Table S1. The predictive performance of balanced and unbalanced dataset

<table>
<thead>
<tr>
<th>Positive dataset : Negative dataset</th>
<th>Sn (%)</th>
<th>Sp (%)</th>
<th>Acc (%)</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1</td>
<td>80.83</td>
<td>78.17</td>
<td>79.50</td>
<td>0.5902</td>
</tr>
<tr>
<td>1:2</td>
<td>60.47</td>
<td>88.94</td>
<td>79.45</td>
<td>0.5215</td>
</tr>
<tr>
<td>1:3</td>
<td>47.79</td>
<td>93.81</td>
<td>82.30</td>
<td>0.4841</td>
</tr>
<tr>
<td>1:4</td>
<td>40.71</td>
<td>94.69</td>
<td>83.89</td>
<td>0.4298</td>
</tr>
<tr>
<td>1:5</td>
<td>34.22</td>
<td>96.70</td>
<td>86.28</td>
<td>0.4141</td>
</tr>
</tbody>
</table>
Table S2. Euclidean distances between the matrix of complete dataset and 60 subsets

<table>
<thead>
<tr>
<th>Datasets</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distances</td>
<td>0.0377</td>
<td>0.0319</td>
<td>0.0306</td>
<td>0.0307</td>
<td>0.0306</td>
<td>0.0260</td>
<td>0.0337</td>
<td>0.0362</td>
</tr>
<tr>
<td>Datasets</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Distances</td>
<td>0.0377</td>
<td>0.0399</td>
<td>0.0320</td>
<td>0.0354</td>
<td>0.0279</td>
<td>0.0275</td>
<td>0.0308</td>
<td>0.0318</td>
</tr>
<tr>
<td>Datasets</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>Distances</td>
<td>0.0325</td>
<td>0.0274</td>
<td>0.0282</td>
<td>0.0366</td>
<td>0.0279</td>
<td>0.0301</td>
<td>0.0346</td>
<td>0.0367</td>
</tr>
<tr>
<td>Datasets</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>Distances</td>
<td>0.0296</td>
<td>0.0267</td>
<td>0.0316</td>
<td>0.0336</td>
<td>0.0355</td>
<td>0.0287</td>
<td>0.0321</td>
<td>0.0277</td>
</tr>
<tr>
<td>Datasets</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>Distances</td>
<td>0.0319</td>
<td>0.0336</td>
<td>0.0267</td>
<td>0.0276</td>
<td>0.0257</td>
<td>0.0283</td>
<td>0.0304</td>
<td>0.0281</td>
</tr>
<tr>
<td>Datasets</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
</tr>
<tr>
<td>Distances</td>
<td>0.0281</td>
<td>0.0372</td>
<td>0.0345</td>
<td>0.0299</td>
<td>0.0338</td>
<td>0.0297</td>
<td>0.0328</td>
<td>0.0297</td>
</tr>
<tr>
<td>Datasets</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>Distances</td>
<td>0.0305</td>
<td>0.0354</td>
<td>0.0293</td>
<td>0.0385</td>
<td>0.0386</td>
<td>0.0322</td>
<td>0.0362</td>
<td>0.0319</td>
</tr>
<tr>
<td>Datasets</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distances</td>
<td>0.0293</td>
<td>0.0343</td>
<td>0.0341</td>
<td>0.0336</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S3. Predictive performances of O-GlcNAcylated sites based on different local window sizes using BPB features on jackknife test

<table>
<thead>
<tr>
<th>Window size</th>
<th>Sn (%)</th>
<th>Sp (%)</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>67.26</td>
<td>74.34</td>
<td>70.80</td>
</tr>
<tr>
<td>13</td>
<td>68.44</td>
<td>73.16</td>
<td>70.80</td>
</tr>
<tr>
<td>15</td>
<td>70.80</td>
<td>73.45</td>
<td>72.12</td>
</tr>
<tr>
<td>17</td>
<td>74.04</td>
<td>71.98</td>
<td>73.01</td>
</tr>
<tr>
<td>19</td>
<td>74.63</td>
<td>74.04</td>
<td>74.34</td>
</tr>
<tr>
<td>21</td>
<td>73.45</td>
<td>73.16</td>
<td>73.30</td>
</tr>
<tr>
<td>23</td>
<td>73.45</td>
<td>73.45</td>
<td>73.45</td>
</tr>
<tr>
<td><strong>25</strong></td>
<td><strong>74.63</strong></td>
<td><strong>74.63</strong></td>
<td><strong>74.63</strong></td>
</tr>
<tr>
<td>27</td>
<td>74.63</td>
<td>74.04</td>
<td>74.34</td>
</tr>
</tbody>
</table>

Table S4. Predictive performances of O-GlcNAcylated sites based on different local window sizes using BRABSB features on jackknife test

