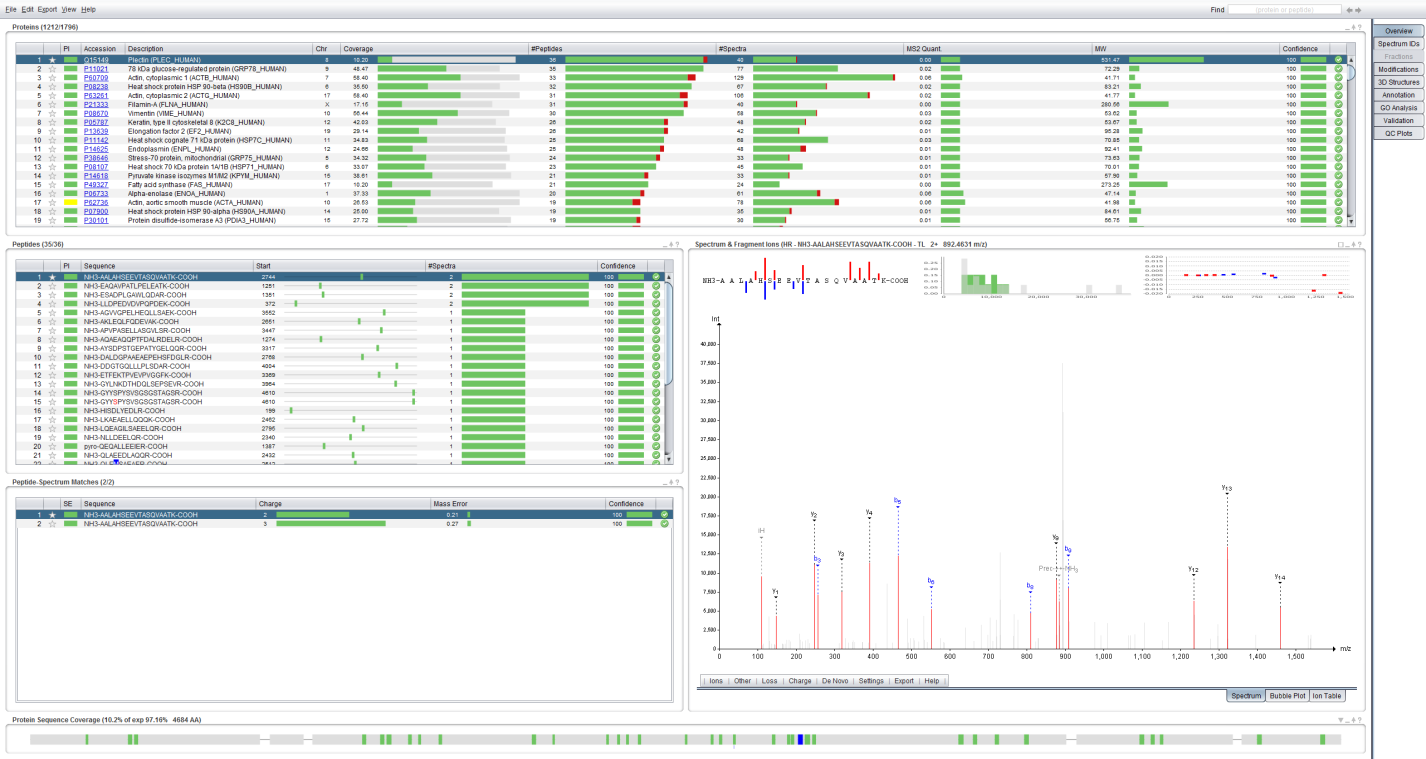
Spectrum Counting Quantification

Spectrum counting based quantification is the simplest quantification method in proteomics. It relies on the rationale that highly abundant peptides will trigger more MS/MS spectra measurements, also likely to present a higher signal to noise ratio. As a result, peptides from abundant proteins are more likely to be identified and in more spectra[1](#_ENREF_1). Two approaches were hence followed: count the number of peptides identified for a given protein – like in the emPAI index[2](#_ENREF_2) method – or count the number of spectra ascribed to a protein[3](#_ENREF_3).

