Micronutrient Deficiencies and the Human Plasma Nutriproteome

Ingo Ruczinski

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

September 15, 2011

Acknowledgments

Shelley Herbrich

Keith West, Bob Cole, Kerry Schulze, Parul Christian, Peter Scholl, Alain Labrique, Jim Yager, John Groopman.

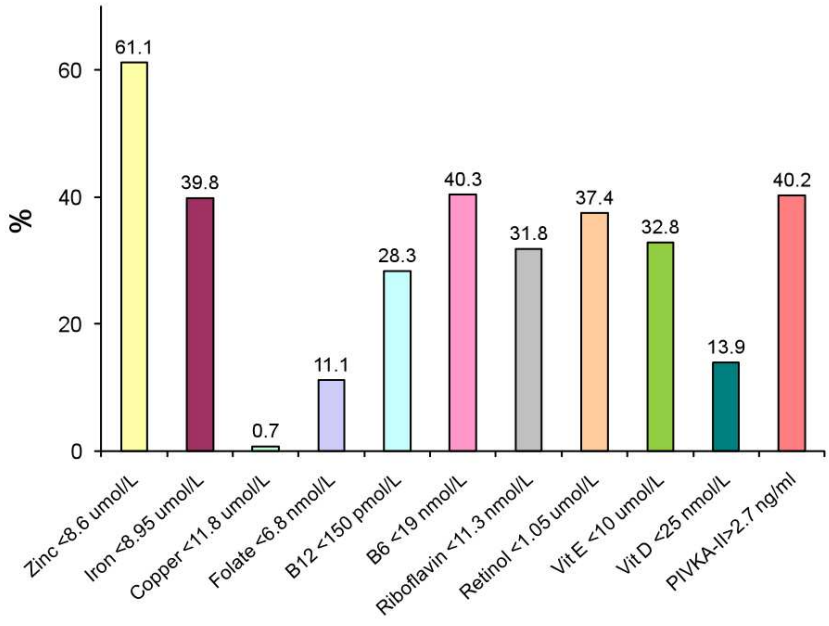
$\longrightarrow$ Funding from the Bill and Melinda Gates Foundation (#5241).

What is Hidden Hunger?

Hidden hunger is unlike the hunger that comes from a lack of food. It is a chronic lack of vitamins and minerals that often has no visible warning signs, so that people who suffer from it may not even be aware of it. Its consequences are nevertheless disastrous: hidden hunger can lead to mental impairment, poor health and productivity, or even death. One in three people in the world suffer from hidden hunger. Women and children from the lower income groups in developing countries are often the most affected.

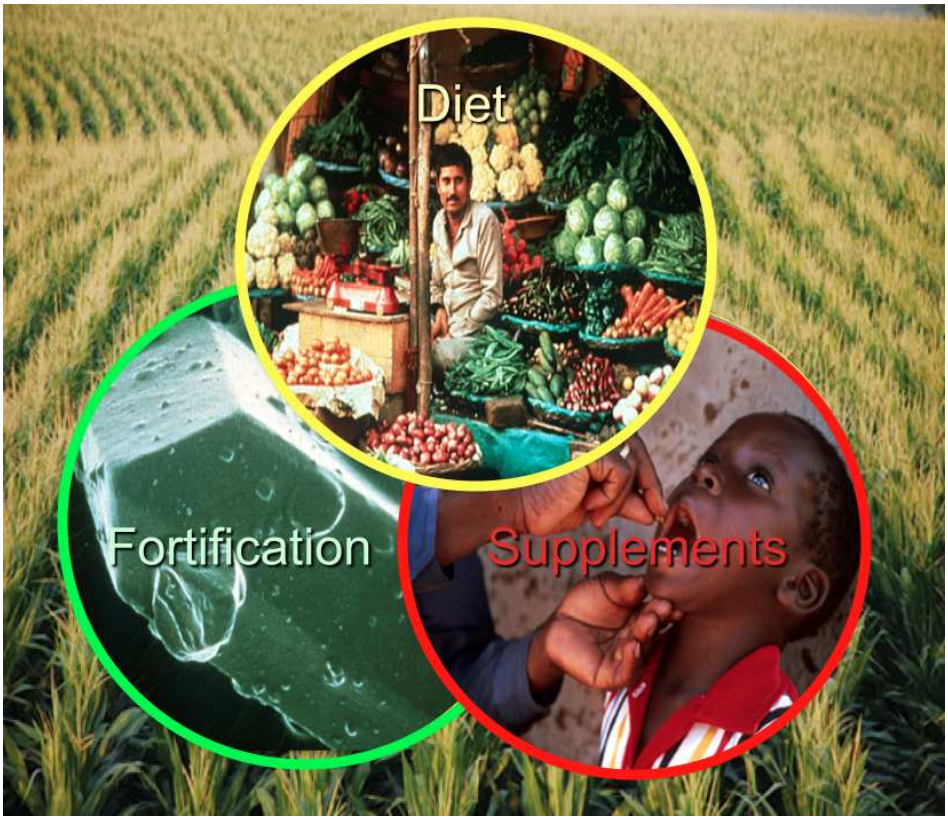
http://www.micronutrient.org/

Prevalence of low micronutrient status



Jiang et al, J Nutr (2005).

