived in Coil and sulfation sites are enriched in  $\beta$ -Strand (Fig. 5b). This might be due to the number of known sulfation and

**Table 2** | Predicted nitration sites in sulfated substrates and vice versa. The hypergeometric distribution was adopted. a. The number of known sulfated or nitrated substrates; b. The number of all tyrosine residues; c. The number of known sulfation or nitration sites; d. The number of totally predicted nitration or sulfation sites; e. The number of predicted nitration sites on known sulfation residues, and vice versa

PTM	Known			Predicted			E-ratio	p-value
	Sub. <sup>a</sup>	Total <sup>b</sup>	Site <sup>c</sup>	Total <sup>d</sup>	Site <sup>e</sup>			
<b>Sulfation</b>	171	1,518	273	525	68	0.83	1.02E-03	
<b>Nitration</b>	539	8,356	1,050	549	46	0.67	8.67E-04	



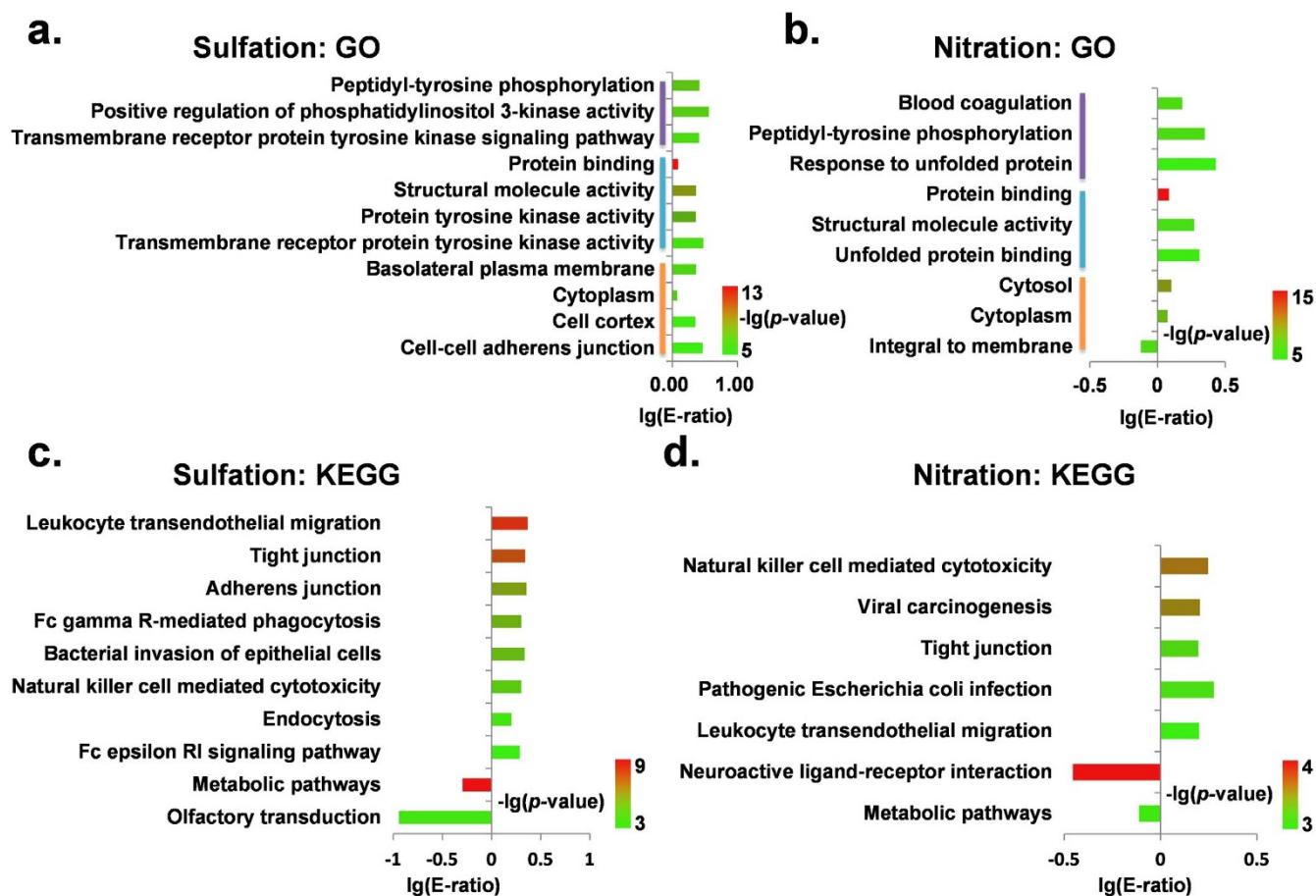
**Table 3** | Predicted sulfation and nitration sites on known phosphorylated substrates, which were collected from *H. sapiens* (HS), *M. musculus* (MM), *D. melanogaster* (DM) and *C. elegans* (CE). The hypergeometric distribution was adopted. *a*. The number of all tyrosine residues; *b*. The number of known phosphorylation sites; *c*. The number of totally predicted sulfation or nitration sites; *d*. The number of predicted sulfation or nitration sites on known phosphorylation residues

Organism	Phosphorylation		Sulfation				Nitration			
	Total <sup>a</sup>	Site <sup>b</sup>	Total <sup>c</sup>	Site <sup>d</sup>	E-ratio	p-value	Total <sup>c</sup>	Site <sup>d</sup>	E-ratio	p-value
<b>HS</b>	126,147	13,730	8,468	1,604	1.74	8.05E-115	13,033	2,051	1.45	1.07E-71
<b>MM</b>	93,152	9,378	6,528	1,167	1.78	3.79E-89	9,864	1,463	1.47	2.68E-56
<b>DM</b>	17,611	904	1,469	114	1.51	4.57E-06	1,863	140	1.46	1.90E-06
<b>CE</b>	4,870	230	550	28	1.08	0.364	546	35	1.36	3.49E-02
<b>Total</b>	241,780	24,242	17,015	2,913	1.71	3.04E-192	25,306	3,689	1.45	1.05E-129