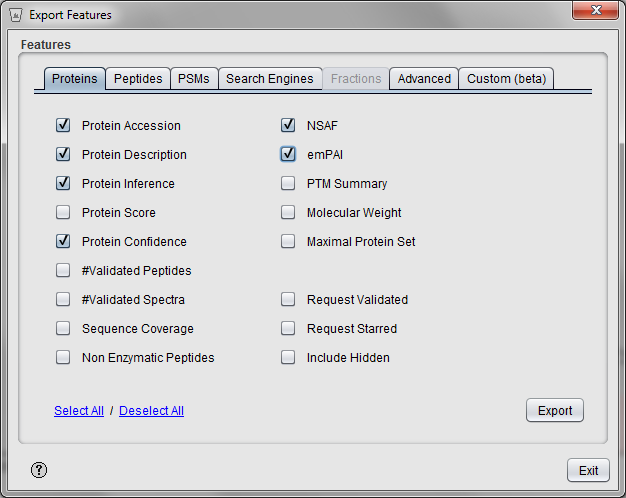
*In your opinion, what are the advantages/shortcomings of spectrum counting? [2.1a]*

Spectrum counting indexes can be exported from almost all identification software. We will now look at an example using PeptideShaker: load the PeptideShaker example dataset as explained in the identification chapter. You should now see the following:

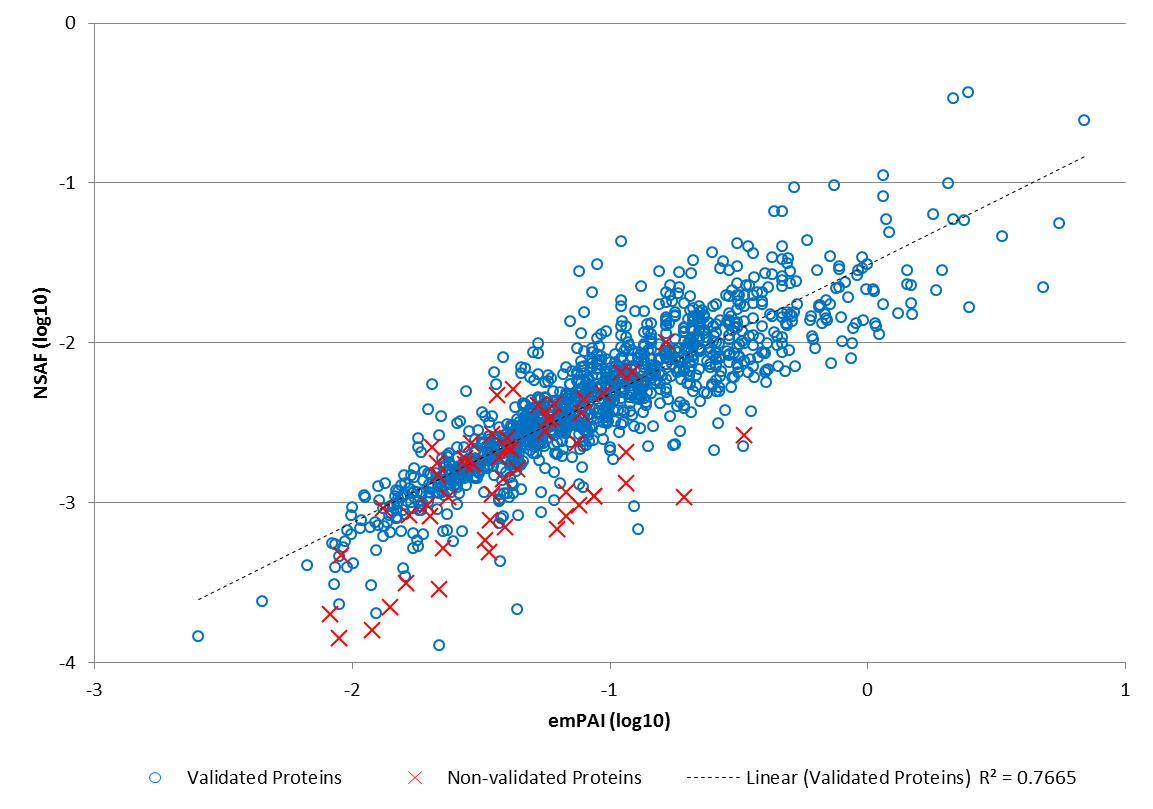


Note that the spectrum counting quantification index has already been calculated for you and is readily displayed in the ‘MS2 Quant.’ column of the protein table. By default the NSAF index is selected as it was found to be more reliable[4](#_ENREF_4). Note also that PeptideShaker corrects the NSAF index automatically for sections of the protein sequence which cannot generate observable peptides and for protein inference problems using the methods detailed in the identification chapter[5](#_ENREF_5).

Export the results via the ‘Export’ menu using the ‘Identification Features’ option, make sure the protein tab is selected, and select the following: ‘Protein Accession’, ‘Protein Description’, ‘Protein Inference’, ‘Protein Confidence’, ‘NSAF’ and ‘emPAI’. Make sure that not only validated proteins are exported by unchecking ‘Request Validated’:



The exported values can be imported into Microsoft Excel for post-processing. If you plot the NSAF index against the emPAI indexes of the target validated and non-validated protein matches, you will get the following:

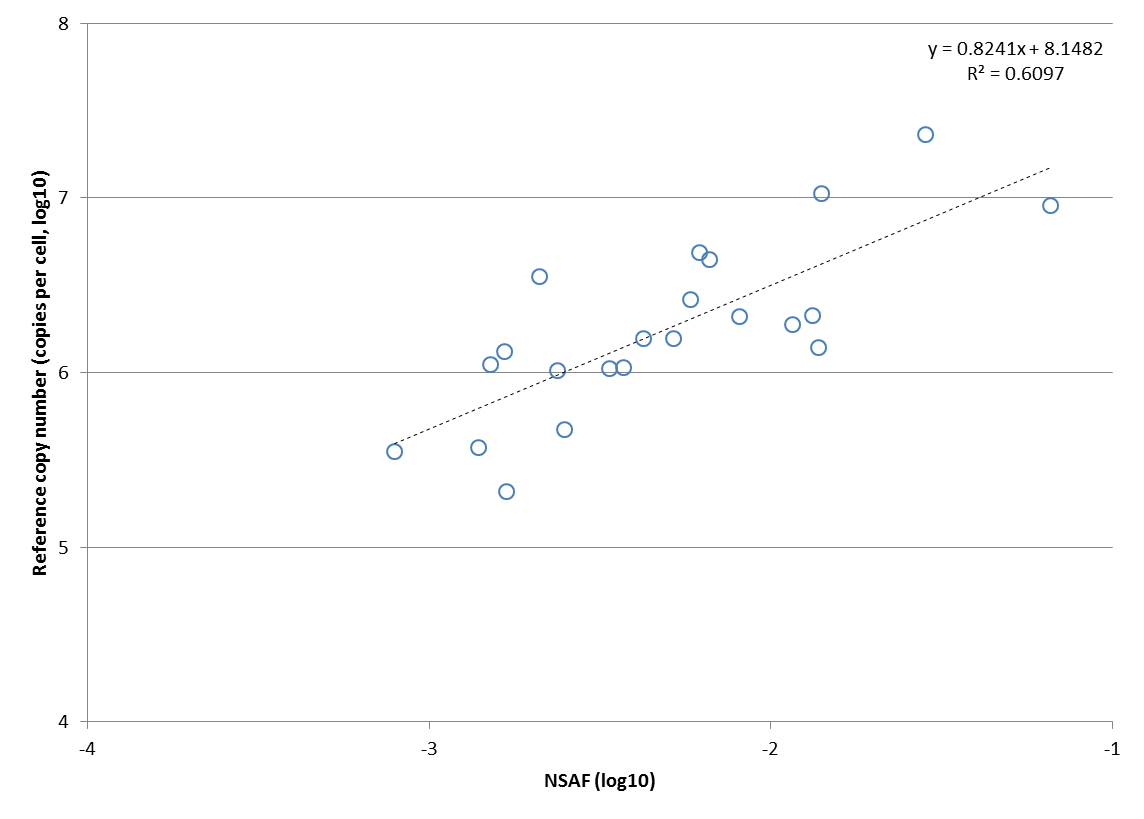


*How do the indexes compare? Why are we using a log scale? [2.1b]*

In order to give an illustrative value to these spectrum counting indexes, we will calibrate these values according to abundance levels from the literature[6](#_ENREF_6). In the