Strategies for preventing micronutrient deficiencies

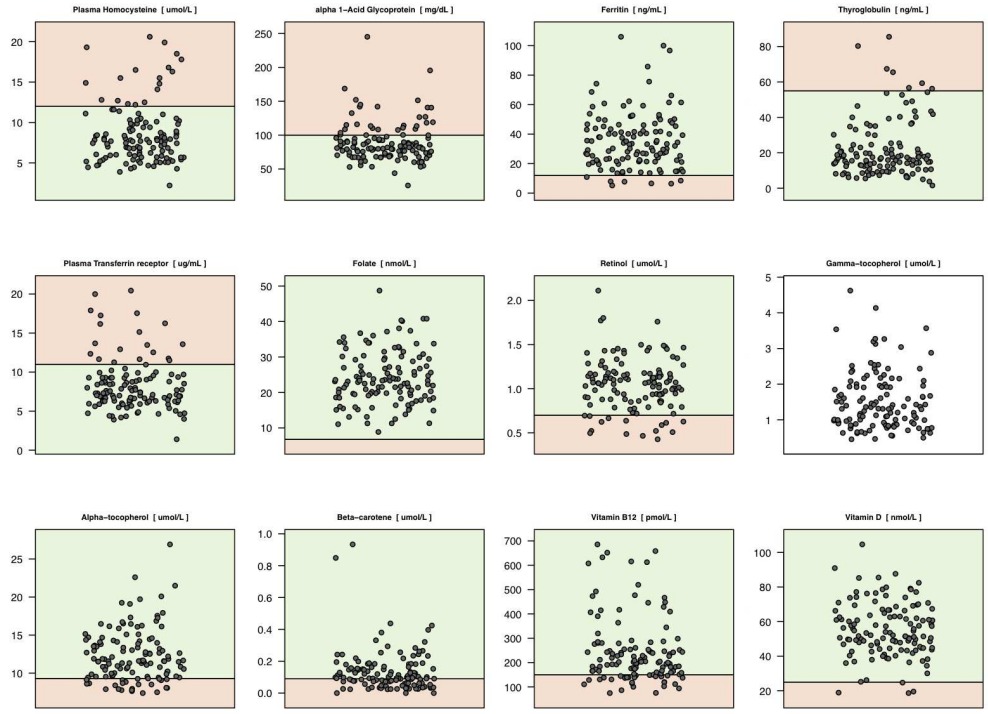


However

Lacking is an ability to routinely and quantitatively assess population status with respect to multiple essential and potentially deficient micronutrients to

- Identify highest risk populations,
- Estimate prevalence and severity,
- Monitor population trends over time,
- Inform, target and guide program development,
- Evaluate and improve programs.

Micronutrient measurements



Proteomics 2003, *3*, 1667–1672 DOI 10.1002/pmic.200300522

Short Communication

Keith A. Baggerly
Jeffrey S. Morris
Jing Wang
David Gold
Lian-Chun Xiao
Kevin R. Coombes

Department of Biostatistics,
UT M.D. Anderson
Cancer Center,
Houston, TX, USA

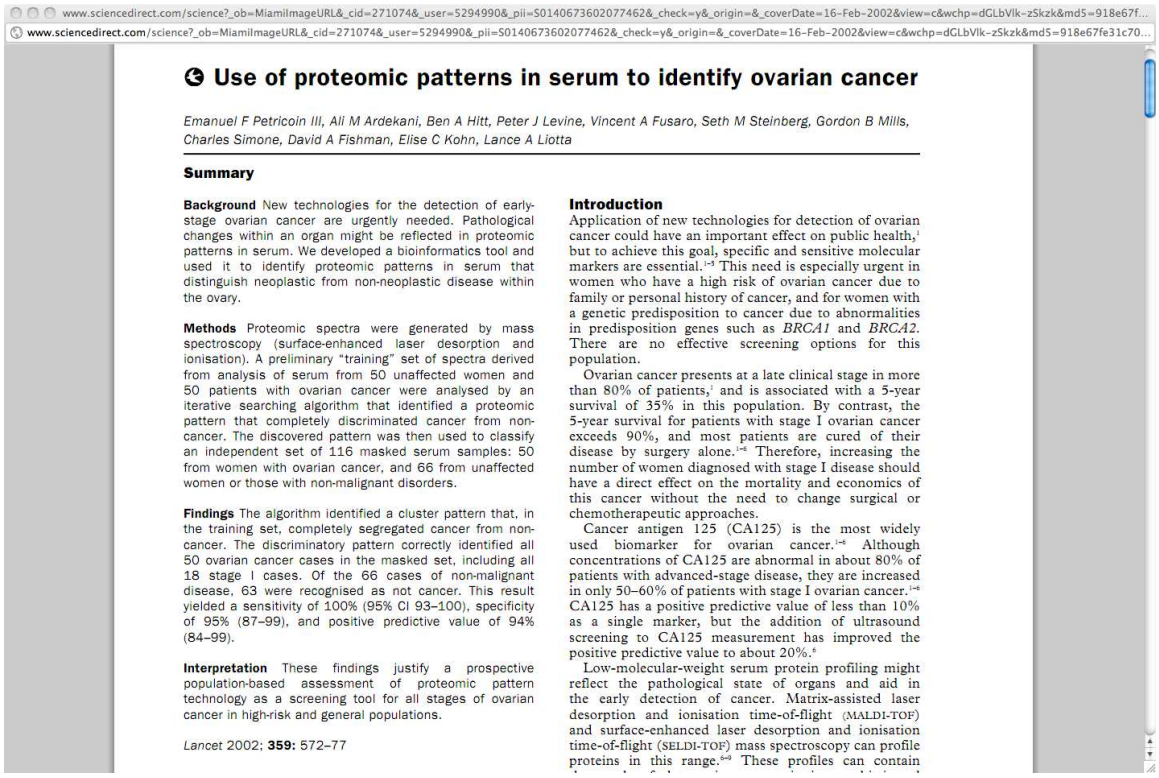
A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples

For our analysis of the data from the First Annual Proteomics Data Mining Conference, we attempted to discriminate between 24 disease spectra (group A) and 17 normal spectra (group B). First, we processed the raw spectra by (i) correcting for additive sinusoidal noise (periodic on the time scale) affecting most spectra, (ii) correcting for the overall baseline level, (iii) normalizing, (iv) recombining fractions, and (v) using variable-width windows for data reduction. Also, we identified a set of polymeric peaks (at multiples of 180.6 Da) that is present in several normal spectra (B1–B8). After data processing, we found the intensities at the following mass to charge (*m/z*) values to be useful discriminators: 3077, 12 886 and 74 263. Using these values, we were able to achieve an overall classification accuracy of 38/41 (92.6%). Perfect classification could be achieved by adding two additional peaks, at 2476 and 6955. We identified these values by applying a genetic algorithm to a filtered list of *m/z* values using Mahalanobis distance between the group means as a fitness function.

Keywords: Cross validation / Data cleaning / Discrimination / Genetic algorithm / Mahalanobis distance
PRO 0522

Sinusoidal noise removal. Visual inspection of the raw spectra revealed systematic distortions, particularly at the high *m/z* values: regular sinusoidal noise affected most of the spectra (Fig. 1). This noise was periodic on the time scale, not on the *m/z* scale. We applied a Fourier transform to several affected spectra, restricting the transform to regions where larger peaks were absent. The period of the noise (roughly 1760 clock ticks) was found to be nearly constant across different fractions and samples, but the phase appeared to be random. We suspect that this phenomenon is linked to the frequency of the alternating current in the power source, but cannot confirm this suspicion without more information. We are certain that it is not due to biology. Sinusoids of the appropriate frequency were fit to the tails of each spectrum, extended to the full spectrum length, and subtracted out. This processing is illustrated in Fig. 2.

The clock is visible in the spectra. Summing the corrected spectra uncovered an unexpected periodic phenomenon – a recurrent dip in intensity every $4096 = 2^{12}$ clock ticks. Smaller, more complicated periodicities occurred at other powers of 2. These periodicities differed from the sinusoidal noise discussed earlier. The sinusoidal noise was random in phase, and so largely canceled between spectra. Here, we were able to detect the new dip because of reinforcement across spectra. Further, this dip was uniformly present in all 41 averaged spectra. Because this phenomenon occurred at powers of 2, we strongly suspect that it is an artifact related to a computer chip inside the instrument recording the data.



www.sciencedirect.com/science?_ob=MiamiImageURL&_cid=271074&_user=5294990&_pii=S0140673602077462&_check=y&_origin=&_coverDate=16-Feb-2002&view=c&wchp=dGLbVlk-zSkzk&md5=918e67f...

www.sciencedirect.com/science?_ob=MiamiImageURL&_cid=271074&_user=5294990&_pii=S0140673602077462&_check=y&_origin=&_coverDate=16-Feb-2002&view=c&wchp=dGLbVlk-zSkzk&md5=918e67fe31c70...

Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

Summary

Background New technologies for the detection of early-stage ovarian cancer are urgently needed. Pathological changes within an organ might be reflected in proteomic patterns in serum. We developed a bioinformatics tool and used it to identify proteomic patterns in serum that distinguish neoplastic from non-neoplastic disease within the ovary.

Methods Proteomic spectra were generated by mass spectroscopy (surface-enhanced laser desorption and ionisation). A preliminary "training" set of spectra derived from analysis of serum from 50 unaffected women and 50 patients with ovarian cancer were analysed by an iterative searching algorithm that identified a proteomic pattern that completely discriminated cancer from non-cancer. The discovered pattern was then used to classify an independent set of 116 masked serum samples: 50 from women with ovarian cancer, and 66 from unaffected women or those with non-malignant disorders.

Findings The algorithm identified a cluster pattern that, in the training set, completely segregated cancer from non-cancer. The discriminatory pattern correctly identified all 50 ovarian cancer cases in the masked set, including all 18 stage I cases. Of the 66 cases of non-malignant disease, 63 were recognised as not cancer. This result yielded a sensitivity of 100% (95% CI 93–100), specificity of 95% (87–99), and positive predictive value of 94% (84–99).

Interpretation These findings justify a prospective population-based assessment of proteomic pattern technology as a screening tool for all stages of ovarian cancer in high-risk and general populations.