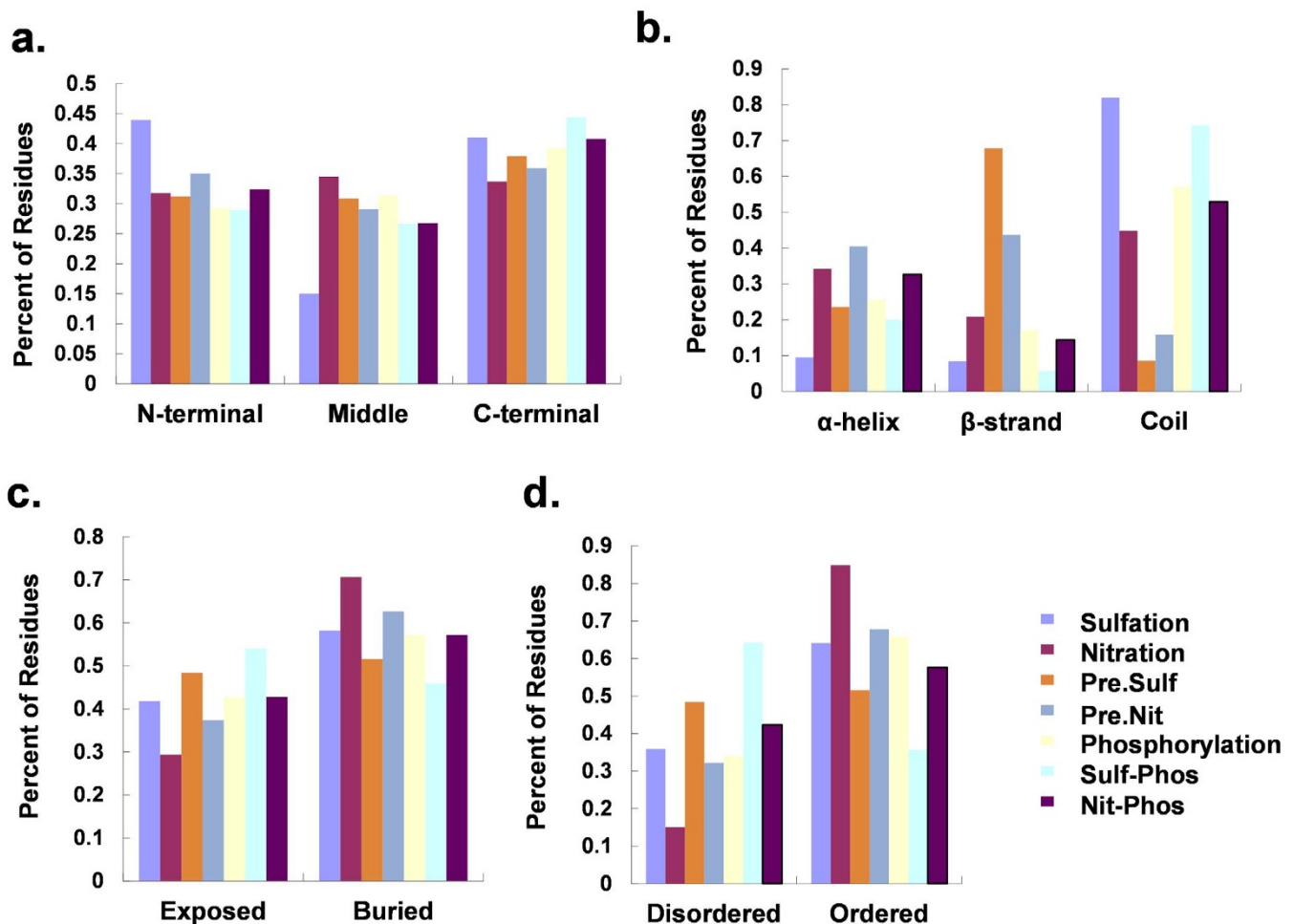
nitration sites is still quite limited, and the secondary structural preferences of two PTMs remain to be further characterized when more sites are identified. Interestingly, although both PTMs prefer not to occur at Coil, they significantly co-occupy with phosphorylation sites at Coil (Fig. 5b). Moreover, from the analysis of the protein surface accessibility, although three types of tyrosine modifications prefer to target buried tyrosines, the result suggested that sulfation-phosphorylation crosstalk preferentially co-occur at exposed tyrosines (Fig. 5c). This result is quite similar with a following analysis of protein disordered regions, in which we revealed that all three PTMs prefer to locate at protein ordered regions. Although the *in situ* crosstalk of nitration and phosphorylation also prefers to localize at ordered tyrosines, the sulfation-phosphorylation crosstalk pref-

erentially target disordered tyrosine residues (Fig. 5d). Taken together, our analyses suggested that sulfation and nitration *in situ* crosstalk with phosphorylation prefer to occur at distinct structural positions in proteins.

**The long-term evolutionary analysis revealed almost no additional natural selection of multiply modified tyrosines.** With experimentally identified PTM sites, we directly detected 2 sulfation-nitration, 3 sulfation-phosphorylation and 183 nitration-phosphorylation site-specific crosstalks (Table S1). In particular, there are ~82% (154) of total *in situ* crosstalks in *H. sapiens* (Table S1). Thus, the long-term evolutionary analysis was only performed for human tyrosine modifications, due to the data



**Figure 4** | Statistical analyses of GO and KEGG annotations for proteins with *in situ* crosstalk between sulfation or nitration and phosphorylation. The enriched GO terms for proteins with *in situ* crosstalk between sulfation (a) or nitration (b) and phosphorylation in comparison with phosphorylated proteins. The enriched KEGG annotations for proteins with *in situ* crosstalk between sulfation (c) or nitration (d) and phosphorylation in comparison with phosphorylated proteins.