file ‘spectrum\_counting.xlsx’, you will find these reference values and in the column ‘spectrum counting 1h’, the NSAF values exported by PeptideShaker.

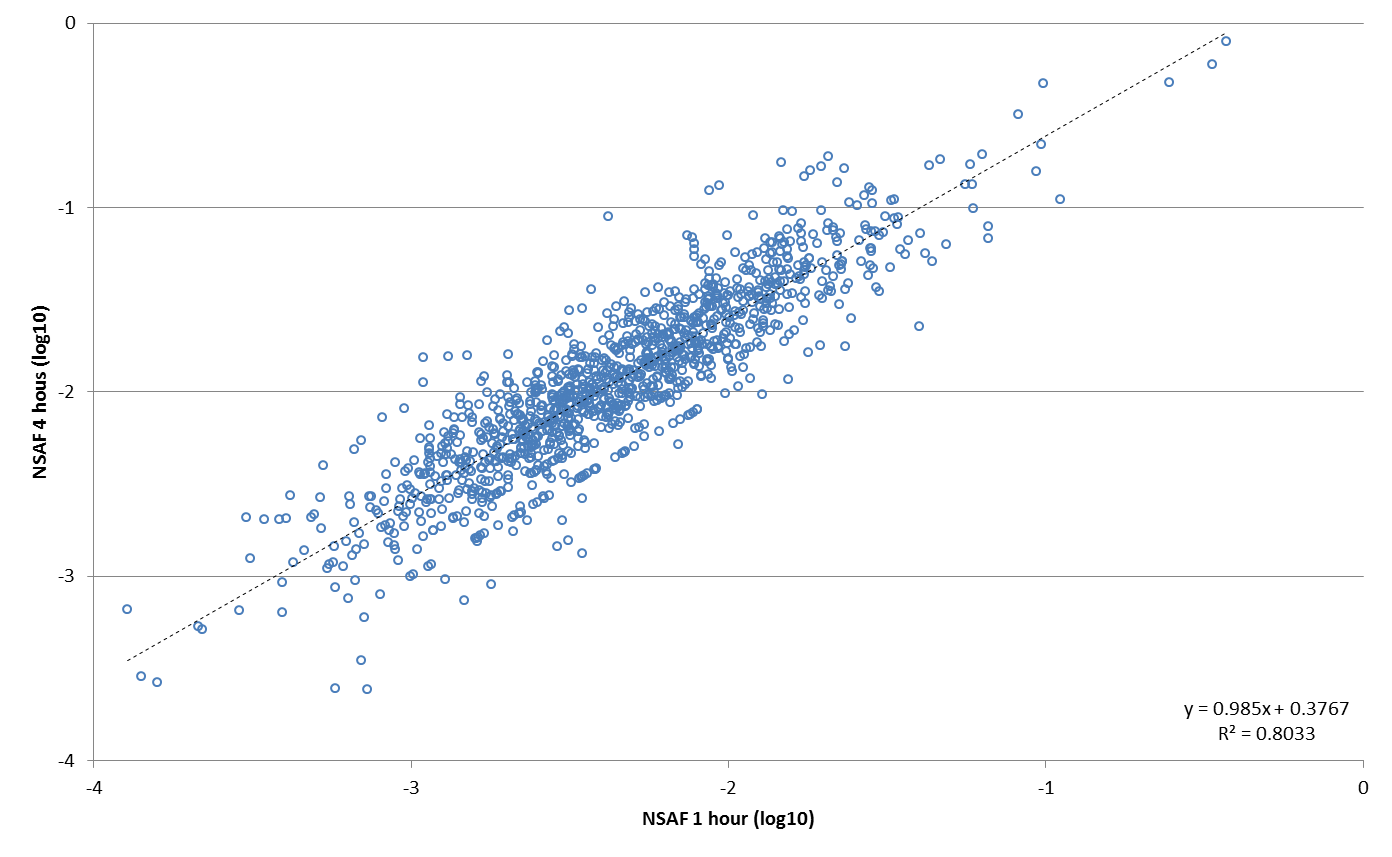
*Why are some values missing? Can we infer a detection limit for our experiment? [2.1c]*

