Systematic analysis of the \textit{in situ} crosstalk of tyrosine modifications reveals no additional natural selection on multiply modified residues

Zhicheng Pan\textsuperscript{1*}, Zexian Liu\textsuperscript{1*}, Han Cheng\textsuperscript{1}, Yongbo Wang\textsuperscript{1}, Tianshun Gao\textsuperscript{1}, Shahid Ullah\textsuperscript{1}, Jian Ren\textsuperscript{2} & Yu Xue\textsuperscript{1}

\textsuperscript{1}Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China, \textsuperscript{2}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 \textit{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 \textit{in situ} crosstalks. From the results, we observed that the \textit{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 \textit{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\textsuperscript{1-3}. In particular, one PTM can crosstalk with other PTMs to synergistically orchestrate specific processes through three distinct mechanisms\textsuperscript{4-13}. First, different PTMs can co-occur in the same proteins and crosstalk in a \textit{cis}-regulatory manner\textsuperscript{4,5}. For example, the phosphorylation of the motif $\Psi$KXXEXXPSP ($\Psi$ 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 (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\textsuperscript{4}. Second, one PTM can regulate another PTM by modifying its cognate enzymes and \textit{vice versa}, in a \textit{trans}-regulatory mode\textsuperscript{7-9}. For example, the E3 ubiquitin ligase complex of Rictor/Cullin-1/Bbx1 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 (Fig. 1b). Third, multiple PTMs can “\textit{in situ}” interplay with each other by competitively modifying same residues\textsuperscript{10-13}. For example, a circadian clock protein PER2 is competitively O-GlcNAcylated and phosphorylated at S662, while the \textit{in situ} crosstalk precisely regulates the PER2 repressor activity\textsuperscript{13} (Fig. 1c). In addition, different types of PTM crosstalks can simultaneously occur and regulate biological functions in a complicated manner. For example, PKC\textsubscript{δ} phosphorylates Caspase-3\textsuperscript{14}, which reciprocally cleaves PKC-δ as a \textit{trans}-crosstalk\textsuperscript{14} (Fig. 1d). Also, p53 can be activated through the PKC\textsubscript{δ}-mediated phosphorylation of S46\textsuperscript{16}, and such a phosphorylation promotes its...
acetylation at K382 through a cis-crosstalk17 (Fig. 1d). Again, lysine acetylation and methylation compete at K382 for modulating the p53 transcription activity18 (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 regions19. 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 phosphorylated20. With a computational simulation, they observed that the overlap of two PTMs is not significant and nearly equal to be random20. 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)21. 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 Escherichia coli, 56% of Saccharomyces cerevisiae, and 57% of mouse liver succinylation sites are overlapped with acetylation22. Interestingly, they also detected that only 27% of succinylation sites are acetylated in human HeLa cells22. 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 PTMs20. 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 variations23. Thus, Gray et al. concluded that mutations of lysines with multi-PTMs undergo much greater purifying selection against singly modified lysine residues23.

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 sulfation24–26, nitration27,28 and tyrosine phosphorylation29 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-

---

**Figure 1 | The examples for different types of PTM crosstalks.** (a) Cis-crosstalk, phosphorylation at S303 of HSF1 promotes its K298 sumoylation4. (b) Trans-crosstalk between ubiquitination of SGK1 by Rictor/Cullin-1/Rbx1 and phosphorylation of Rictor by SGK118. (c) In situ crosstalk, protein PER2 is competitively O-GlcNAcylated and phosphorylated at S66221. (d) A complex crosstalk among PKCδ, Caspase-3, and p53 that different types of PTM crosstalks can simultaneously occur and regulate biological functions.
ylation and nitration can co-occur at Y125, Y133 and Y136 of α-synuclein (UniProt ID: P37840)\textsuperscript{30,31}. Moreover, human Gastrin (P01350) was identified to be phosphorylated by ν-Src at Y87\textsuperscript{22}, the site which is also modulated by sulfation\textsuperscript{32}.

From the scientific literature, public databases and our previous studies\textsuperscript{32}, 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\textsuperscript{27}, 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 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\textsuperscript{34} (Fig. 2a). Although a previous analysis proposed that the +2 position didn’t contain any information with a limited data set\textsuperscript{26}, 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 found 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)\textsuperscript{27} (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\textsuperscript{33} and SulfoSite\textsuperscript{36} 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\textsuperscript{27,28}, whereas sulfation was proposed to mostly occur in the secretory pathway\textsuperscript{24–26}. 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\textsuperscript{27}. In order to analyze and compare the functional abundance and diversity of sulfation and PTN, we downloaded the gene ontology (GO) (March 31\textsuperscript{st}, 2012) association files from the GOA database at the EBI (http://www.ebi.ac.uk/goa/). There were 44,741 human proteins, 65 sulfated proteins and 326 nitrated proteins annotated with at least one GO term.