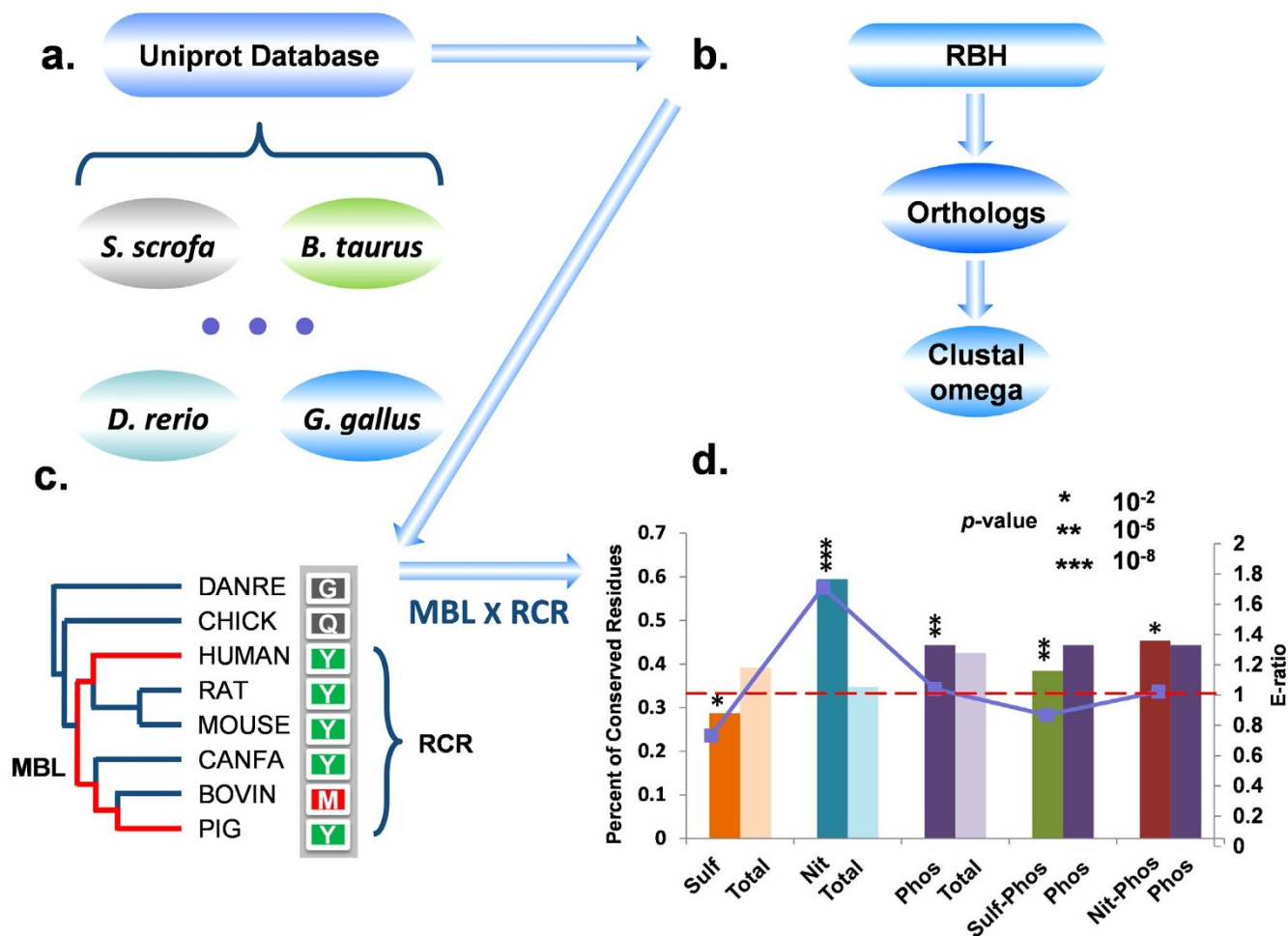
**Figure 5** | The sequence and structure preferences of known or predicted tyrosine modification sites. Sulfation, Nitration, and Phosphorylation denote known modified tyrosines. Pre. Sulf., predicted sulfation sites in phosphorylated proteins; Pre. Nit., predicted nitration sites in phosphorylated proteins; Sulf.-Phos., *in situ* crosstalk of sulfation and phosphorylation in phosphorylated proteins; Nit.-Phos., *in situ* crosstalk of nitration and phosphorylation in phosphorylated proteins. (a) Position distribution of modified tyrosines in N-terminal, Middle, or C-terminal regions in protein sequences. (b) Distribution of modified tyrosines in  $\alpha$ -helix,  $\beta$ -Strand, Coil of the secondary structure. (c) Distribution of tyrosine modification residues in exposed and buried regions. (d) Distribution of modified tyrosines in disordered and ordered regions.

limitation. Totally, we took 101 sulfotyrosines of 880 total tyrosines in 48 known sulfated proteins, 564 nitrotyrosines of 5233 tyrosines in 328 known nitrated substrates, and 13,730 phosphotyrosines of 126,147 tyrosine residues in 5876 known phosphorylated proteins from *H. sapiens*. Also, 1604 and 2051 predicted sulfation and nitration sites on 13,730 known human phosphorylation sites were considered. Because the phosphoregulation and phosphoproteome rapidly evolve<sup>40,41</sup>, here we focused on analyzing the potentially natural selection of tyrosines after the speciation of vertebrates. We obtained the proteome sets of several other vertebrates from the UniProt database (Fig. 6a). Then we computed pairwise orthologs among these species, and further multi-aligned the orthologous proteins together for each cluster of orthologous groups (COGs)<sup>42</sup> (Fig. 6b). As previously described<sup>43</sup>, we calculated  $RCS_Y$  values of modified and unmodified tyrosines, and only residues with  $RCS_Y = 1$  were regarded as conserved tyrosines (Fig. 6c).

