Reproducibility of differential proteomic technologies in CPTAC fractionated xenografts

Authors

David L. Tabb[1]*, Xia Wang[2], Steven A. Carr[3], Karl R. Clauser[3], Philipp Mertins[3], Matthew C. Chambers[1], Jerry D. Holman[1], Jing Wang[1], Bing Zhang[1], Lisa J. Zimmerman[4], Xian Chen[5], Harsha P. Gunawardena[5], Sherri R. Davies[6], Matthew J.C. Ellis[6,7], Shunqiang Li[6], R. Reid Townsend[6], Emily S. Boja[8], Karen A. Ketchum[9], Christopher R. Kinsinger[8], Mehdi Mesri[8], Henry Rodriguez[8], Tao Liu[10], Sangtae Kim[10], Jason E. McDermott[10], Samuel H. Payne[10], Vladislav A. Petyuk[10], Karin D. Rodland[10], Richard D. Smith[10], Feng Yang[10], Daniel W. Chan[11], Bai Zhang[11], Hui Zhang[11], Zhen Zhang[11], Jian-Ying Zhou[11], and Daniel C. Liebler[4]

1. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37232
2. Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221
3. Proteomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142
4. Department of Biochemistry, Vanderbilt University, Nashville, TN 37232
5. Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599
6. Department of Medicine, Washington University, St. Louis, MO 63110
7. Current address: Smith Breast Center, Baylor College of Medicine, Houston, TX 77030
8. Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, MD 20892
9. Enterprise Science and Computing, International, Rockville, MD 20850
10. Division of Biological Sciences, Pacific Northwest National Laboratory, Richland, WA 99352
11. JHMI and Division of Clinical Chemistry, Johns Hopkins University, Baltimore, MD 21231

Contents

Authors 1
Supporting Method A: Instrument-specific protocol summary 3
Supporting Method B: Bayesian hierarchical model 5
Supporting Figure 1: Checking the variance constancy assumption 6
Supporting Figure 2: Kappa evaluation of differential gene overlap 7
Supporting Figure 3: Rank correlation plots for repeatability 8
Supporting Figure 4: Alternate analysis showing consistency for OVelos@45 (iTRAQ) 11
Supporting Figure 5: Comparing SPC and XIC gene differences from common data 13
Supporting Figure 6: Comparing differential genes within technologies 14
References 16

Supporting Method A: Instrument-specific protocol summary

OVelos@10 employed a 24-fraction bRPLC technique but added an unfractionated RPLC experiment for each sample, totaling 25 LC-MS/MS experiments per iTRAQ 4plex. Each LC-MS/MS experiment lasted an average of 5340 seconds. This site used the 114 and 115 channels for WHIM2, while the 116 and 117 channels represented WHIM16. The instrument produced a total of six replicates, starting on these dates: 4/29/2013, 5/18/2013, 6/20/2013, 8/9/2013, 9/24/2013, and 10/15/2013. In every replicate, the final fraction was collected within four days of the initial fraction. The WHIM replicates were "interstitial" in that each replicate was run between samples from TCGA; the WHIMs were intended as quality controls for the TCGA samples.

OVelos@45 also employed a 24-fraction bRPLC technique but without the unfractionated RPLC experiment, yielding 24 LC-MS/MS experiments per iTRAQ 4plex. Each LC-MS/MS experiment lasted an average of 5950 seconds. The same mapping of WHIMs to channels was applied as at site 10. The instrument produced a total of five replicates, starting on these dates: 3/3/2013, 5/19/2013, 6/21/2013, 8/13/2013, and 9/23/2013. In contrast to site 10, site 45 replaced some of the fractions in each replicate due to quality assessment with re-runs that took place as much as seven months later. Like site 10, the data were collected interstitially. Notably, OVelos@45 performed full process replicates, starting from tissue, in each experiment.

QExac@56 was also an iTRAQ site, but it employed a Q-Exactive instrument with 24 bRPLC fractions and a flow-through LC-MS/MS experiment (denoted as fraction 'A') for each replicate. Each LC-MS/MS experiment spanned an average of 6600 seconds. Five replicates were produced on these dates: 3/3/2013, 5/19/2013, 6/21/2013, 8/13/2013, and 9/23/2013. In every case, the final fraction was collected no more than three days after the initial fraction for a replicate. On the first, fourth, and fifth replicates, site 56 alternated iTRAQ channels so that WHIM2 was measured on 114 and 116 reporter ions while WHIM16 was measured on 115 and 117 reporter ions. On the second and third replicates, the site used 114 and 115 for WHIM2 and 116 and 117 for WHIM16. Site 56 also collected its WHIM data interstitially with TCGA samples. As was true for OVelos@45, site 56 began each process replicate with a separate aliquot of powderized tumor material.

OVelos@65 collected label-free data, with fifteen bRPLC fractions to each WHIM sample (so each comparison required thirty LC-MS/MS experiments). Each LC-MS/MS experiment spanned 4800 seconds. This site collected ten WHIM replicates interstitially during TCGA sample analysis and another six WHIM replicates interstitially during subsequent non-TCGA sample sets. The first ten replicates were collected between 8/2/2012 and 2/16/2013 on a regular rotation (the block structure included WHIM16, five TCGA samples, WHIM2, and then five TCGA samples). The final six replicates were collected between 2/25/2013 and 6/25/2013. In every case, the final fraction was collected within four days of the initial fraction. Producing sixteen replicates of two samples without multiplexing, each comprising fifteen RAW files, led to this instrument producing 480 of the 1095 files included in this analysis.

OElite@65 also collected label-free data in fifteen bRPLC fractions, using a very similar method to the above Orbitrap Velos at the same site. Its LC-MS/MS time spanned an average of 4200 seconds, using the shortest time (17.5 hours) per sample of any instrument in the study. Instead of collecting these

data in the interstices of other large experiments, these four replicates for both samples were collected as a quality control for an instrument that had delivered lower sensitivity than expected. Its replicates were collected back-to-back between 1/9/2014 and 1/21/2014.

