Interpreting MQ Phosphopeptide Results
Format of results

• Usually we analyse phosphoproteomic data with MaxQuant
• You get the following result files either as single files or in a combined Excel file:
  1. XX-XXX\textsuperscript{1}\_ProteinGroups
  2. XX-XXX\textsuperscript{1}\_Phosphosites (STY)
• On request we can also provide other MQ output files

1 = Your project number
Phosphosites vs. Phosphopeptides

• Phosphosites are **NOT** Phosphopeptides!!!
• A phosphosite is a serine/threonine/tyrosine with a distinct location in the protein sequence, which is phosphorylated
• Quantitative information for each site is derived from ALL detected peptides, which are phosphorylated at this position (=combined/condensed information)
Example

Protein X (P 19,32)

Protein X (P19)

Protein X (P32)

Phosphopeptides
Example (evidence table)

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Intensity</th>
<th>OR for SILAC based approaches</th>
<th>Ratio H/L (normalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RILILS(P)GAAPS(P)EE</td>
<td>300000</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td>RILILS(P)GAAPSEE</td>
<td>100000</td>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td>RILILSGAAPS(P)EE</td>
<td>500000</td>
<td></td>
<td>1.7</td>
</tr>
</tbody>
</table>

This information can be found in the MQ output file: evidence.txt
Example (phosphosite table)

This information can be found in the MQ output file: Phospho (STY)Sites.txt
Phospho (STY) table

• This table (+/- the proteingroups table) is usually sufficient to make qualified assumptions about changes in the phosphorylation status of proteins between samples

• Key quantitative values are:
  – Occupancy (mainly applicable for SILAC data)
  – Intensity [1/2/3] OR Ratio H/L [1/2/3]

__1, __2, __3? What does that mean?
There is quantitative information derived from single, double or triple (and higher) phosphorylated peptides.

Using this intensities is MUCH better then looking just at the summed site intensities, because in a biological system the same protein may have distinct functions when it is differentially phosphorylated.
**Occupancy**

- Fraction of protein for which the site is modified (Value from 0 - 1 ➔ 0 - 100% modified)
  - This value is calculated from three separate ratios:
    - Ratio H/L proteingroup
    - Ratio H/L modified peptide (here: phosphorylation)
    - Ratio H/L unmodified counterpart peptide
  - Occupancy is calculated for every labeling state if all three ratios are present
Occupancy example

Untreated (light)

Treated (heavy)

<table>
<thead>
<tr>
<th>Positions within proteins</th>
<th>Occupancy L</th>
<th>Occupancy H</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Quantitative site information
If I don’t have occupancy values I must look at two types of data:

1. **Phosphosite quant data (Phosphosites (STY).txt)**
   - A. Intensities__1/2/3 ➔ label free data
   - B. Ratio H/L (norm)__1/2/3 ➔ SILAC data

2. **ProteinGroup quant data (ProteinGroups.txt)**
   - A. LFQ Intensity ➔ label-free data
   - B. Ratio H/L (norm) ➔ SILAC data
Example

<table>
<thead>
<tr>
<th>Positions with in proteins</th>
<th>Amino acid</th>
<th>Intensity untreated</th>
<th>Intensity untreated_1</th>
<th>Intensity untreated_2</th>
<th>Intensity untreated_3</th>
<th>Intensity treated</th>
<th>Intensity treated_1</th>
<th>Intensity treated_2</th>
<th>Intensity treated_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>S</td>
<td>400000</td>
<td>100000</td>
<td>300000</td>
<td>0</td>
<td>125000</td>
<td>50000</td>
<td>75000</td>
<td>0</td>
</tr>
</tbody>
</table>

From this we could assume that serine at position 19 is more phosphorylated under the untreated condition.
Example

<table>
<thead>
<tr>
<th>Positions with in proteins</th>
<th>Amino acid</th>
<th>Intensity untreated</th>
<th>Intensity untreated 1</th>
<th>Intensity untreated 2</th>
<th>Intensity untreated 3</th>
<th>Intensity treated</th>
<th>Intensity treated 1</th>
<th>Intensity treated 2</th>
<th>Intensity treated 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>S</td>
<td>400000</td>
<td>100000</td>
<td>300000</td>
<td>0</td>
<td>125000</td>
<td>50000</td>
<td>75000</td>
<td>0</td>
</tr>
</tbody>
</table>

But what about the protein level? Maybe the whole protein is less abundant in the treated sample?
Example

<table>
<thead>
<tr>
<th>Positions with in proteins</th>
<th>Amino acid</th>
<th>Intensity untreated</th>
<th>Intensity untreated__1</th>
<th>Intensity untreated__2</th>
<th>Intensity untreated__3</th>
<th>Intensity treated</th>
<th>Intensity treated__1</th>
<th>Intensity treated__2</th>
<th>Intensity treated__3</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>S</td>
<td>400000</td>
<td>100000</td>
<td>300000</td>
<td>0</td>
<td>125000</td>
<td>50000</td>
<td>75000</td>
<td>0</td>
</tr>
</tbody>
</table>

So you need to also look at the quant protein data (proteingroups.txt) before making any assumptions!!!