In our results, phosphorylation sites are slightly more conserved than unmodified tyrosines ( $p$ -value  $< 10^{-5}$ ), and the ratio of modified/unmodified site (E-ratio) was only 1.04 (Fig. 6d). The result was consistent with previous studies, which demonstrated that the conservation of phosphorylated and non-phosphorylated tyrosines is quite similar<sup>40,44</sup>. However, we further observed that known nitration sites are significantly more conserved than non-nitrated ones (Fig. 6d,  $p$ -value  $< 10^{-8}$ , E-ratio=1.71), and the result was also

consistent with a previous report<sup>45</sup>. In contrast, sulfation sites are significantly less conserved than unmodified tyrosines (Fig. 6d,  $p$ -value  $< 0.01$ , E-ratio=0.73). In this regard, the conservation of distinct tyrosine modifications is quite different. Adding more or less species for the evolutionary analysis didn't influence the results, as well as changing the  $RCS_Y$  threshold from 5/8 to 1.

Furthermore, we analyzed the conservation of sulfation-phosphorylation and nitration-phosphorylation crosstalk sites against known phosphorylation sites (Fig. 6d). Unexpectedly, the tyrosines modified by both sulfation and phosphorylation are statistically less conserved than phosphorylation sites ( $p$ -value  $< 10^{-5}$ , E-ratio=0.87). Although nitration-phosphorylation crosstalk sites look like to be significantly more conserved than phosphorylation sites, the E-ratio of crosstalk/phosphorylation sites was only 1.02 (Fig. 6d,  $p$ -value  $< 0.01$ ). In this regard, the evolutionary pressure on tyrosines with nitration and phosphorylation is extremely weak against phosphorylation sites, and no signature of additional natural selection was observed for the sulfation-phosphorylation crosstalk. This might be due to tyrosine residues undergo stronger purifying selection than other types of amino acids, and both phosphorylated or non-phosphorylated tyrosines slowly evolve with a similar rate<sup>40,44</sup>. Thus, additionally evolutionary pressure is not necessary or difficult to be detected. Again, the change of either the number of species or the  $RCS_Y$  threshold didn't influence the results (Data not shown).



**Figure 6** | The computational procedure for analyzing the long-term evolution of tyrosines. (a) Proteome sets of eight vertebrates were obtained from the UniProt database<sup>55</sup>. (b) The orthologs were pairwise detected<sup>42</sup>, and the multiple sequence alignment was performed by Clustal Omega<sup>40</sup>. (c) The  $RCS_Y$  is calculated by the number of tyrosines appearing in the column of multi-alignment of orthologs where the species with the MBL containing the tyrosine residues<sup>43</sup>. (d) The conservation levels of different types of modified or unmodified tyrosines. The lines with dots represent the enrichment ratio between the two datasets. Total, total tyrosines.

**The analysis of human genetic variations revealed no functional constraint on tyrosines with multiple PTMs.** As previously described<sup>23</sup>, we mapped all experimentally identified and computationally predicted tyrosine modification sites in this study to the HGMD database<sup>46</sup>, but only detected 2 sulfation-phosphorylation and 6 nitration-phosphorylation crosstalk sites that are associated with human inherited diseases (Table S4). We also mapped these sites to the CanProVar database<sup>47</sup>, but still only obtained 3 sulfation-phosphorylation and 4 nitration-phosphorylation crosstalk sites that are potentially implicated in human cancers (Table S5). Again, we mapped the sites to human rare variations (allele frequency < 1%) (Table 4)<sup>23</sup>. The statistical results suggested that neither sulfation-phosphorylation nor nitration-phosphorylation crosstalk sites could significantly enrich more disease-associated or rare variations (Table 4). Indeed, human genetic variations or mutations will unambiguously disrupt the *in situ* crosstalks of tyrosine modifications for the specific sites, but no additionally functional constraint was observed for the crosstalks.

## Discussion

Accumulative studies have exhibited the ubiquity and importance of PTM crosstalks<sup>5–7,9</sup>. By reciprocally modifying upstream enzymes<sup>7,9</sup>, by modifying amino acids to change the modification states of adjacent sites<sup>5,6</sup>, or by competitively regulating the same residue<sup>10–12,19,30–33</sup>, various PTMs can precisely orchestrate specific

biological processes by crosstalks. In this regard, experimental identification of PTM crosstalks is fundamental for understanding the complexities and regulatory mechanisms of PTMs.

With GPS-YNO2<sup>27</sup> and a newly developed tool of GPS-TSP, we directly predicted potential nitration sites in sulfated proteins and *vice versa*. The statistical results demonstrated that *in situ* crosstalk between sulfation and nitration was significantly under-represented (Table 2). This observation might be associated with the different involvements of pathways for sulfation and nitration. However, different sequence preferences might be the basic cause of the low overlap between the two PTMs. By comparing GO terms of sulfated and nitrated proteins, we observed that most of significantly different GO terms are in sulfation (Fig. 3c). Thus, we proposed that the regulation of sulfation is much more specific than nitration. Also, the performance of GPS-YNO2 for the prediction of nitration sites is limited as 87.51% of *Ac*, 71.33% of *Sn*, and 89.84% of *Sp*<sup>27</sup>, while the performance of GPS-TSP with the default threshold is 90.23% of *Ac*, 89.60% of *Sn*, and 90.36% of *Sp* (Table 1). Thus, the sequence profile of sulfation is much more stringent than nitration.