Lancet 2002; **359:** 572–77

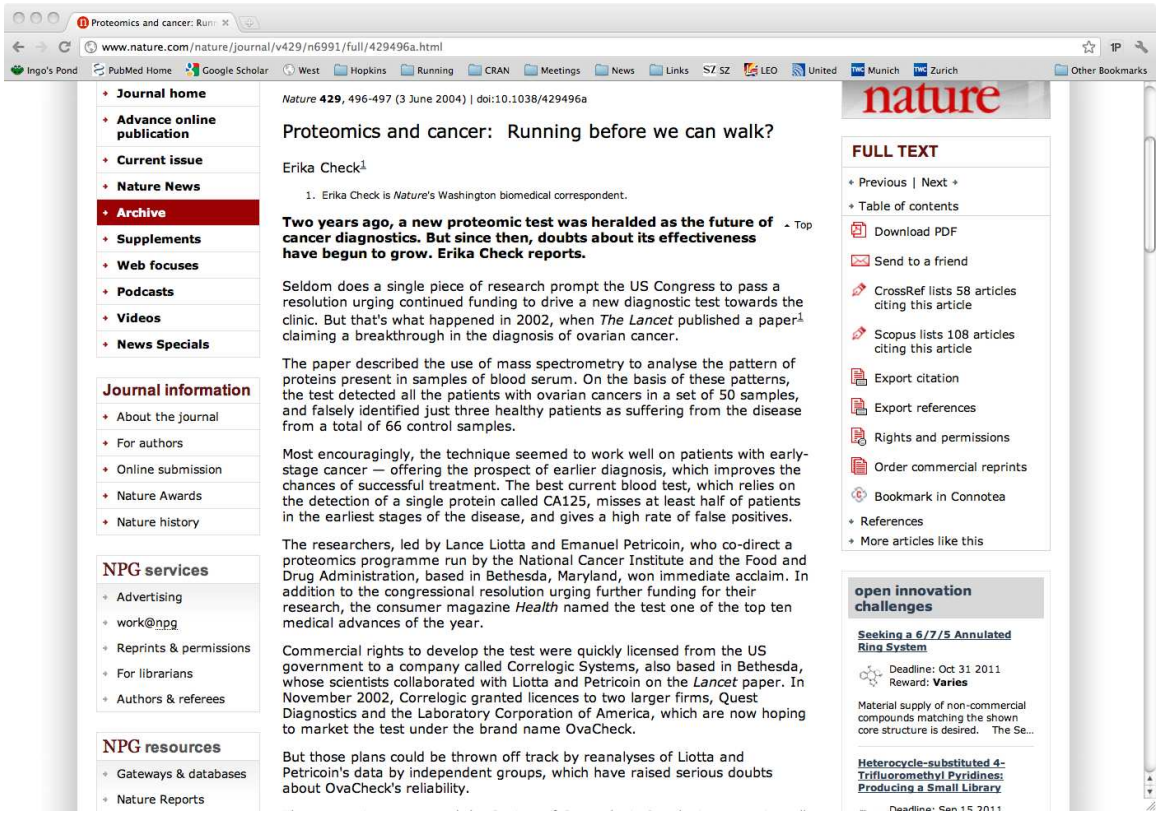
Introduction

Application of new technologies for detection of ovarian cancer could have an important effect on public health,[1] but to achieve this goal, specific and sensitive molecular markers are essential.[1–3] This need is especially urgent in women who have a high risk of ovarian cancer due to family or personal history of cancer, and for women with a genetic predisposition to cancer due to abnormalities in predisposition genes such as *BRCA1* and *BRCA2*. There are no effective screening options for this population.

Ovarian cancer presents at a late clinical stage in more than 80% of patients,[1] and is associated with a 5-year survival of 35% in this population. By contrast, the 5-year survival for patients with stage I ovarian cancer exceeds 90%, and most patients are cured of their disease by surgery alone.[1–4] Therefore, increasing the number of women diagnosed with stage I disease should have a direct effect on the mortality and economics of this cancer without the need to change surgical or chemotherapeutic approaches.

Cancer antigen 125 (CA125) is the most widely used biomarker for ovarian cancer.[1–4] Although concentrations of CA125 are abnormal in about 80% of patients with advanced-stage disease, they are increased in only 50–60% of patients with stage I ovarian cancer.[1–4] CA125 has a positive predictive value of less than 10% as a single marker, but the addition of ultrasound screening to CA125 measurement has improved the positive predictive value to about 20%.[4]

Low-molecular-weight serum protein profiling might reflect the pathological state of organs and aid in the early detection of cancer. Matrix-assisted laser desorption and ionisation time-of-flight (MALDI-TOF) and surface-enhanced laser desorption and ionisation time-of-flight (SELDI-TOF) mass spectroscopy can profile proteins in this range.[5–7] These profiles can contain



Proteomics and cancer: Running before we can walk?

www.nature.com/nature/journal/v429/n6991/full/429496a.html

Ingo's Pond | PubMed Home | Google Scholar | West | Hopkins | Running | CRAN | Meetings | News | Links | SZ SZ | LEO | United | Munich | Zurich | Other Bookmarks

- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements
- Web focuses
- Podcasts
- Videos
- News Specials

Journal information

- About the journal
- For authors
- Online submission
- Nature Awards
- Nature history

NPG services

- Advertising
- work@npg
- Reprints & permissions
- For librarians
- Authors & referees

NPG resources

- Gateways & databases
- Nature Reports

Nature **429**, 496-497 (3 June 2004) | doi:10.1038/429496a

Proteomics and cancer: Running before we can walk?

Erika Check[1]

1. Erika Check is *Nature*'s Washington biomedical correspondent.

Two years ago, a new proteomic test was heralded as the future of cancer diagnostics. But since then, doubts about its effectiveness have begun to grow. Erika Check reports.

Top

Seldom does a single piece of research prompt the US Congress to pass a resolution urging continued funding to drive a new diagnostic test towards the clinic. But that's what happened in 2002, when *The Lancet* published a paper[1] claiming a breakthrough in the diagnosis of ovarian cancer.

The paper described the use of mass spectrometry to analyse the pattern of proteins present in samples of blood serum. On the basis of these patterns, the test detected all the patients with ovarian cancers in a set of 50 samples, and falsely identified just three healthy patients as suffering from the disease from a total of 66 control samples.