If you plot the reference numbers against the NSAF value, you should see the following:

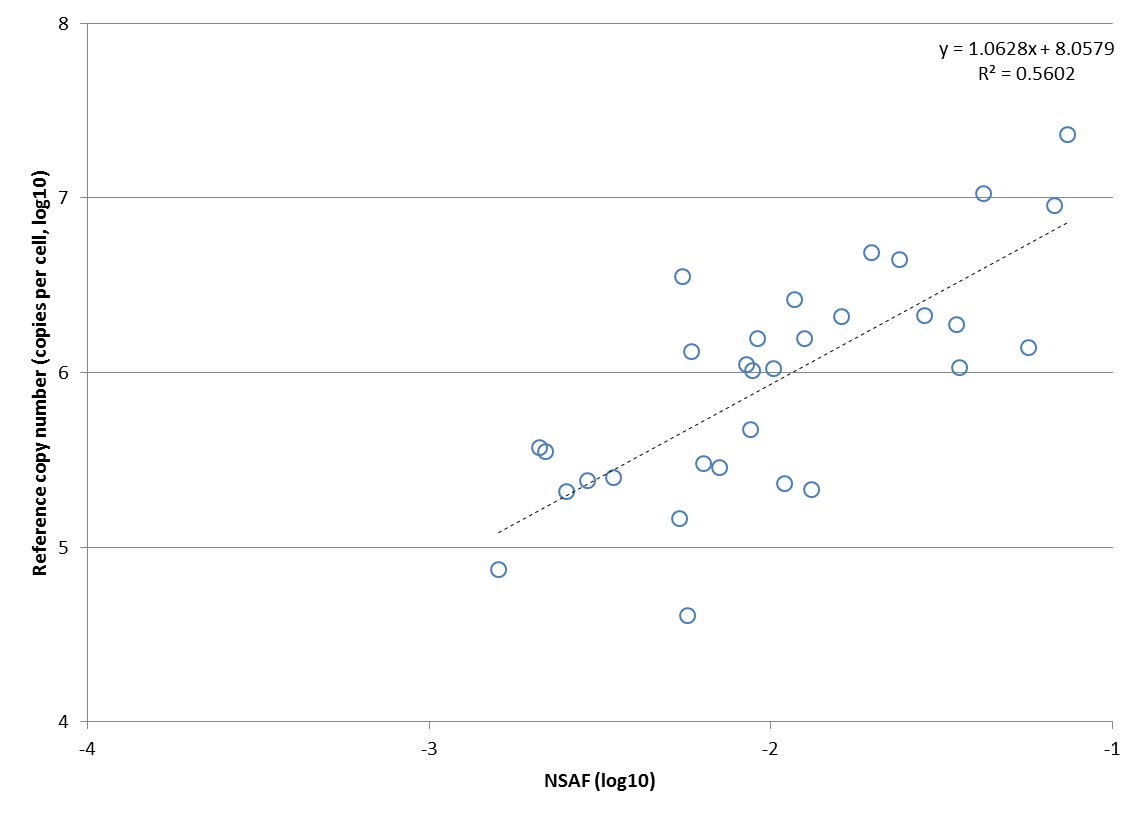


As you can see, we can use a simple formula to translate the NSAF index into a protein count using a linear regression (R2=0.6097).

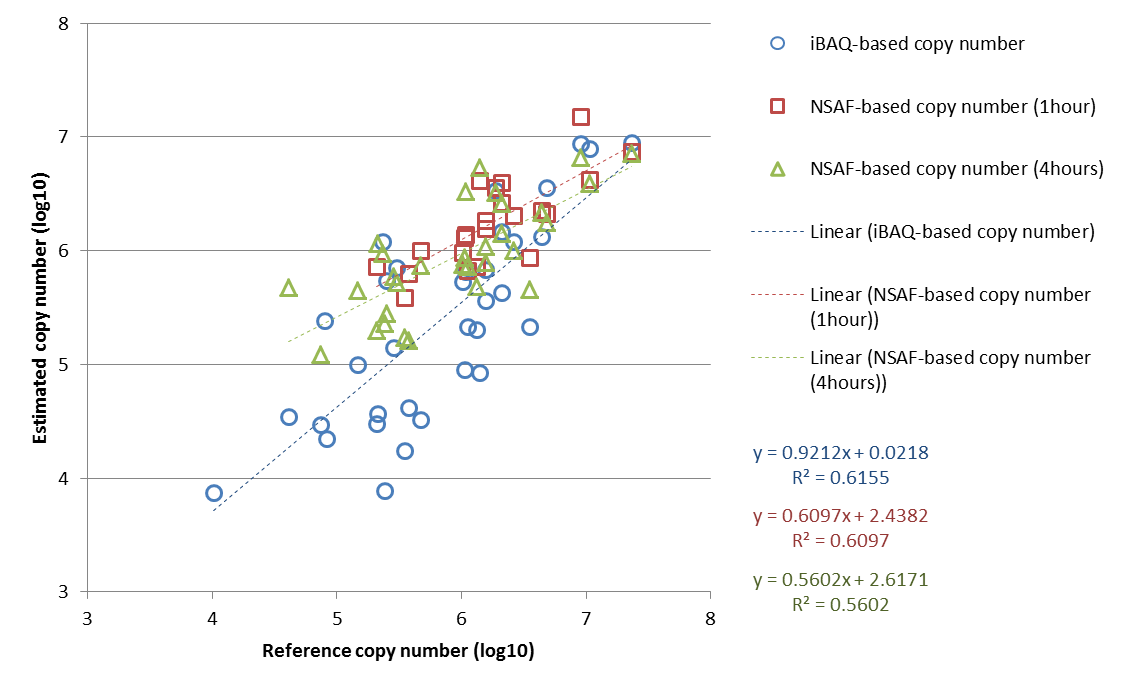
In order to increase our proteome coverage, we measured the same sample with a longer gradient (4 hours) - the PeptideShaker export is provided in the excel sheet. As you can see, 3480 protein matches were validated for this run, three times more than in the previous dataset (1212). If you plot the commonly identified proteins, you will observe the good correlation between the datasets:



Conduct the same normalization as for the previous dataset in order to obtain copy numbers. You should obtain the following:

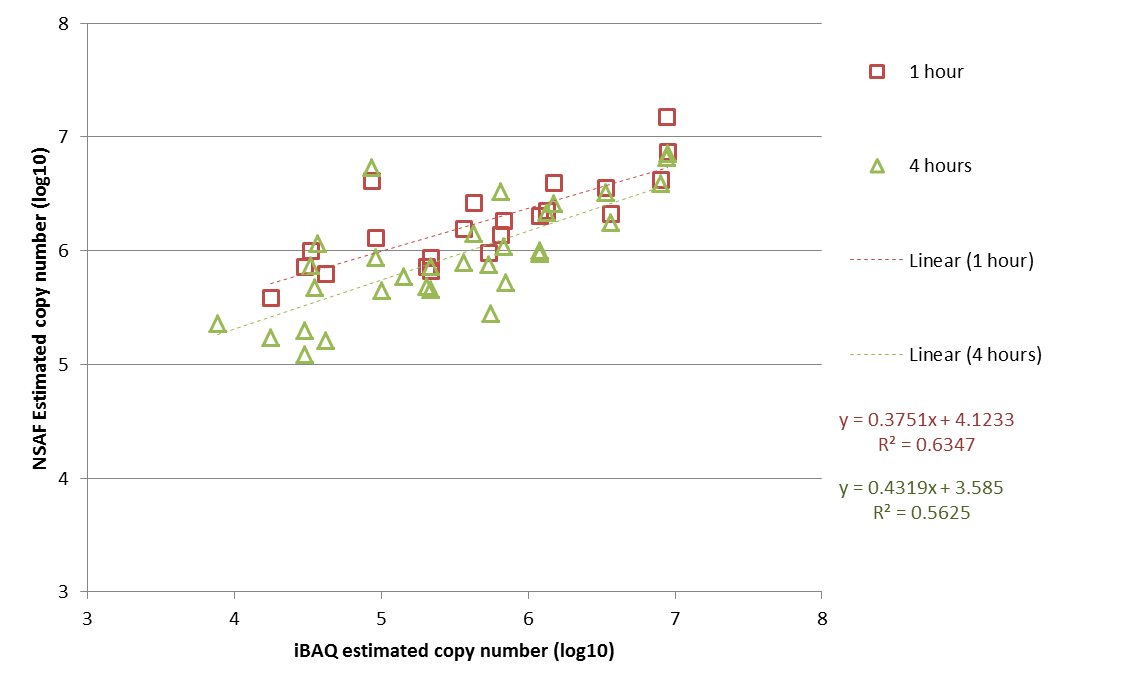


In the ‘iBAQ-based copy number’, we provide the protein abundances obtained from MS1 intensity-based quantification[7](#_ENREF_7). Plotting the values obtained from the three datasets results in:



*What are the performances of the quantification on the different datasets? [2.1d]*

When comparing the spectrum counting approach to the intensity based quantification, you obtain the following:



Note that the R2 values for spectrum counting compared to intensity based quantification are very similar as the ones compared to the reference. We can hence assume that this variability is due to the quantification error and not the reference.

*What are the advantages and shortcomings of spectrum counting in comparison to intensity based quantification? [2.1e]*

References

1. Liu, H., Sadygov, R.G. & Yates, J.R., 3rd A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical chemistry* **76**, 4193-4201 (2004).

2. Ishihama, Y. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP* **4**, 1265-1272 (2005).

3. Powell, D.W. et al. Cluster analysis of mass spectrometry data reveals a novel component of SAGA. *Molecular and cellular biology* **24**, 7249-7259 (2004).

4. Colaert, N., Gevaert, K. & Martens, L. RIBAR and xRIBAR: Methods for reproducible relative MS/MS-based label-free protein quantification. *Journal of proteome research* **10**, 3183-3189 (2011).

5. Burkhart, J.M. et al. The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood* **120**, e73-82 (2012).

6. Zeiler, M., Straube, W.L., Lundberg, E., Uhlen, M. & Mann, M. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP* **11**, O111 009613 (2012).

7. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology* **7**, 548 (2011).