QExac@98 collected label-free data using only five bRPLC fractions, though with each LC-MS/MS experiment spanning 14,400 seconds. This instrument and OVelos@65 used the same overall amount of time per sample, but QExac@98 split that time among far fewer fractions. This workflow was intended to simplify the comparison of extracted ion chromatograms (XICs) across experiments. The first three replicates were separated by a week (6/1/2014, 6/7/2014, 6/14/2014), with WHIM2 immediately preceding WHIM16. The fourth replicate was run right on the heels of the third (6/15/2014). The fifth through tenth replicates were run back-to-back between 7/16/2014 and 8/1/2014.

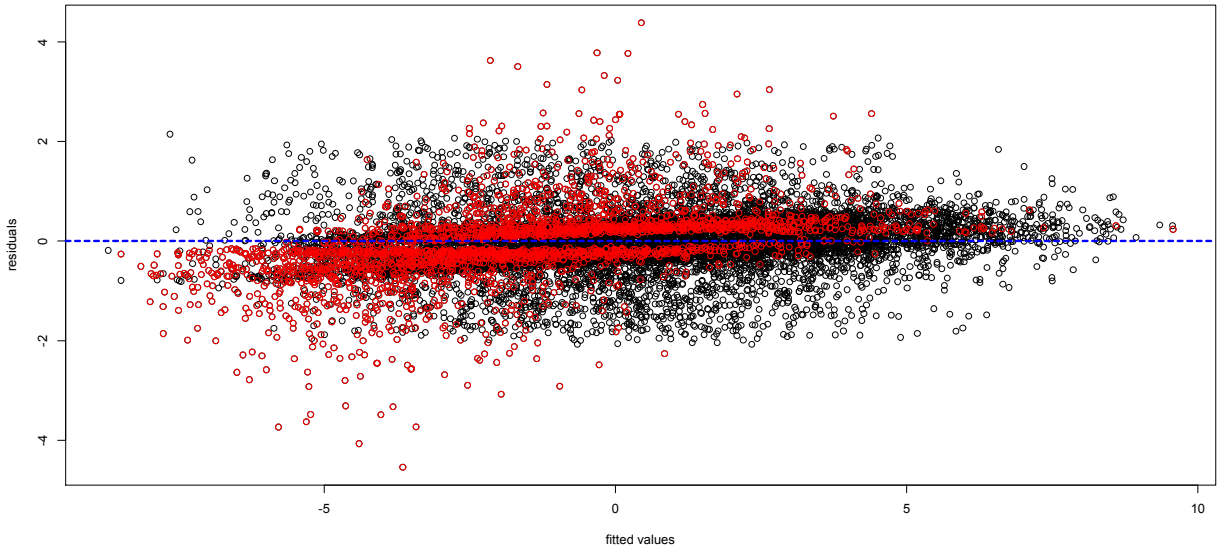
Supporting Method B: Bayesian hierarchical model

A benefit provided by Bayesian approach is that it can incorporate helpful information in statistical modeling through the prior distribution. Before the comparison between any two groups of abundance values, there are naturally three expected outcomes: down-regulated, no change, or up-regulated. Thus, a three-part prior was assumed on the log-ratios of genes, a normal distribution $N(0, \delta^2)$ for non-differentially expressed genes and two uniform distributions for differentially expressed genes ($U(\omega_+, \kappa_+)$ for higher expression in WHIM2 and the $U(-\kappa_-, -\omega_-)$ for higher expression in WHIM16), where $\kappa_+, \kappa_-, \omega_+, \omega_-$ are all positive, and $U(a, b)$ represents a uniform distribution with lower bound a and upper bound b. Thus, the parameters κ_+, ω_-, and κ_-, ω_+ set the upper and lower limits in the uniform mixtures. We set the magnitude of κ_+ and κ_- large enough to cover the expected maximum range of the log ratios. The limits set by ω_+ and ω_- determine the boundaries to classify a gene as higher in WHIM2 or higher in WHIM16 or unchanging. The range of this limit was discussed by Baladandayuthapani et al[1]. Here we set both ω_+ and ω_- to 0.1. With the value of 0.1, essentially any genes with less than 1.1 fold ratio on the normalized abundance were not classified as being differentially expressed. The variance in the prior normal distribution (δ^2) was imposed as an inverse gamma prior with a small mean[1] and was estimated in the model. This type of uniform-normal mixture prior has been widely used in differentiation analysis for microarray data analysis[2] and comparative genomic hybridization data[1].

The parameters in the model were estimated by sample averages of the posterior output from Markov chain Monte Carlo sampling[3,4]. Moreover, the model allowed for simultaneous testing of the large number of genes through the calculation of posterior probabilities of their no-change (null) and up- or down-regulated (non-null) status, with genes in the non-null group being those differentiated between the sample types (WHIM2 or WHIM16). To compare each of the n genes, the posterior probability that a gene belonged to the null status was computed and denoted as $p_i, i = 1, \ldots, n$. To control the FDR of differentiation analysis, a threshold was set on the n posterior probabilities as a cutoff point to classify genes into null and non-null status[5]. Particularly, the n posterior probabilities were first ranked from the smallest to the highest as $p_{(1)}, p_{(2)}, \ldots, p_{(n)}$ with $p_{(1)}$ the minimum and $p_{(n)}$ the maximum. To control FDR at a level α, let k be the largest set of the top k smallest posterior probabilities, $p_{(1)}, p_{(2)}, \ldots, p_{(\mathrm{k})}$, which has an average posterior probability less than or equal to α. The genes corresponding to these k smallest posterior probabilities were classified as differentially expressed (non-null status) and the remaining genes as non-differentially expressed (null status).