Most encouragingly, the technique seemed to work well on patients with early-stage cancer — offering the prospect of earlier diagnosis, which improves the chances of successful treatment. The best current blood test, which relies on the detection of a single protein called CA125, misses at least half of patients in the earliest stages of the disease, and gives a high rate of false positives.

The researchers, led by Lance Liotta and Emanuel Petricoin, who co-direct a proteomics programme run by the National Cancer Institute and the Food and Drug Administration, based in Bethesda, Maryland, won immediate acclaim. In addition to the congressional resolution urging further funding for their research, the consumer magazine *Health* named the test one of the top ten medical advances of the year.

Commercial rights to develop the test were quickly licensed from the US government to a company called Correlogic Systems, also based in Bethesda, whose scientists collaborated with Liotta and Petricoin on the *Lancet* paper. In November 2002, Correlogic granted licences to two larger firms, Quest Diagnostics and the Laboratory Corporation of America, which are now hoping to market the test under the brand name OvaCheck.

But those plans could be thrown off track by reanalyses of Liotta and Petricoin's data by independent groups, which have raised serious doubts about OvaCheck's reliability.

nature

FULL TEXT

- Previous | Next
- Table of contents
- Download PDF
- Send to a friend
- CrossRef lists 58 articles citing this article
- Scopus lists 108 articles citing this article
- Export citation
- Export references
- Rights and permissions
- Order commercial reprints
- Bookmark in Connotea
- References
- More articles like this

open innovation challenges

Seeking a 6/7/5 Annulated Ring System

Deadline: Oct 31 2011
Reward: **Varies**

Material supply of non-commercial compounds matching the shown core structure is desired. The Se...

Heterocycle-substituted 4-Trifluoromethyl Pyridines: Producing a Small Library

Deadline: Sep 15 2011

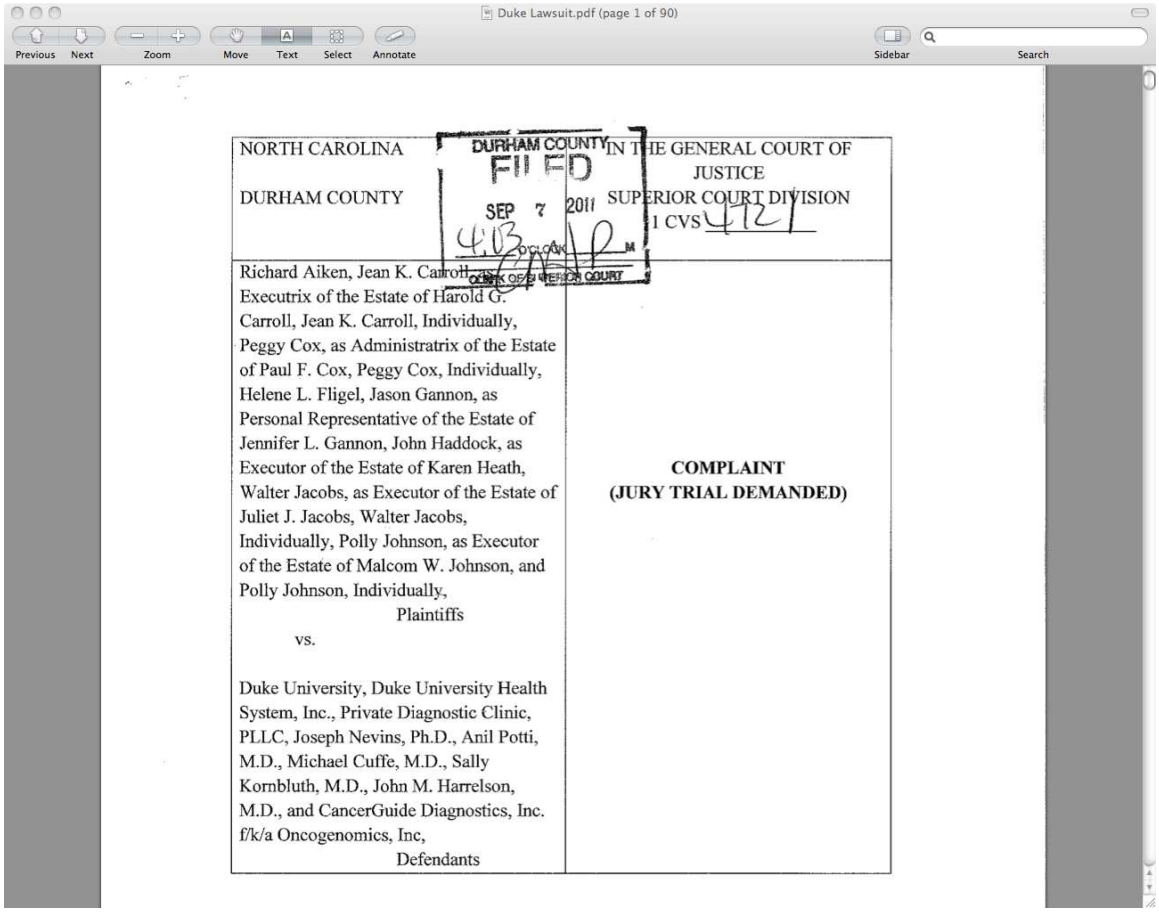
DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