<table>
<thead>
<tr>
<th>Window size</th>
<th>Sn (%)</th>
<th>Sp (%)</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>74.04</td>
<td>75.52</td>
<td>74.78</td>
</tr>
<tr>
<td>13</td>
<td>71.39</td>
<td>74.04</td>
<td>72.71</td>
</tr>
<tr>
<td>15</td>
<td>73.45</td>
<td>72.86</td>
<td>73.16</td>
</tr>
<tr>
<td>17</td>
<td>75.81</td>
<td>73.45</td>
<td>74.63</td>
</tr>
<tr>
<td>19</td>
<td>75.22</td>
<td>75.22</td>
<td>75.22</td>
</tr>
<tr>
<td>21</td>
<td>74.93</td>
<td>74.04</td>
<td>74.48</td>
</tr>
<tr>
<td>23</td>
<td>77.29</td>
<td>76.40</td>
<td>76.84</td>
</tr>
<tr>
<td><strong>25</strong></td>
<td><strong>76.40</strong></td>
<td><strong>77.58</strong></td>
<td><strong>76.99</strong></td>
</tr>
<tr>
<td>27</td>
<td>75.81</td>
<td>77.29</td>
<td>76.55</td>
</tr>
</tbody>
</table>
Table S5. Predictive performances of O-GlcNAcylated sites based on different local window sizes using ANBPB features on jackknife test

<table>
<thead>
<tr>
<th>Window size</th>
<th>Sn (%)</th>
<th>Sp (%)</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>70.80</td>
<td>73.16</td>
<td>71.98</td>
</tr>
<tr>
<td>13</td>
<td>72.86</td>
<td>74.63</td>
<td>73.75</td>
</tr>
<tr>
<td>15</td>
<td>74.93</td>
<td>73.75</td>
<td>74.34</td>
</tr>
<tr>
<td>17</td>
<td>77.29</td>
<td>73.75</td>
<td>75.52</td>
</tr>
<tr>
<td>19</td>
<td>76.40</td>
<td>75.22</td>
<td>75.81</td>
</tr>
<tr>
<td>21</td>
<td>78.47</td>
<td>73.45</td>
<td>75.96</td>
</tr>
<tr>
<td><strong>23</strong></td>
<td><strong>79.94</strong></td>
<td><strong>77.58</strong></td>
<td><strong>78.76</strong></td>
</tr>
<tr>
<td>25</td>
<td>79.35</td>
<td>76.99</td>
<td>78.17</td>
</tr>
<tr>
<td>27</td>
<td>77.58</td>
<td>76.70</td>
<td>77.14</td>
</tr>
</tbody>
</table>

Table S6. Predictive performances of O-GlcNAcylated sites based on different local window sizes using relative RANS features on jackknife test

<table>
<thead>
<tr>
<th>Window size</th>
<th>Sn (%)</th>
<th>Sp (%)</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>64.60</td>
<td>72.57</td>
<td>68.58</td>
</tr>
<tr>
<td>13</td>
<td>67.55</td>
<td>70.21</td>
<td>68.88</td>
</tr>
<tr>
<td>15</td>
<td>68.73</td>
<td>71.98</td>
<td>70.35</td>
</tr>
<tr>
<td>17</td>
<td>69.91</td>
<td>72.27</td>
<td>71.09</td>
</tr>
<tr>
<td>19</td>
<td>68.44</td>
<td>72.86</td>
<td>70.65</td>
</tr>
<tr>
<td>21</td>
<td>68.73</td>
<td>73.16</td>
<td>70.94</td>
</tr>
<tr>
<td>23</td>
<td>70.21</td>
<td>71.68</td>
<td>70.94</td>
</tr>
<tr>
<td><strong>25</strong></td>
<td><strong>70.50</strong></td>
<td><strong>72.27</strong></td>
<td><strong>71.39</strong></td>
</tr>
<tr>
<td>27</td>
<td>69.03</td>
<td>73.16</td>
<td>71.09</td>
</tr>
</tbody>
</table>
Table S7. Comparison of O-GlcNAcPRED with YinOYang and OGlcnAcScan on the independent test dataset

<table>
<thead>
<tr>
<th>UniProt ID</th>
<th>O-GlcNAcylation site</th>
<th>YinOYang</th>
<th>OGlcnAcScan</th>
<th>O-GlcNAcPRED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q15750</td>
<td>S395</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>S396</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S399</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>T400</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S401</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q7Z589</td>
<td>S238</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>T535</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>S557</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>P48382</td>
<td>S439</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>S442</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q9UHD9</td>
<td>T113</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>S116</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>T124</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Q16186</td>
<td>S213</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S220</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>T221</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>T222</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>T225</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Q99996</td>
<td>T2794</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>T2797</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q6UN15</td>
<td>T205</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>T208</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Q15723</td>
<td>T375</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Key</td>
<td>T376</td>
<td>S400</td>
<td>S500</td>
<td>S572</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Prediction methods</td>
<td>Sn (%)</td>
<td>Sp (%)</td>
<td>Acc (%)</td>
<td>MCC</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>YinOYang</td>
<td>34.33</td>
<td>89.36</td>
<td>88.85</td>
<td>0.0725</td>
</tr>
<tr>
<td>O-GlcNAcscan</td>
<td>31.34</td>
<td>92.45</td>
<td>91.89</td>
<td>0.0847</td>
</tr>
<tr>
<td>O-GlcNAcPRED</td>
<td><strong>56.72</strong></td>
<td>64.77</td>
<td>64.70</td>
<td>0.0428</td>
</tr>
</tbody>
</table>

Table S8. Performances comparisons among O-GlcNAcPRED, YinOYang and O-GlcNAcscan