Supporting Figure 1: Checking the variance constancy assumption

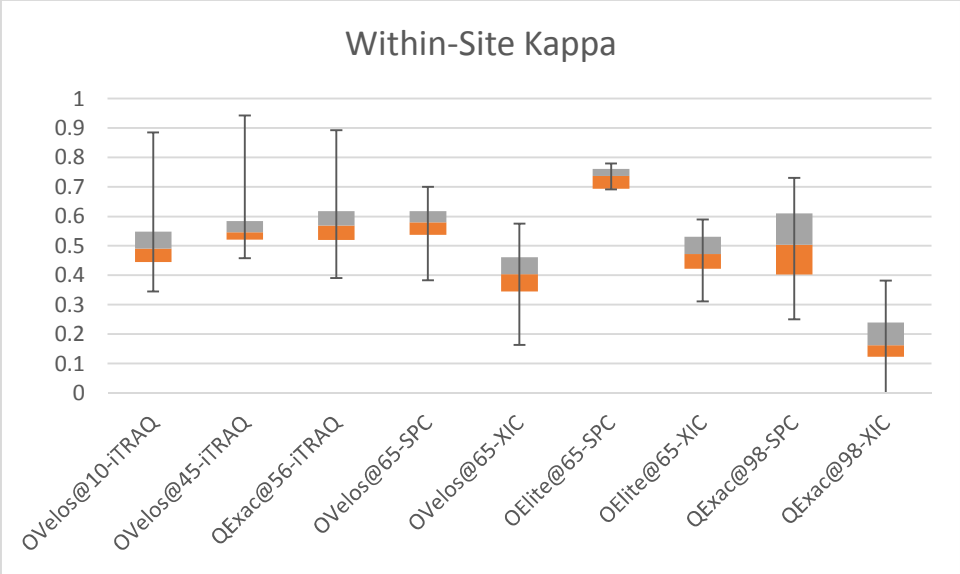
In one-versus-one comparison, it is assumed that non-differential genes will have normally distributed intensities with a constant variance in XIC and iTRAQ analysis, irrespective of intensity magnitude. We plotted the fitted values against the residuals. If variances are constant with regard to the intensity levels, we expect to see the residuals randomly scattered around the horizontal reference line (y=0). As an example, the plot from QExac@98 replicate A is shown (Red = differences at 5% FDR). It seems that the equal variance assumption is approximately satisfied. Note that we only used the non-differentially expressed genes in the variance estimation (black dots).

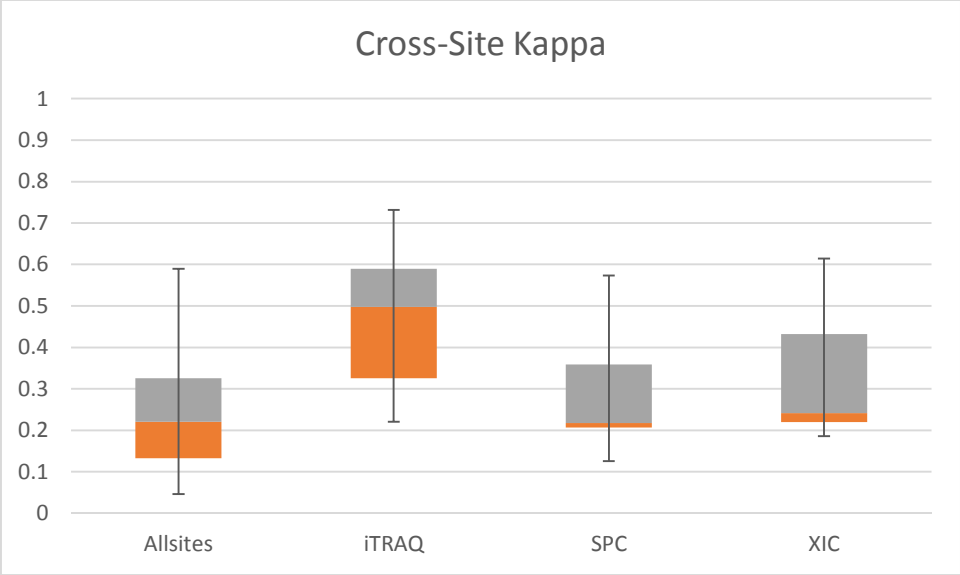


Supporting Figure 2: Kappa evaluation of differential gene overlap

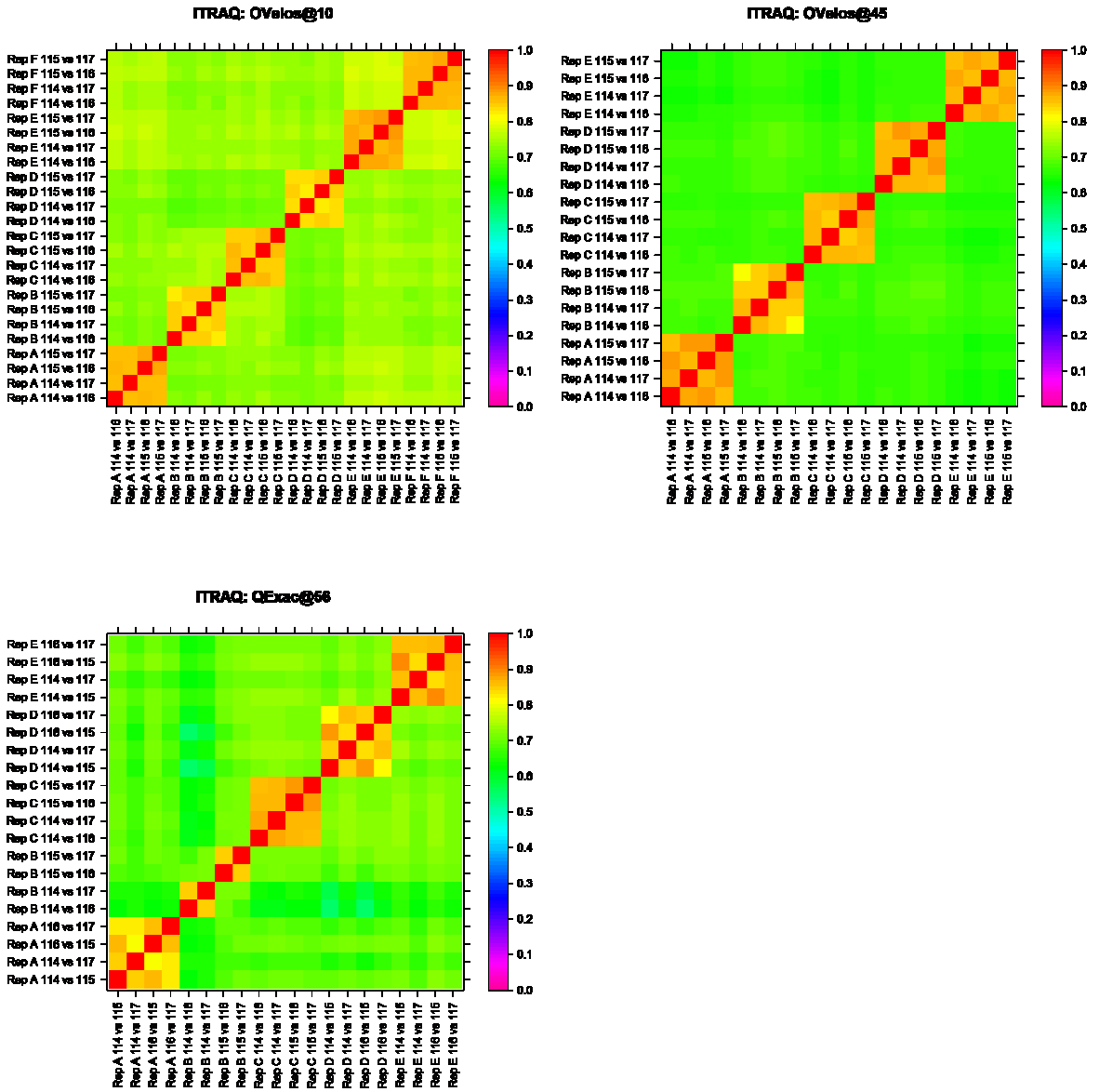
For each instrument, we compared by Cohen's kappa statistic the individual replicates to determine how well their set of differential genes overlapped with the differences elicited by every other replicate produced by that instrument. For iTRAQ comparisons, the highest kappa metrics were produced by two different uses of the same channel (such as for QExac@56, where replicate C 115vs116 compared to replicate C115vs117 produced a high value of 0.89).