With the hypergeometric distribution\textsuperscript{26}, we statistically analyzed the enriched biological processes, molecular functions and cellular components with GO annotations for sulfated (Fig. 3a, p-value < 10\textsuperscript{-15}) and nitrated (Fig. 3b, p-value < 10\textsuperscript{-10}) 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 previous experimental observations\textsuperscript{26–28} (Table 3). 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 (χ\textsuperscript{2}) test\textsuperscript{22} (Fig. 3c, p-value < 10\textsuperscript{-8}). 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\textsuperscript{27}, 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).
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). 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 < 10⁻⁵) or nitration (Fig. 4b, p-value < 10⁻⁵) 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

| 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 |
|           |           | 91.54% | 83.17% | 93.18% | 0.7161 |
|           |           | 90.48% | 87.12% | 91.13% | 0.7034 |
| Sulfonator|           | 87.63% | 61.79% | 93.02% | 0.5588 |
| SulfoSite |           | 88.45% | 69.73% | 91.09% | 0.5403 |

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.
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) pathways39, 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 nitration sites slightly prefer to locate at C-terminal, as well as other types of tyrosine modifications and their crosstalks (Fig. 5a).

| PTM       | Known| Predicted |
|-----------|------|-----------|
| Sulfation | 171  | 1,518     |
| Nitration | 539  | 8,356     |

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.

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 β-Strand (Fig. 5b). This might be due to the number of known sulfation and
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...
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 \( H. \) 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\(^{40,41}\), 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)\(^{42}\) (Fig. 6b). As previously described\(^{43}\), we calculated RCS\(_Y\) values of modified and unmodified tyrosines, and only residues with \( RCS_Y \geq 1\) 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\(^{44,45}\). 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\(^{46}\). 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\(^{44,45}\). 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 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.
The analysis of human genetic variations revealed no functional constraint on tyrosines with multiple PTMs. As previously described, we mapped all experimentally identified and computationally predicted tyrosine modification sites in this study to the HGMD database, 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, 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). 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. By reciprocally modifying upstream enzymes, by modifying amino acids to change the modification states of adjacent sites, or by competitively regulating the same residue, 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-YNO 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-YNO for the prediction of nitration sites is limited as 87.51% of Ac, 71.33% of Sn, and 89.84% of Sp, 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-
yalatable 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 overlapping rate between sulfation and nitration for the crosstalk with phosphorylation because of nearly random co-occurrence20, we didn’t detect additional evolutionary pressure still remains to be elusive19-23. In contrast with previous studies that the evolutionary analyses were performed at the genome- or proteome-wide level40,41,44, Gray et al. defined a protein-specific measure, ϵ, to estimate the long-term natural selection on individual proteins23,52. 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% difference23. 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 previous23 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 database46, and only 104 (0.28%) and 17 (0.43%) variations were mapped to lysines with single and multiple PLMs, respectively23. 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%, respectively23. 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 HGMD46, GWASdb53 and 1000 Genomes project51 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-occurrence23, 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 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

| Resource          | Phosphorylation | Sulf.-Phos. | Nit.-Phos. |
|-------------------|-----------------|-------------|------------|
|                   | Num. | All | Num. | All | E-ratio | p-value | Num. | All | E-ratio | p-value |
| HGMG             | 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 |

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 46, CanProVar 47, and SNPs of 1000 Genomes project 51. a. Num., the number of modified tyrosines which can be mapped to public databases; b. All, the number of all modified tyrosines.
sequences were retrieved from the UniProt database\(^2\). As previously described\(^2\), 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\(^5\) to reduce the redundancy of homologous sites. 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/download.php.