By Keith A. Baggerly* and Kevin R. Coombes†

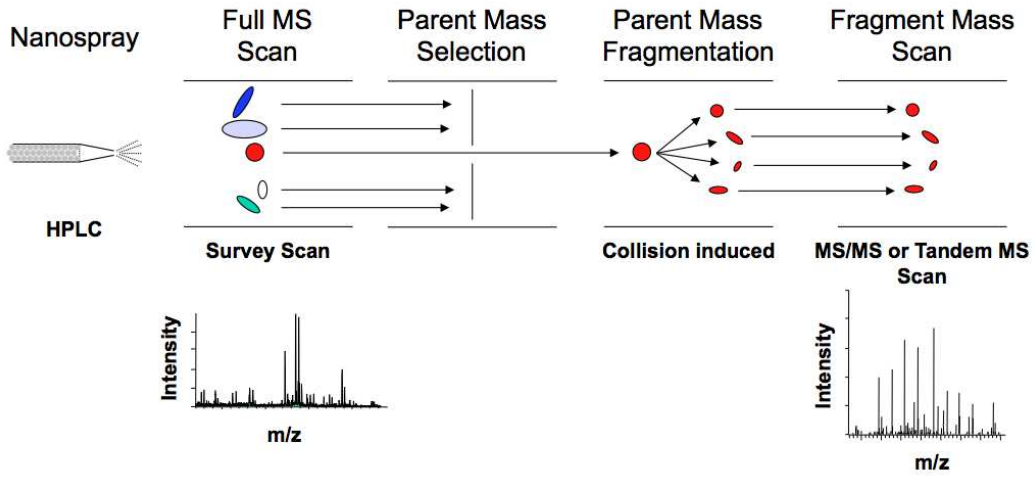
U.T. M.D. Anderson Cancer Center

High-throughput biological assays such as microarrays let us ask very detailed questions about how diseases operate, and promise to let us personalize therapy. Data processing, however, is often not described well enough to allow for exact reproduction of the results, leading to exercises in "forensic bioinformatics" where aspects of raw data and reported results are used to infer what methods must have been employed. Unfortunately, poor documentation can shift from an inconvenience to an active danger when it obscures not just methods but errors. In this report, we examine several related papers purporting to use microarray-based signatures of drug sensitivity derived from cell lines to predict patient response. Patients in clinical trials are currently being allocated to treatment arms on the basis of these results. However, we show in five case studies that the results incorporate several simple errors that may be putting patients at risk. One theme that emerges is that the most common errors are simple (e.g., row or column offsets); conversely, it is our experience that the most simple errors are common. We then discuss steps we are taking to avoid such errors in our own investigations.

Annals of Applied Statistics

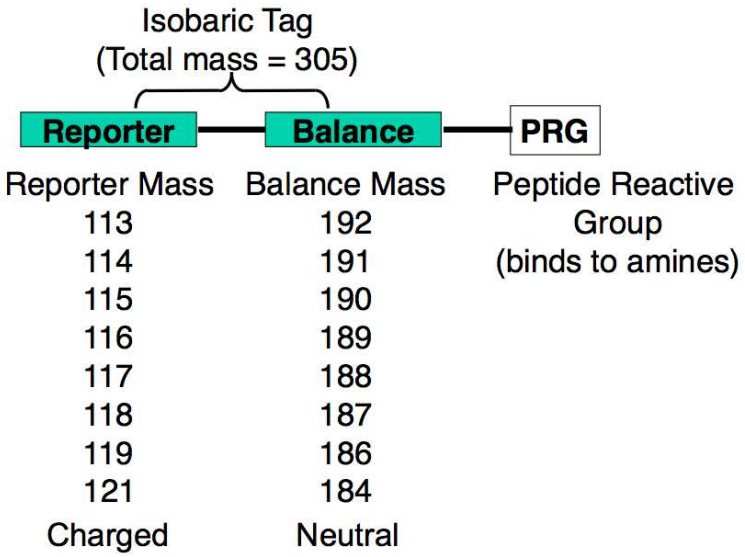


Tandem mass spectrometry



iTRAQ

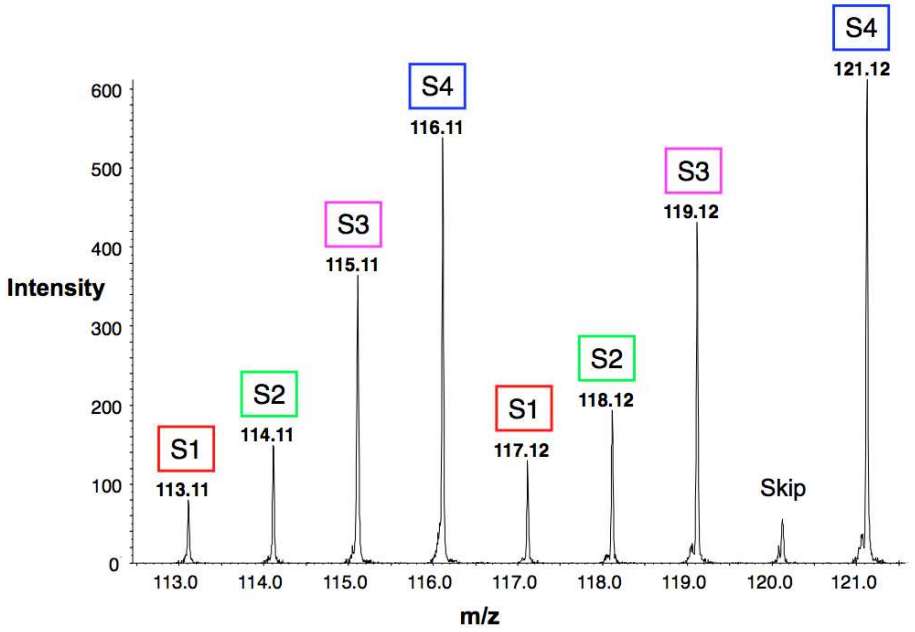
Isobaric Tag for Relative and Absolute Quantitation