In each plot, the bars represent the interquartiles and median, and the whiskers represent the range from the maximum to the minimum for each instrument.

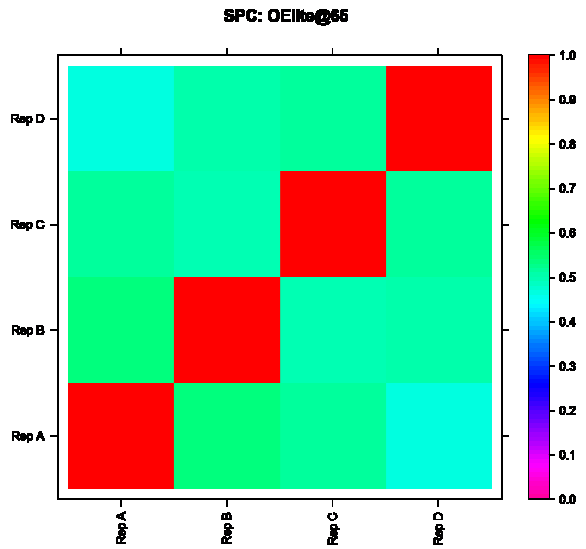


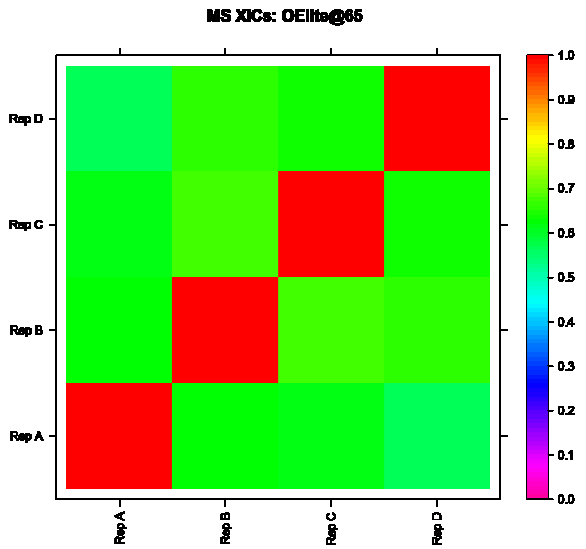


Supporting Figure 3: Rank correlation plots for repeatability

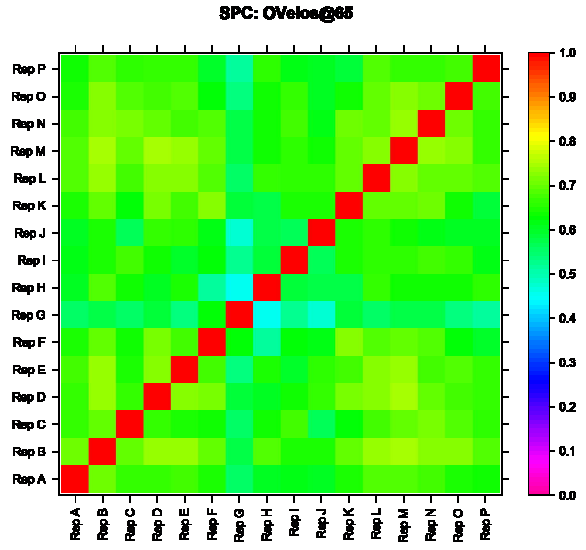


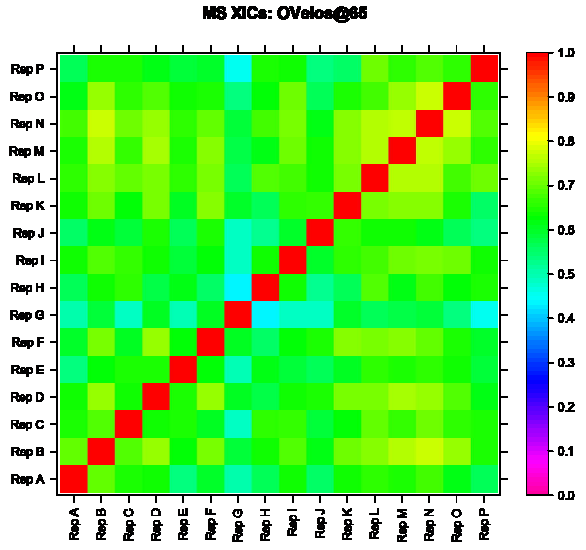
Because each iTRAQ experiment included two channels for both WHIM2 and WHIM16, four different pairs of intensities can be compared within each 4plex. These data showed considerably stronger correlation in differential posterior probabilities for comparisons within a 4plex than for comparisons outside a given 4-plex.