By computationally predicting sulfation and nitration sites in known tyrosine phosphorylated substrates, we observed that up to 24.5% of known phosphorylation sites might also be modified by either sulfation or nitration, while statistical results suggested that both sulfation and nitration prefer to *in situ* crosstalk with phosphorylation at the same phosphorylation sites rather than non-phosphor-





**Table 4 |** The statistical analysis of human variations that change singly or multiply modified tyrosines. The experimentally identified or predicted tyrosine modification sites were mapped to the HGMD<sup>46</sup>, CanProVar<sup>47</sup>, and SNPs of 1000 Genomes project<sup>54</sup>. *a.* Num., the number of modified tyrosines which can be mapped to public databases; *b.* All, the number of all modified tyrosines

Resource	Phosphorylation		Sulf.-Phos.		E-ratio	<i>p</i> -value	Nit.-Phos.		E-ratio	<i>p</i> -value
	Num. <sup>a</sup>	All <sup>b</sup>	Num.	All			Num.	All		
HGMD	31	13730	2	1604	0.55	0.28	6	2051	1.30	0.31
CanProVar	40	13730	3	1604	0.64	0.30	4	2051	0.68	0.27
1000 Genome	64	13730	8	1604	1.07	0.48	12	2051	1.26	0.24

ylatable tyrosines (Table 3, Table S2 & S3). Interestingly, the overlapping rate between sulfation and nitration for the crosstalk with phosphorylation is quite low. Thus, the results suggested that sulfation and nitration preferentially crosstalk with phosphorylation in distinct biological processes and functions (Fig. 4). Notably, our analysis of GO terms demonstrated that sulfation and phosphorylation moderately prefer to co-occur at the same residues of proteins located in cytoplasm (GO:0005737) (Fig. 4a, *p*-value < 10<sup>-5</sup>, E-ratio = 1.19). The result seems conflicting with known evidence that sulfation exclusively occurs in secreted pathways because two tyrosylprotein sulfotransferases, TPST-1 and TPST-2, locate in the trans-Golgi network, and almost all of sulfated proteins were identified in extracellular or membrane-related regions<sup>24–26</sup>. Actually, from the 42 human sulfated substrates manually curated in this study, it was observed that only 5 proteins localized exclusively in extracellular positions, while 20 proteins have no extracellular localizations. Our hypothesis was that the proteins with sulfation-phosphorylation crosstalks might also locate at other subcellular localizations beyond cytoplasm. Indeed, after we removed proteins with additional GO terms on golgi, extracellular or membrane in both phosphorylated and crosstalk substrates, the statistical significance didn't exist any longer (*p*-value > 0.05).

Undoubtedly, our results will be helpful for further experimental consideration. For example, the molecular masses of a nitro group (·NO<sub>2</sub>) is ~46 Da, while the phosphate and sulfate groups are 79.9663 and 79.9568 Da, respectively<sup>48</sup>. In this regard, unambiguously distinguishing tyrosine phosphorylation and sulfation is particularly difficult and error-prone by currently experimental approaches. Indeed, the human trypsin 1 (UniProt ID: P07477) was first identified to be phosphorylated at Y151<sup>49</sup>, whereas a later experiment verified that the site is sulfated<sup>50</sup>. Although it's not possible that nearly 2900 phosphorylation sites are actually sulfated (Table 3), great attention should be paid to experimentally identifying either phosphorylation or sulfation sites, given the preference of *in situ* crosstalk between sulfation and phosphorylation. Since currently systematical profiling of sulfation sites is yet to be carried out, the predictions based on the limited small-scale experimental data could serve as a start for the studies of crosstalks between sulfation and phosphorylation. In addition, because the large-scale data set of tissue-specific sulfation or nitrations sites is still not available, whether the preference of *in situ* crosstalks of tyrosine modifications will be changed in different tissues or organs remains to be dissected. However, our current results still clearly demonstrated that different types of tyrosine modifications can have distinct preferences for the *in situ* crosstalk.

In his classical essay, Theodosius Dobzhansky stated “nothing in biology makes sense, except in the light of evolution”<sup>51</sup>. Thus, the analysis of PTMs should be performed under the framework of evolutionary biology. Indeed, a number of evolutionary studies have characterized additional natural selection on the phosphorylation and phosphoproteome against unmodified residues, although only a weak evolutionary pressure was observed<sup>40,41,44</sup>. Recently, more controversial viewpoints have emerged toward the residues with multiple PTMs, and whether the *in situ* crosstalk of PTMs with an

additionally evolutionary pressure still remains to be elusive<sup>19–23</sup>. In contrast with previous studies that the evolutionary analyses were performed at the genome- or proteome-wide level<sup>40,41,44</sup>, Gray *et al.* defined a protein-specific measure,  $\alpha$ , to estimate the long-term natural selection on individual proteins<sup>23,52</sup>. Although such a new approach was used, the analysis demonstrated that lysines with multiple PTMs are slightly more conserved than singly modified lysines, with only a 6% difference<sup>23</sup>. Also, our long-term evolutionary analysis of the *in situ* crosstalks among tyrosine modifications revealed that only nitration-phosphorylation but not sulfation-phosphorylation crosstalk sites are slightly more conserved than phosphorylation sites, with a very small E-ratio of 1.02 (Fig. 6d). In this regard, although the statistical significance exists in the previous<sup>23</sup> and our results, only very weak to no evolutionary pressure was detected for multiply modified sites from the long-term analyses.

For the short-term evolutionary analysis, Gray *et al.* mapped 37,720 singly and 3961 multiply modified lysines to disease-associated mutations in the HGMD database<sup>46</sup>, and only 104 (0.28%) and 17 (0.43%) variations were mapped to lysines with single and multiple PLMs, respectively<sup>23</sup>. Also, we found 2 (0.12%) sulfation- and 6 (0.29%) nitration-phosphorylation crosstalk sites occur at HGMD mutations (Table 4). Thus, only an extremely small proportion of singly or multiple PTM sites are associated with human inherited diseases, whatever the statistical value is significant or not. Moreover, Gray *et al.* hypothesized that if additional negative selection exists in multiply modified lysines, they would prefer to occur at variations with lower allele frequencies. Indeed, the fractions of rare variations (allele frequency < 1%) for unmodified, singly, doubly and triply modified lysines are 88.8%, 91.1%, 93.8% and 100%, respectively<sup>23</sup>. For tyrosine modifications, we only detected one common variation (allele frequency > 1%) on phosphorylation sites, and didn't find any other common variations on crosstalk sites. Thus, it seems that residues with multiple PTMs prefer to occur at rare variations and undergo additional evolutionary pressure. To draw this conclusion, a prerequisite is that multiply modified with multiple PTMs can really enrich the human variations or mutations. However, our results on variations in the HGMD<sup>46</sup>, GWASdb<sup>53</sup> and 1000 Genomes project<sup>54</sup> demonstrated that human variations were not over-represented in multiply modified tyrosines (Table 4). Taken together, no additional evolutionary pressure was detected on sites with multiple PTMs from the short-term analysis.