For the analysis of the in situ cotranslational sulfation, nitration and phosphorylation, we took 1050 known nitration sites of 339 proteins from a previously published study\(^3\). The experimentally identified phosphorylation sites were taken from several major databases, including PhosphoPep v2.058, Phospho.ELM 8.3 (released in April 2010)\(^3\), SysPTM v1.1.\(^3\), PhosphoSitePlus\(^8\), and HIPDP 9.0\(^9\). We obtained 5876, 4302, 657 and 199 tyrosine phosphorylations 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 no sulfation in yeast\(^10\). Totally, 24,242 tyrosine phosphorylation sites were collected from 11,034 proteins, which were also available for download at http://tsp.biocuckoo.org/download.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:\(^10\):

\[
Ac = \frac{TP + TN}{TP + FP + TN + FN}, \quad Sn = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}, \quad MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TN + FN) \times (TP + FN) \times (TN + FP)}}
\]

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 AUC (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 v.1.1 (http://www.cbs.dtu.dk/services/NetSurfP/), which calculates a specificity (Sp) for each of the three major types of secondary structures including α-Helix, β-Strand and Coil, respectively\(^11\). 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%.

The data sets of human genetic mutations and variations were downloaded 153,839 human disease-associated single nucleotide polymorphisms (SNPs) from the ClinVar database (March 20\(^{th}\) 2012) in the NCBI ftp server, including 75,827 records with 15,641 non-redundant missense mutations. Moreover, we downloaded 73,270 cancer-associated variants from the CanProVar database (September 20\(^{th}\), 2012)\(^12\). In addition, the SNPs of the 1000 Genomes project were downloaded (June 1\(^{st}\) 2010)\(^13\), 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).

1. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnology* 21, 255–261 (2003).
2. Walsh, C. T., Garneau-Tsokidova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversification. *Angew Chem Int Ed Engl* 44, 7342–7372 (2005).
3. Walsh, G. & Jeffers, R. Post-translational modifications in the context of eukaryotic proteins. *Nature Biotechnology* 24, 1241–1252 (2006).
4. 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).
5. Hietakangas, V. et al. Phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A* 103, 45–50 (2006).
6. 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).
7. Peters, J. M. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Molecular Cell* 7, 644–656 (2001).
8. Gao, D. et al. Rictor forms a complex with Cullin-1 to promote SGK1 ubiquitination and destruction. *Molecular Cell* 39, 797–808 (2010).
9. Lopez-Ortin, C. & Hunter, T. The regulatory crosstalk between kinases and proteases in cancer. *Nature Reviews Cancer* 10, 278–292 (2010).
10. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annual Review of Biochemistry* 80, 825–858 (2011).
11. Kourzarides, T. SnapShot: Histone-modifying enzymes. *Cell* 128, 802 (2007).
12. Yang, X. J. & Seto, E. Lysine acetylation: codified crosstalk with other posttranslational modifications. *Molecular Cell* 31, 449–461 (2008).
13. Kaakki, K. et al. GlcNAcose-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. *Cell Metabolism* 17, 291–302 (2013).
14. Voss, O. H., Kim, S., Wewers, M. D. & Dosef, A. J. Regulation of monocytie apoptosis by the protein kinase Cdelta-dependent phosphorylation of caspase-3. *The Journal of Biological Chemistry* 280, 17371–17379 (2005).
15. 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).
16. Yoshida, K., Liu, H. & Mikol, 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).
17. 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).
18. Shi, X. et al. Modulation of p53 function by SET8-mediated methylation at lysine 382. *Molecular Cell* 27, 636–646 (2007).
19. Wang, Z. et al. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates nutrient cytokinesis. *Sci Signal* 3, ra2 (2010).
20. Trinidad, J. C. et al. Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synaphe. *Molecular & Cellular Proteomics : MCP* 11, 215–229 (2012).
21. Park, J. et al. SIRT5-mediated lysine desuccinylation impacts diverse metabolic pathways. *Molecular Cell* 59, 419–430 (2015).
22. Weintert, B. T. et al. Lysine succinylation is a frequently occurring modification in proarykotes and eukaryotes and extensively overlaps with acetylation. *Cell Reports* 4, 842–851 (2013).
23. 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).
24. Keoho, J. W. & Bertozzi, C. R. Tyrosine sulfation: a modulator of extracellular protein-protein interactions. *Chemistry & biology* 7, R57–61 (2000).
25. 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).
26. Monigatti, F., Hekking, B. & Steen, H. Protein sulfation analysis—A primer. *Biochim Biophys Acta 1764*, 1904–1913 (2006).
27. Liu, Z. et al. GPS-YNO2: computational prediction of tyrosine nitration sites in proteins. *Molecular Biosystems* 7, 1197–1204 (2011).
28. Ghesquire, B. et al. In vivo protein-bound tyrosine nitration characterized by diaxonal chromatography. *Molecular & Cellular Proteomics* : MCP 8, 2642–2652 (2009).
29. Lim, W. A. & Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* 142, 661–667 (2010).