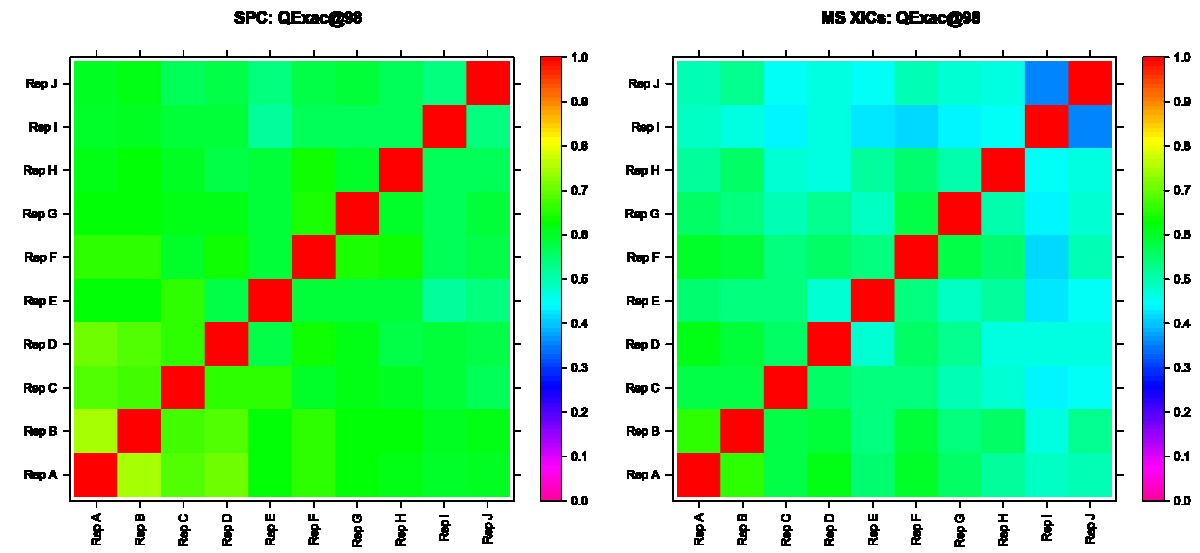


The data for OElite@65 produced better correlations for extracted ion current comparisons than for spectral count comparisons. Site 65 primarily employs a spectral count paradigm for data analysis.