More interestingly, although no evolutionary pressure was detected for the *in situ* crosstalk of O-GlcNAcylation and phosphorylation because of nearly random co-occurrence<sup>20</sup>, we didn't detect any additional natural selection on significantly co-occupied sulfation- and nitration-phosphorylation crosstalk sites. Thus, whether different PTMs prefer to *in situ* crosstalk or not, little or no additional evolutionary pressure exists on residues with multiple PTMs against singly modified sites.

## Methods

**The data sets of tyrosine modifications.** By searching the PubMed with multiple keywords such as “tyrosine sulfation”, “sulfation” and “sulfated”, we collected 273 experimentally identified tyrosine sulfation sites in 171 proteins, and integrated the dataset into the DOSS database (<http://tsp.biocuckoo.org/database.php>). The protein



sequences were retrieved from the UniProt database<sup>55</sup>. As previously described<sup>27</sup>, we regarded the known sulfation sites as positive data (+), while other non-sulfated tyrosines were taken as negative data (-). Because the redundancy of homologous sites in the positive data (+) leads to overestimated prediction, we used CD-HIT (<http://weizhong-lab.ucsd.edu/cd-hit/>) to cluster the protein sequences<sup>56</sup>, followed by re-alignment with BLAST packages and manual check of proteins with  $\geq 40\%$  identity<sup>57</sup>. The redundant sulfation sites at the same position in the homologous proteins according to the alignment results were cleared. Finally, a non-redundant dataset for training was constructed with 202 positive sites and 1,027 negative sites from 116 substrates. The manually collected dataset and the non-redundant training dataset were provided for download at <http://tsp.biocuckoo.org/down.php>.

For the analysis of the *in situ* crosstalk among sulfation, nitration and phosphorylation, we took 1050 known nitration sites of 539 proteins from a previously published study<sup>27</sup>. The experimentally identified phosphorylation sites were taken from several major databases, including PhosphoPep v2.0<sup>58</sup>, Phospho.ELM 8.3 (released in April 2010)<sup>58</sup>, SysPTM 1.1<sup>59</sup>, PhosphoSitePlus<sup>60</sup>, and HPRD 9.0<sup>61</sup>. We obtained 5876, 4302, 657 and 199 tyrosine phosphoproteins from *H. sapiens*, *M. musculus*, *D. melanogaster* and *C. elegans*. The phosphorylation information in *S. cerevisiae* was not used, because experimental efforts suggested there is not sulfation in yeast<sup>3,62</sup>. Totally, 24,242 tyrosine phosphorylation sites were collected from 11,034 proteins, which were also available for download at <http://tsp.biocuckoo.org/down.php>.

**Performance evaluation.** To evaluate the performance of GPS-TSP for the prediction of tyrosine sulfation sites, four measurements of accuracy (*Ac*), sensitivity (*Sn*), specificity (*Sp*), and Mathew Correlation Coefficient (*MCC*) were adopted and defined as below<sup>27,63</sup>:

$$Ac = \frac{TP + TN}{TP + FP + TN + FN}, Sn = \frac{TP}{TP + FN}, Sp = \frac{TN}{TN + FP},$$

and

$$MCC = \frac{(TP \times TN) \times (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}}.$$

The leave-one-out (LOO) validation and 4-, 6-, 8- and 10-fold cross-validations were carried out. The Receiver Operating Characteristic (ROC) curves and AROC (area under ROC) values were also calculated.

**The sequence and structure analysis.** To assess the position distribution of modified tyrosine residues in protein sequences, we equally separated a protein into three fragments, namely, N-terminal, Middle and C-terminal, and directly counted the number of residues in each part. The secondary structures were predicted NetSurfP server ver. 1.1 (<http://www.cbs.dtu.dk/services/NetSurfP/>), which calculates a probability score for each of three major types of secondary structures including  $\alpha$ -Helix,  $\beta$ -Strand and Coil, respectively<sup>64</sup>. The secondary structure of a modified tyrosine was assigned into the type with the highest probability score. Also, NetSurfP predicts both Relative Surface Accessibility (RSA) and Absolute Surface Accessibility (ASA), and tyrosines were classified as buried or exposed based on a threshold of 25% exposure<sup>64</sup>. Moreover, the disordered regions in proteins were directly predicted by RONN (<https://www.strubi.ox.ac.uk/RONN/>)<sup>65</sup>.

**The long-term evolutionary analysis.** To analyze the conservation of modified and unmodified tyrosines in *H. sapiens*, we further downloaded 41,118, 17,864, 28,853, 51,192, 26,457, 24,214, and 26,138 proteins for *D. rerio* (DANRE), *G. gallus* (CHICK), *R. norvegicus* (RAT), *M. Musculus* (MOUSE), *C. familiaris* (CANFA), *B. Taurus* (BOVIN), and *S. scrofa* (PIG) from the UniProt database (November, 2013)<sup>55</sup>, respectively. As previously described<sup>13</sup>, the approach of reciprocal best hits (RBHs) was used by pairwise detecting orthologs among the eight species, and the blastall program in the BLAST package was chosen. All orthologous proteins among different species were multi-aligned by Clustal Omega (<http://www.clustal.org/omega/>)<sup>40</sup>. In a multiple sequence alignment (MSA), if a human tyrosine is sulfated, nitrated and/or phosphorylated, the corresponding column was regarded as the modified position. Other columns containing human tyrosines were taken as the unmodified positions. To calculate the conservation of a human tyrosine residue from a MSA, we adopted a previously reported measurement of Residue Conservation Score (RCS), which is expressed as Residue Conservation Ratio (RCR)<sup>43</sup>. Given a tyrosine column, the *RCS<sub>Y</sub>* can be calculated as below:

$$RCS_Y = \frac{N_Y}{N}$$

The *N* is the number of sequences with the maximum branch length (MBL), which is the maximum branch distance between any two species that contain a conserved tyrosine. The *N<sub>Y</sub>* is the number of tyrosines appears in the column. The tyrosine residues with *RCS<sub>Y</sub>* = 1 were regarded as conserved tyrosines. The change of the *RCS<sub>Y</sub>* threshold (0.625 ~ 1) didn't influence the final results. The species tree of the eight organisms were taken from the Interactive Tree Of Life (iTOL, <http://itol.embl.de/>)<sup>40</sup>.