Reporter Mass	Balance Mass	Peptide Reactive Group (binds to amines)
113	192	
114	191	
115	190	
116	189	
117	188	
118	187	
119	186	
121	184	
Charged	Neutral	

iTRAQ

Isobaric Tag for Relative and Absolute Quantitation



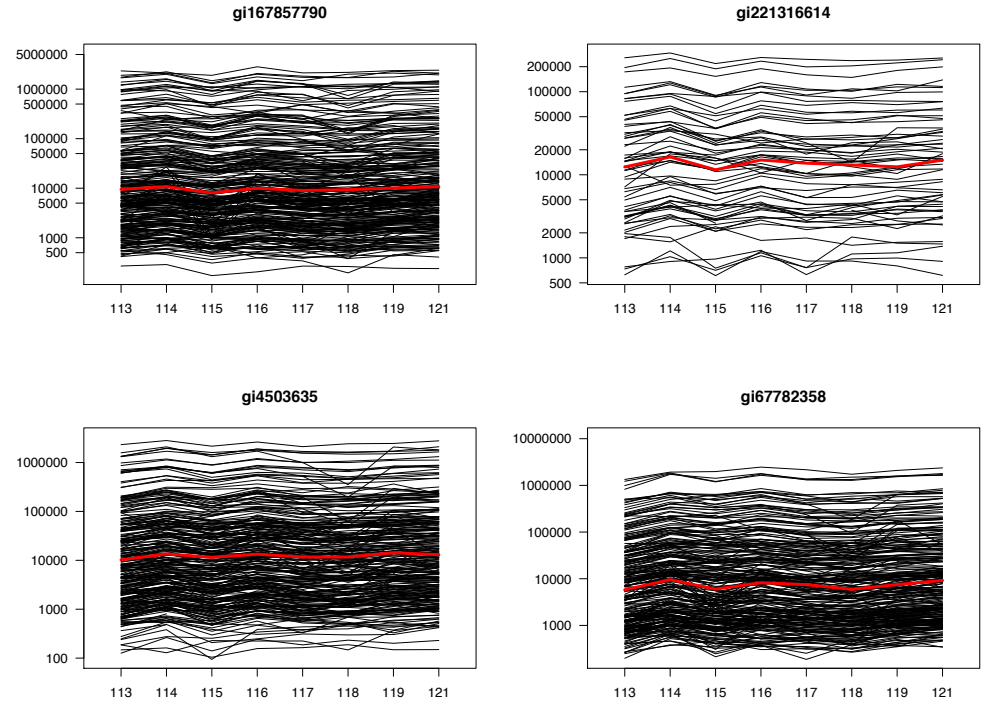
Spectra files

Sequence	Protein Access	Quan Usage	Quan Info	114/113	115/113	116/113	117/113	118/113	119/113	121/113
sLLEDIR	gi9955968	Used	Unique	1.38	1.053	1.204	0.863	0.896	0.993	1.234
sLLEDIR	gi9955968	Used	Unique	1.691	1.286	1.27	1.117	1.102	1.092	1.487
sLLEDIR	gi9955968	Used	Unique	1.191	0.988	1.16	1.061	1.088	1.054	1.385
sLLEDIR	gi9955968	Used	Unique	1.59	1.044	1.156	1.043	1.096	0.988	1.248
sLLEDIR	gi9955968	Used	Unique	1.899	1.476	1.632	1.399	1.246	1.263	1.716
sLLEDIR	gi9955968	Used	Unique	1.579	1.373	1.507	1.148	1.134	1.139	1.648
sLLEDIR	gi9955968	Used	Unique	1.933	1.317	1.442	1.264	1.354	1.295	1.611
sLLEDIR	gi9955968	Used	Unique	1.829	1.616	1.673	1.361	1.344	1.208	1.796
sLLEDIR	gi9955968	Used	Unique	1.69	1.311	1.6	1.17	1.177	1.181	1.664
sLLEDIR	gi9955968	Used	Unique	1.5	1.104	1.167	1.014	0.897	1.009	1.282
sLLEDIR	gi9955968	Used	Unique	1.679	1.315	1.481	1.159	1.056	1.214	1.578
eDRAPLLk	gi9845507	Used	Unique	1.316	0.928	1.385	1.176	0.946	1.213	1.387
eDRAPLLk	gi9845507	Used	Unique	1.688	1.258	1.633	1.499	1.352	1.501	1.395
eDRAPLLk	gi9845507	Not Used	Redundant	1.651	1.138	1.342	1.397	1.451	1.333	1.481
eDRAPLLk	gi9845507	Not Used	Redundant	1.532	2.702	4.066	3.167	2.319	3.346	3.938
eDRAPLLk	gi9845507	Not Used	Redundant	1.606	1.247	1.69	1.349	0.745	1.38	1.48
eDRAPLLk	gi9845507	Not Used	Redundant	1.891	1.13	1.847	1.357	0.98	1.135	1.56
eDRAPLLk	gi9845507	Not Used	Redundant	1.768	1.198	1.798	1.534	1.035	1.636	1.754
eDRAPLLk	gi9845507	Not Used	Redundant	2.108	2.653	4.883	3.333	2.936	3.291	4.065
eDRAPLLk	gi9845507	Not Used	Redundant	1.723	1.332	1.983	1.605	1.179	1.552	1.76
eDRAPLLk	gi9845507	Not Used	Redundant	1.947	1.409	2.456	2.055	1.513	1.692	2.222
eDRAPLLk	gi9845507	Not Used	Redundant	2.503	1.865	2.536	2.196	1.178	2.468	2.354
eDRAPLLk	gi9845507	Not Used	Redundant	4.45	0.334	4.475	1.936	0.989	1.966	3.576
eDRAPLLk	gi9845507	Not Used	Redundant	1.115	1.004	1.687	1.175	0.887	1.803	1.714
eDRAPLLk	gi9845507	Not Used	Redundant	3.506	0.297	2.309	1.329	0.897	1.515	2.143
eDRAPLLk	gi9845507	Not Used	Redundant	1.876	1.219	1.796	1.572	1.114	1.526	1.722
fkDFLLYLR	gi9845238	Used	Unique	0.936	1.307	1.337	1.332	1.47	1.358	1.454
fkDFLLYLR	gi9845238	Used	Unique	1.04	1.46	1.564	1.6	1.644	1.483	1.692
iLGYTLk	gi9789987	Not Used	Redundant	1.492	1.343	1.558	1.102	1.43	1.041	1.196
iDVLk	gi94818781	Not Used	Redundant	1.718	1.417	1.581	1.236	1.242	1.267	1.631
gQDFEk	gi94818781	Not Used	Redundant	1.595	1.024	1.306	1.071	1.039	1.238	1.519
eGQAVAVPSSk	gi94721327	Used	Unique	1.206	1.052	1.248	1.069	1.189	1.195	1.347
eGQAVAVPSSk	gi94721327	Used	Unique	1.242	1.146	1.208	1.055	1.051	1.173	1.235
eGQAVAVPSSk	gi94721327	Used	Unique	1.298	1.026	1.202	1.011	1.213	1.213	1.362