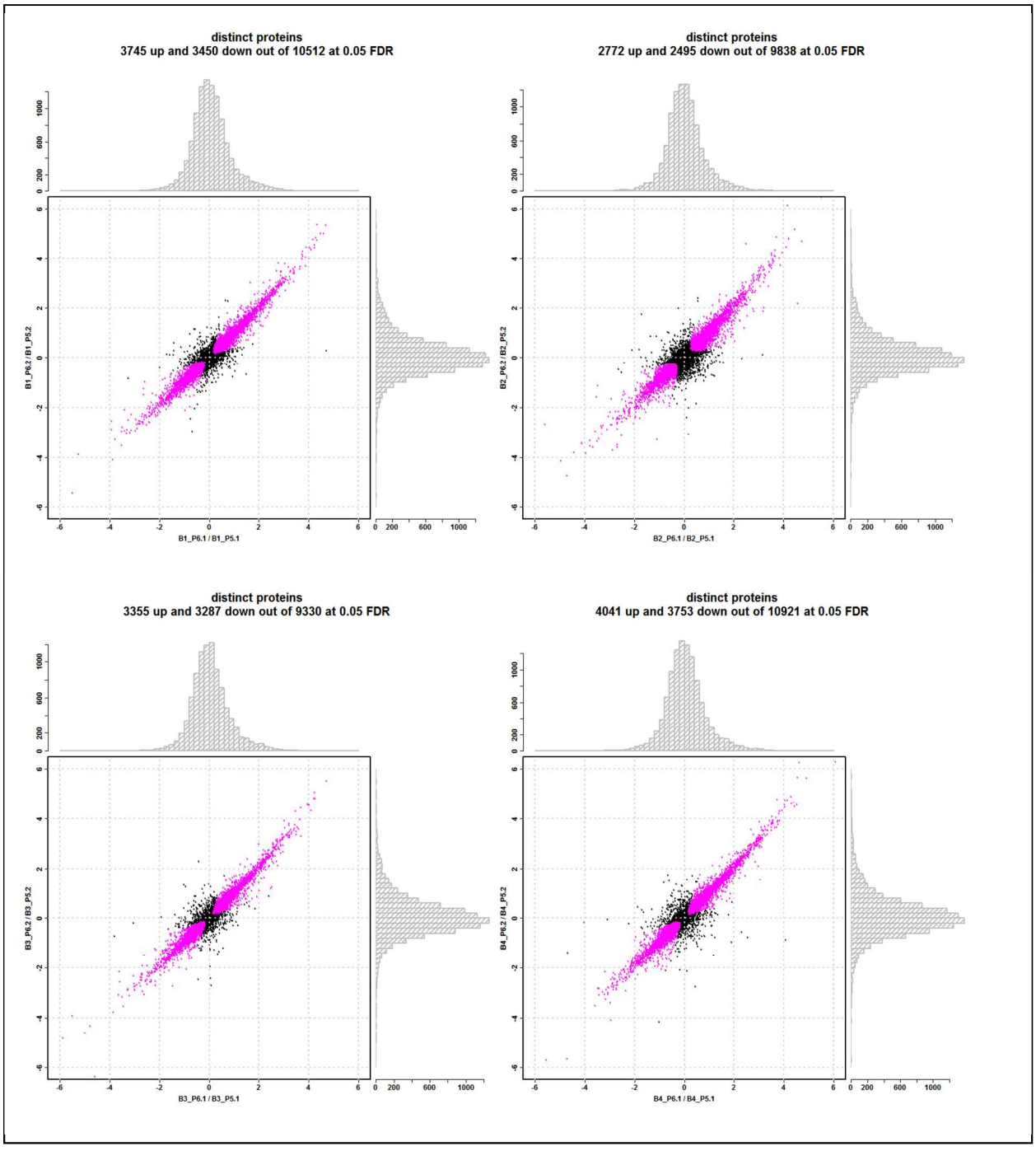


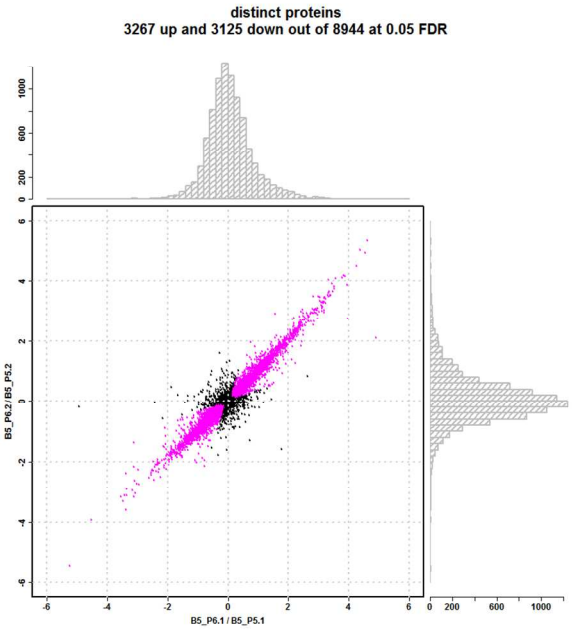
OVelos@65 data supported SPC and XIC analysis equivalently well. Data collection spanned ten months for these sixteen replicates. Replicate 'G' appeared to produce relatively low correlations with other replicates.



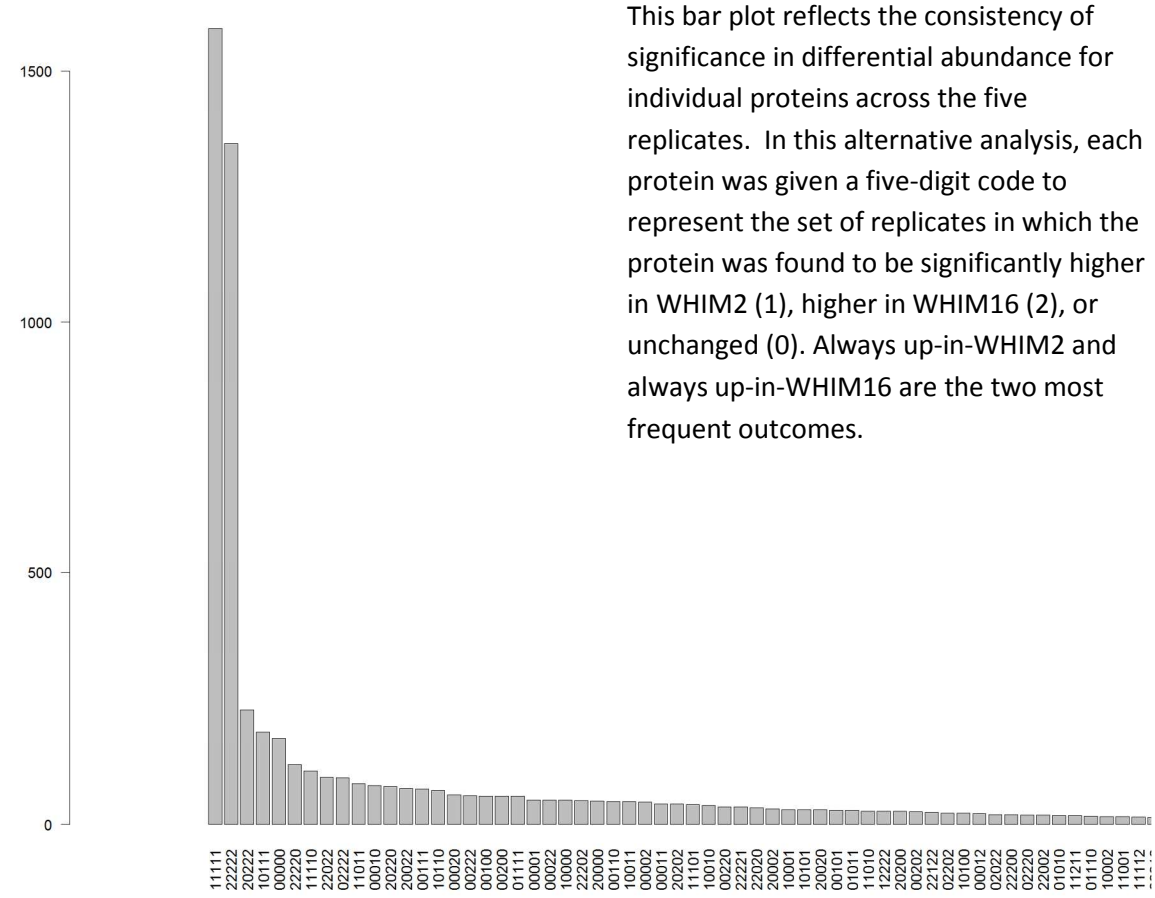
Although experimental methodology for site 98 was designed around XIC differentiation, SPC analysis appeared to produce more reliable correlation, especially in the final six replicates that suffered reduced sensitivity.

