Description of controlled spike-in experiment

September 19, 2011

1 Experimental design

The experiment aimed at evaluating the ability of significance analysis to detect known fold changes, as well as the extent of false positive changes among proteins spiked in constant concentrations. It was conducted in a case-control experimental design with six mixtures that share a same complex background. To evaluate the sensitivity of detecting known fold changes, six proteins were spiked into the background in varying concentrations, according to the Latin Square design[1] in **Figure 1a**. The design is advantageous as it allows us to evaluate the ability to detect a range of fold changes over a series of proteins, baseline concentrations, and runs, while limiting the total number of mixtures. To evaluate the specificity of the models, six additional proteins were spiked in different but constant concentrations to all the mixtures as in **Figure 1b**. Each protein was represented by 2 peptides, and each peptide was represented by 3 transitions.

2 Sample selection

Each spike-in mixture contained three components: (1) 24 crude synthetic unlabeled ("endogenous") peptides (purchased from JPT technology, Germany), representing twelve *S. cerevisiae* proteins, with 2 peptides per protein; (2) the isotopically labeled "reference" peptides with the same sequences of amino acids (purchased from JPT technology, Germany); and (3) a complex peptide background generated from whole worm extracts.

3 Sample preparation

Six different mixtures were generated from these components for the purposes of relative quantification. Twelve endogenous peptides from six proteins were mixed at varying concentrations in a Latin Square design (**Figure 1a**). These peptides serve as positive controls, and represent proteins with true changes in concentration across mixtures. The second set of twelve endogenous peptides from the remaining six proteins was mixed at constant concentrations (**Figure 1b**). These peptides serve as negative controls, and represent proteins in constant concentration across mixtures.

The maximum concentration of each endogenous peptide in a mixture was 50 fmol. This concentration gives signal intensities, depending on each individual peptide, in the range of 10^5 to 10^6 counts per second[2]. The remaining concentrations were obtained by a relative fractionation. Among the controls, the relative concentrations of each peptide between two mixtures varied from two fold up to a maximum of 512 fold, in order to resemble the natural dynamic range encountered in biological samples, as well as to test the dynamic range of the triple quadrupole mass spectrometry instrument. The reference peptides, labeled with a heavy Arginine (R) or Lysine (K), were further spiked into each sample at constant concentrations, equal to one-tenth of the maximum[3].

The resulting mixture of synthetic peptides was spiked into a complex and non-fractionated peptide background prepared from a whole *C. elegans* lysate, which generated from mixed-stage, wild-type *C. elegans* worms[4, 5].

(a) Proteins in varying concentrations

	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5	Mixture 6
YBR132C	Max	Max/512	Max/256	Max/128	Max/32	Max/8
YBR144C	Max/8	Max	Max/512	Max/256	Max/128	Max/32
YBR147W	Max/32	Max/8	Max	Max/512	Max/256	Max/128
YBR184W	Max/128	Max/32	Max/8	Max	Max/512	Max/256
YBR203W	Max/256	Max/128	Max/32	Max/8	Max	Max/512
YBR284W	Max/512	Max/256	Max/128	Max/32	Max/8	Max

(b) Proteins in constant concentration

	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5	Mixture 6
YBR168W	Max	Max	Max	Max	Max	Max
YBR186W	Max	Max	Max	Max	Max	Max
YBR204C	Max/32	Max/32	Max/32	Max/32	Max/32	Max/32
YBR228W	Max/32	Max/32	Max/32	Max/32	Max/32	Max/32
YBR250W	Max/256	Max/256	Max/256	Max/256	Max/256	Max/256
YBR270C	Max/256	Max/256	Max/256	Max/256	Max/256	Max/256

Figure 1: Experimental design of the controlled spike-in experiment. The maximal concentration in each mixture, denoted "Max", was 50 fmol for all proteins. **(a)** Concentrations of proteins spiked in a Latin Square design. **(b)** Concentrations of proteins spiked in constant concentration.

4 Mass spectrometric analysis

The targeted mass spectrometry analysis was optimized as in (Picotti et. al., 2010). The order of the mixtures was randomized, and each mixture was profiled in two mass spectrometry runs. The procedure was repeated a second time to acquire a second set of replicate runs. Quantitative measurements were performed on 4000 QTRAP instrument in time-scheduled SRM mode using software Analyst 1.5 (Applied Biosystems/MDS Sciex), and peptide separation using Tempo Nano LC system (Applied Biosystems/MDS Sciex).

5 Signal processing

Each transition was quantified by its peak apex automatically, using MultiQuant software package (Applied Biosystems/MDS Sciex). Due to sample evaporation, the strength of the signal varied between runs in the second replicate set. The quantified transitions were manually curated, resulting in the total of 49 transitions and a small number of missing values.

1. Montgomery, D. *Design and analysis of experiments* (Wiley, 2005).

2. Picotti, P. *et al.* High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature Methods* **10**, 43–46 (2010).

3. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Science of the USA* **100**, 6940–6945 (2003).

4. Jovanovic, M. *et al.* A quantitative targeted proteomics approach to validate predicted microrna targets in *C. elegans*. *Nature Methods* **10**, 837–842 (2010).

5. Schrimpf, S. P. *et al.* Comparative functional analysis of the caenorhabditis elegans and drosophila melanogaster proteomes. *PLoS Biology* **7**, e48 (2009).