**The data sets of human genetic mutations and variations.** We downloaded 153,839 human disease-associated single nucleotide polymorphisms (SNPs) from the

GWASdb<sup>53</sup>. Also, the human inherited disease-associated mutations with annotated information of the Human Gene Mutation Database (HGMD)<sup>46</sup> were obtained from the ClinVar dataset (March 20<sup>th</sup>, 2012) in the NCBI ftp server, including 78,827 records with 15,641 non-redundant missense mutations. Moreover, we downloaded 73,270 cancer-associated variants from the CanProVar database (September 20<sup>th</sup>, 2012)<sup>47</sup>. In addition, the SNPs of the 1000 Genomes project were downloaded (June 1<sup>st</sup> 2013, phase I release v3)<sup>54</sup>. Totally, we obtained 325,159 coding SNPs (cSNPs) with reference allele frequencies from four ancestry-based super population groups (AMR, Ad Mixed American; ASN, East Asian; AFR, African; EUR, European).

- Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnology* **21**, 255–261 (2003).
- Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342–7372 (2005).
- Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnology* **24**, 1241–1252 (2006).
- Hietakangas, V. *et al.* Phosphorylation of serine 303 is a prerequisite for the stress-inducible SUMO modification of heat shock factor 1. *Molecular and Cellular Biology* **23**, 2953–2968 (2003).
- Hietakangas, V. *et al.* PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A* **103**, 45–50 (2006).
- Lu, Z., Cheng, Z., Zhao, Y. & Volchenboum, S. L. Bioinformatic analysis and post-translational modification crosstalk prediction of lysine acetylation. *Plos One* **6**, e28228 (2011).
- Peters, J. M. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Reviews. Molecular Cell Biology* **7**, 644–656 (2006).
- Gao, D. *et al.* Rictor forms a complex with Cullin-1 to promote SGK1 ubiquitination and destruction. *Molecular Cell* **39**, 797–808 (2010).
- Lopez-Otin, C. & Hunter, T. The regulatory crosstalk between kinases and proteases in cancer. *Nature Reviews Cancer* **10**, 278–292 (2010).
- Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annual Review of Biochemistry* **80**, 825–858 (2011).
- Kouzarides, T. SnapShot: Histone-modifying enzymes. *Cell* **128**, 802 (2007).
- Yang, X. J. & Seto, E. Lysine acetylation: codified crosstalk with other posttranslational modifications. *Molecular Cell* **31**, 449–461 (2008).
- Kaasik, K. *et al.* Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. *Cell Metabolism* **17**, 291–302 (2013).
- Voss, O. H., Kim, S., Wewers, M. D. & Doseff, A. I. Regulation of monocyte apoptosis by the protein kinase Cdelta-dependent phosphorylation of caspase-3. *The Journal of Biological Chemistry* **280**, 17371–17379 (2005).
- Endo, K. *et al.* Proteolytic cleavage and activation of protein kinase C [micro] by caspase-3 in the apoptotic response of cells to 1-beta-D-arabinofuranosylcytosine and other genotoxic agents. *The Journal of Biological Chemistry* **275**, 18476–18481 (2000).
- Yoshida, K., Liu, H. & Miki, Y. Protein kinase C delta regulates Ser46 phosphorylation of p53 tumor suppressor in the apoptotic response to DNA damage. *The Journal of Biological Chemistry* **281**, 5734–5740 (2006).
- Hofmann, T. G. *et al.* Regulation of p53 activity by its interaction with homeodomain-interacting protein kinase-2. *Nature Cell Biology* **4**, 1–10 (2002).
- Shi, X. *et al.* Modulation of p53 function by SET8-mediated methylation at lysine 382. *Molecular Cell* **27**, 636–646 (2007).
- Wang, Z. *et al.* Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* **3**, ra2 (2010).
- Trinidad, J. C. *et al.* Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Molecular & Cellular Proteomics* : MCP **11**, 215–229 (2012).
- Park, J. *et al.* SIRT5-mediated lysine desuccinylation impacts diverse metabolic pathways. *Molecular Cell* **50**, 919–930 (2013).
- Weinert, B. T. *et al.* Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. *Cell Reports* **4**, 842–851 (2013).
- Gray, V. E., Liu, L., Nirankari, R., Hornbeck, P. V. & Kumar, S. Signatures of natural selection on mutations of residues with multiple posttranslational modifications. *Molecular Biology and Evolution* (2014).
- Kehoe, J. W. & Bertozzi, C. R. Tyrosine sulfation: a modulator of extracellular protein-protein interactions. *Chemistry & biology* **7**, R57–61 (2000).
- Stone, M. J., Chuang, S., Hou, X., Shoham, M. & Zhu, J. Z. Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *New Biotechnology* **25**, 299–317 (2009).
- Monigatti, F., Hekking, B. & Steen, H. Protein sulfation analysis--A primer. *Biochim Biophys Acta* **1764**, 1904–1913 (2006).
- Liu, Z. *et al.* GPS-YNO2: computational prediction of tyrosine nitration sites in proteins. *Molecular Biosystems* **7**, 1197–1204 (2011).
- Ghesquiere, B. *et al.* In vitro and in vivo protein-bound tyrosine nitration characterized by diagonal chromatography. *Molecular & Cellular Proteomics* : MCP **8**, 2642–2652 (2009).
- Lim, W. A. & Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661–667 (2010).