Supporting Figure 4: Alternate analysis showing consistency for OVelos@45 (iTRAQ)



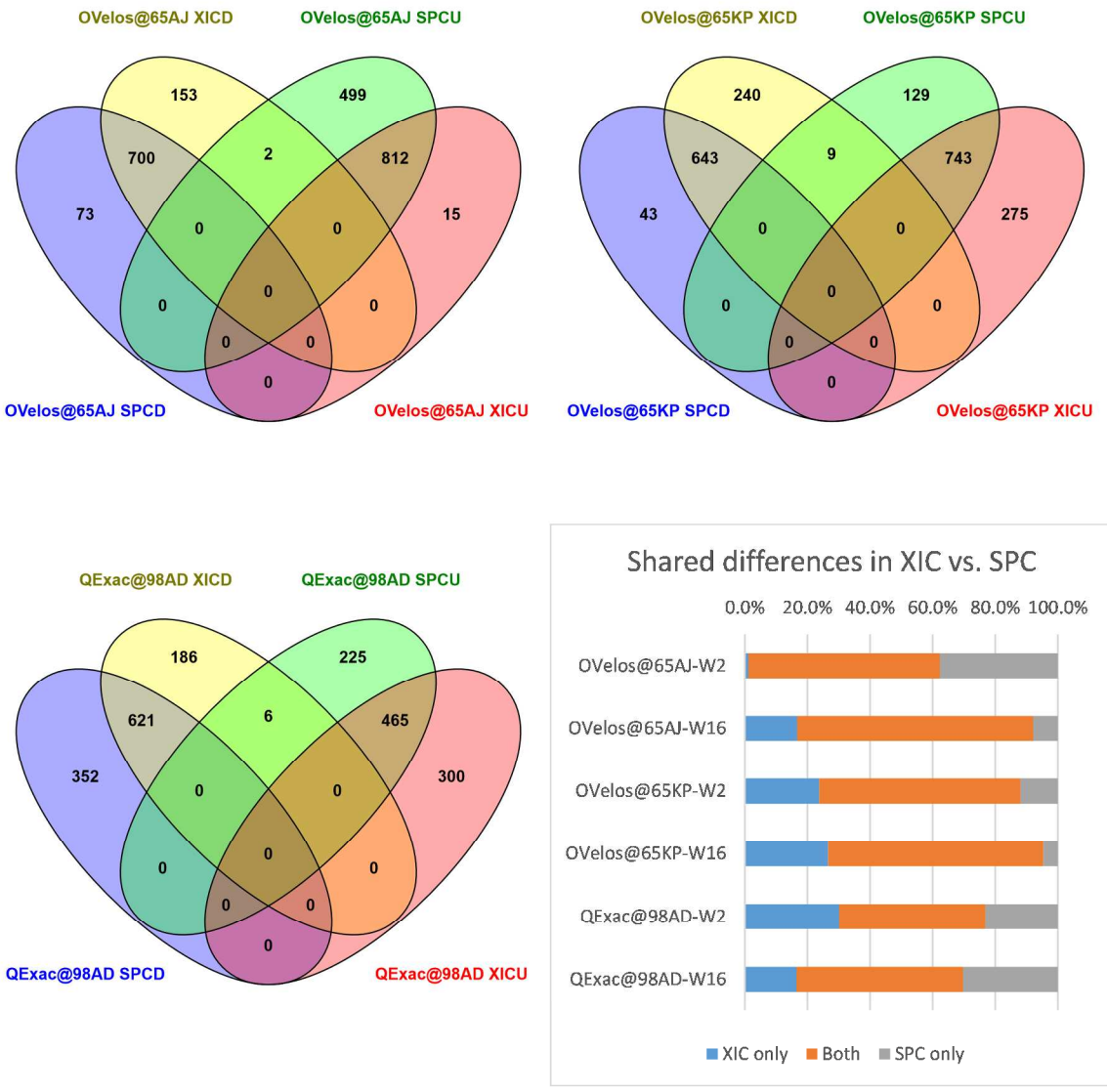


These five plots correspond to moderated paired t-test analyses of five interstitial CompRef replicates. Axes correspond to replicates of $\log_2$ ratios of WHIM2 to WHIM16 within one iTRAQ experiment. The x-axis represents the $\log_2$ ratio of channel 116 to 114, and the y-axis represents the $\log_2$ ratio of channel 117 to 115. The proteins passing the 0.05 FDR significance threshold according to Benjamini-Hochberg adjusted p-values of the moderated paired t-test are colored magenta. Overall, strong agreement between replicates was observed within individual iTRAQ experiments.



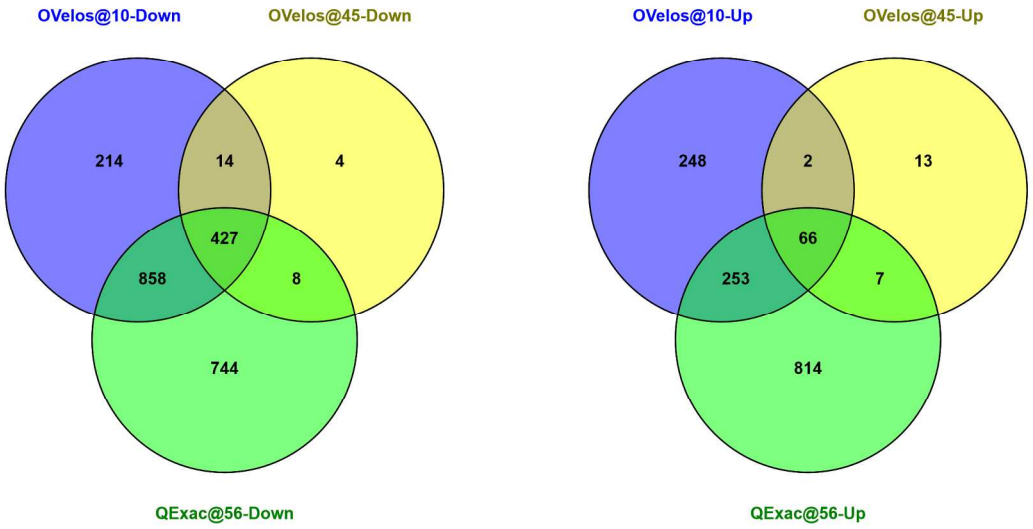
This bar plot reflects the consistency of significance in differential abundance for individual proteins across the five replicates. In this alternative analysis, each protein was given a five-digit code to represent the set of replicates in which the protein was found to be significantly higher in WHIM2 (1), higher in WHIM16 (2), or unchanged (0). Always up-in-WHIM2 and always up-in-WHIM16 are the two most frequent outcomes.

Supporting Figure 5: Comparing SPC and XIC gene differences from common data

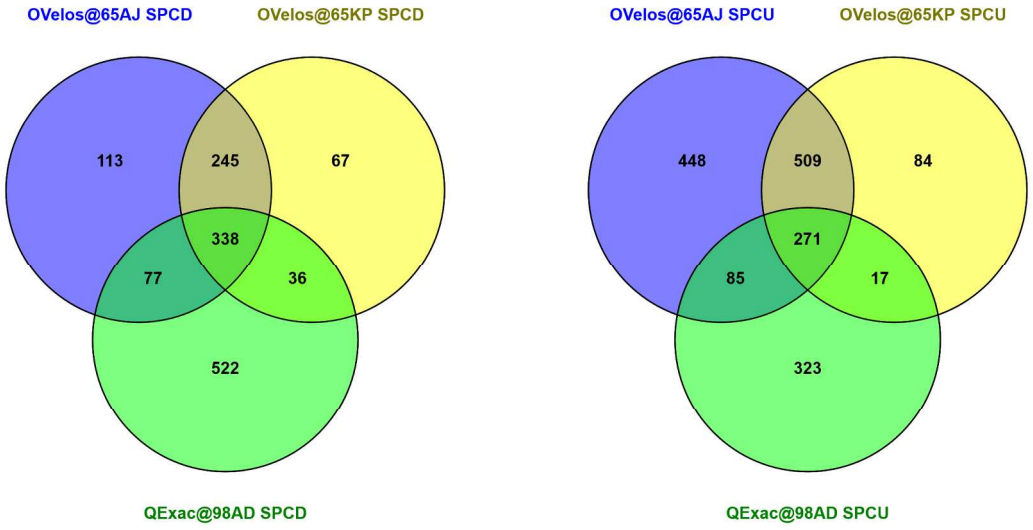


Label-free data sets may be analyzed on the basis of spectral counts or by precursor ion extracted ion chromatograms for differentiating gene expression. These plots compare the lists of genes found to significantly differ from the IDPicker assemblies with both spectral counts and intensity sums reported for each gene. The "XICD" list for an instrument includes the genes that are significantly down by XIC assessment (higher in WHIM16), while the "XICU" list includes the genes that are significantly up (higher in WHIM2). These two analysis approaches may be different enough that the resulting gene lists can be used to confirm each other.

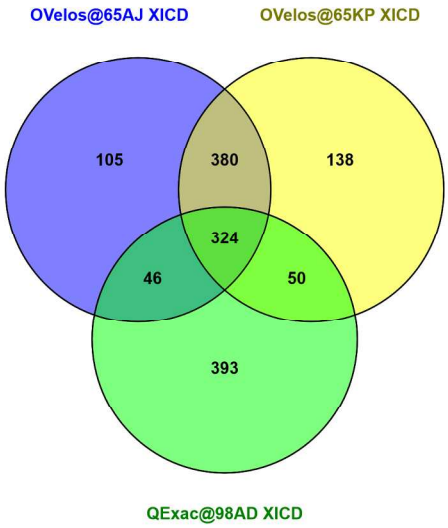
Supporting Figure 6: Comparing differential genes within technologies

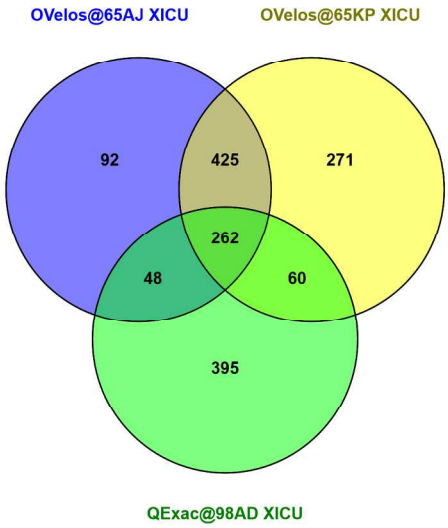


Three iTRAQ instruments took part in the study. OVelos@45 was hampered by somewhat reduced sensitivity and a decreased fraction of differential genes. QExac@56, on the other hand, featured excellent sensitivity. These Venn diagrams give a solid overlap between QExac@56 and OVelos@10. As in the prior Supporting Figure, "up" indicates higher expression in WHIM2, while "down" reflects higher expression in WHIM16.



Among label-free instruments, SPC analysis produced a strong correspondence between early and late sets from OVelos@65, and a strong performance from the four initial replicates on QExac@98 showed the power of sensitivity. OElite@65 was omitted for its low sensitivity.





When the same data are analyzed by XIC rather than SPC, however, the final six replicates of OVelos@65 look superior to the first ten replicates, perhaps because of the diminished repeatability seen in the sixth replicate. Although the QExac@98 methods were intended for XIC analysis, the SPC analysis produces a larger number of differential genes.

References

(1) Baladandayuthapani, V.; Ji, Y.; Talluri, R.; Nieto-Barajas, L. E.; Morris, J. S. Bayesian Random Segmentation Models to Identify Shared Copy Number Aberrations for Array CGH Data. *J. Am. Stat. Assoc.* **2010**, *105* (492), 1358–1375.

(2) Parmigiani, G.; Garrett, E. S.; Anbazhagan, R.; Gabrielson, E. A statistical framework for expression-based molecular classification in cancer. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2002**, *64* (4), 717–736.

(3) Choi, H.; Fermin, D.; Nesvizhskii, A. I. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics MCP* **2008**, *7* (12), 2373–2385.

(4) Gelman, A. *Bayesian data analysis*, Third edition.; Chapman & Hall/CRC texts in statistical science; CRC Press: Boca Raton, 2014.

(5) Newton, M. A.; Noueiry, A.; Sarkar, D.; Ahlquist, P. Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostat. Oxf. Engl.* **2004**, *5* (2), 155–176.