30. Takahashi, T., Yamashita, H., Nakamura, T., Nagano, Y. & Nakamura, S. Tyrosine 125 of alpha-synuclein plays a critical role for dimerization following nitrate stress. *Brain Research* **938**, 73–80 (2002).
31. Negro, A., Brunati, A. M., Donella-Deana, A., Massimino, M. L. & Pinna, L. A. Multiple phosphorylation of alpha-synuclein by protein tyrosine kinase Syk prevents eosin-induced aggregation. *FASEB Journal* **16**, 210–212 (2002).
32. Songyang, Z. & Cantley, L. C. Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends in Biochemical Sciences* **20**, 470–475 (1995).
33. Rehfeld, J. F., Hansen, C. P. & Johnsen, A. H. Post-poly(Glu) cleavage and degradation modified by O-sulfated tyrosine: a novel post-translational processing mechanism. *The EMBO Journal* **14**, 389–396 (1995).
34. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Research* **14**, 1188–1190 (2004).
35. Monigatti, F., Gasteiger, E., Bairoch, A. & Jung, E. The Sulfinator: predicting tyrosine sulfation sites in protein sequences. *Bioinformatics* **18**, 769–770 (2002).
36. Chang, W. C. *et al.* Incorporating support vector machine for identifying protein tyrosine sulfation sites. *Journal of Computational Chemistry* **30**, 2526–2537 (2009).
37. Barrell, D. *et al.* The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Research* **37**, D396–403 (2009).
38. Diella, F., Gould, C. M., Chica, C., Via, A. & Gibson, T. J. Phospho.ELM: a database of phosphorylation sites--update 2008. *Nucleic Acids Research* **36**, D240–244 (2008).
39. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32**, D277–280 (2004).
40. Gnad, F. *et al.* PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology* **8**, R250 (2007).
41. Landry, C. R., Levy, E. D. & Michnick, S. W. Weak functional constraints on phosphoproteomes. *Trends Genet* **25**, 193–197 (2009).
42. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
43. Minguéz, P. *et al.* Deciphering a global network of functionally associated post-translational modifications. *Molecular systems biology* **8**, 599 (2012).
44. Chen, S. C., Chen, F. C. & Li, W. H. Phosphorylated and nonphosphorylated serine and threonine residues evolve at different rates in mammals. *Molecular Biology and Evolution* **27**, 2548–2554 (2010).
45. Ng, J. Y., Boelen, L. & Wong, J. W. Bioinformatics analysis reveals biophysical and evolutionary insights into the 3-nitrotyrosine post-translational modification in the human proteome. *Open Biology* **3**, 120148 (2013).
46. Stenson, P. D. *et al.* The Human Gene Mutation Database: 2008 update. *Genome Medicine* **1**, 13 (2009).
47. Li, J., Duncan, D. T. & Zhang, B. CanProVar: a human cancer proteome variation database. *Human Mutation* **31**, 219–228 (2010).
48. Gharib, M. *et al.* Artifactual sulfation of silver-stained proteins: implications for the assignment of phosphorylation and sulfation sites. *Molecular & Cellular Proteomics : MCP* **8**, 506–518 (2009).
49. Gaboriaud, C., Serre, L., Guy-Crotte, O., Forest, E. & Fontecilla-Camps, J. C. Crystal structure of human trypsin 1: unexpected phosphorylation of Tyr151. *Journal of Molecular Biology* **259**, 995–1010 (1996).
50. Sahin-Toth, M., Kukor, Z. & Nemoda, Z. Human cationic trypsinogen is sulfated on Tyr154. *The FEBS Journal* **273**, 5044–5050 (2006).
51. Dobzhansky, T. Nothing in biology makes sense, except in the light of evolution. *Amer Biol Teacher* **35**, 125–129 (1973).
52. Gray, V. E. & Kumar, S. Rampant purifying selection conserves positions with posttranslational modifications in human proteins. *Molecular Biology and Evolution* **28**, 1565–1568 (2011).
53. Li, M. J. *et al.* GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Research* **40**, D1047–1054 (2012).
54. Amoutzias, G. D. *et al.* Posttranslational regulation impacts the fate of duplicated genes. *Proc Natl Acad Sci U S A* **107**, 2967–2971 (2010).
55. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research* **38**, D142–148 (2010).
56. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
57. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Research* **36**, W5–9 (2008).
58. Bodenmiller, B. *et al.* PhosphoPeP--a database of protein phosphorylation sites in model organisms. *Nature Biotechnology* **26**, 1339–1340 (2008).
59. Li, H. *et al.* SysPTM - a systematic resource for proteomic research of post-translational modifications. *Molecular & Cellular Proteomics : MCP* (2009).
60. Hornbeck, P. V., Chabra, L., Kornhauser, J. M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 (2004).
61. Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic Acids Research* **37**, D767–772 (2009).
62. Moore, K. L. The biology and enzymology of protein tyrosine O-sulfation. *The Journal of Biological Chemistry* **278**, 24243–24246 (2003).
63. Liu, Z. *et al.* GPS-PUP: computational prediction of pupylation sites in prokaryotic proteins. *Molecular Biosystems* **7**, 2737–2740 (2011).
64. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M. & Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology* **9**, 51 (2009).
65. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369–3376 (2005).

## Acknowledgments

This work was supported by grants from the National Basic Research Program (973 project) (2012CB910101, and 2013CB9339002), Natural Science Foundation of China (31171263, 81272578 and J1103514), International Science & Technology Cooperation Program of China (2014DFB30020), China Postdoctoral Science Foundation (2014M550392), and Fundamental Research Funds for the Central Universities (HUST: 2013TS080, 2014YQ003).

## Author contributions

Y.X. designed and supervised experiments. Z.P. and Z.L. performed experiments and data analysis. H.C., Y.W., T.G., S.U. and J.R. contributed to data analysis. Y.X., Z.P. and Z.L. wrote the manuscript with contributions of all authors.

## Additional information

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

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

**How to cite this article:** Pan, Z. *et al.* Systematic analysis of the *in situ* crosstalk of tyrosine modifications reveals no additional natural selection on multiply modified residues. *Sci. Rep.* **4**, 7331; DOI:10.1038/srep07331 (2014).



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 in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>