Spectra files

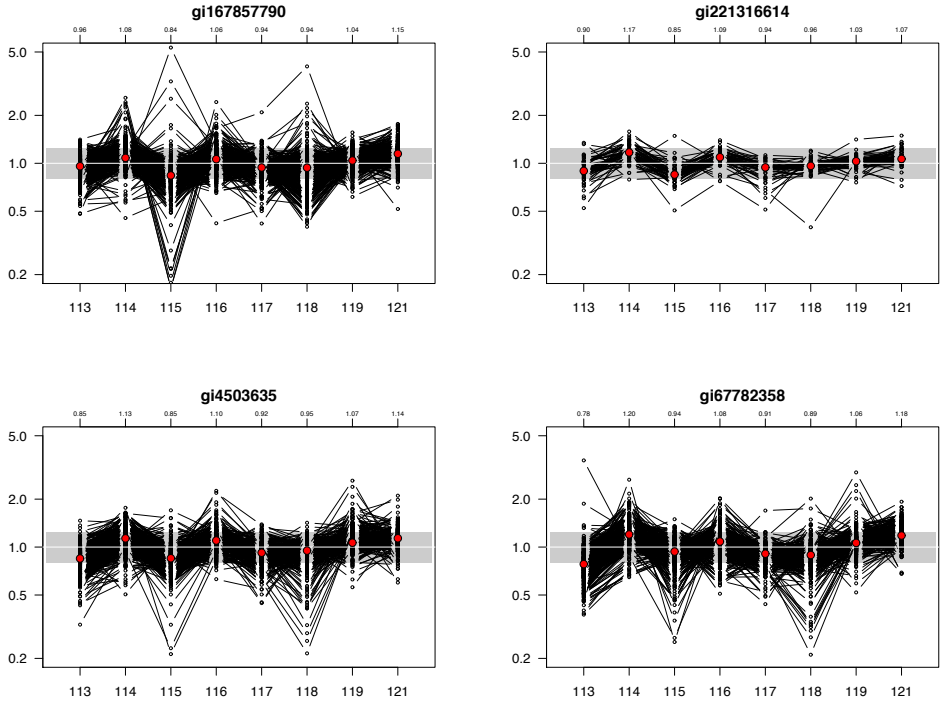
Sequence	Protein Access	Quan Usage	Quan Info	113	114	115	116	117	118	119
sLLEDIR	gi9955968	Used	Unique	2.79E+04	3.85E+04	2.94E+04	3.36E+04	2.41E+04	2.50E+04	2.77E+04
sLLEDIR	gi9955968	Used	Unique	2.96E+05	5.00E+05	3.81E+05	3.76E+05	3.31E+05	3.26E+05	3.23E+05
sLLEDIR	gi9955968	Used	Unique	2.46E+04	2.93E+04	2.43E+04	2.86E+04	2.61E+04	2.68E+04	2.59E+04
sLLEDIR	gi9955968	Used	Unique	3.88E+03	6.16E+03	4.05E+03	4.48E+03	4.04E+03	4.25E+03	3.83E+03
sLLEDIR	gi9955968	Used	Unique	6.21E+05	1.18E+06	9.17E+05	1.01E+06	8.69E+05	7.74E+05	7.85E+05
sLLEDIR	gi9955968	Used	Unique	5.82E+05	9.19E+05	7.99E+05	8.77E+05	6.68E+05	6.60E+05	6.63E+05
sLLEDIR	gi9955968	Used	Unique	4.88E+04	9.43E+04	6.42E+04	7.03E+04	6.16E+04	6.60E+04	6.31E+04
sLLEDIR	gi9955968	Used	Unique	4.52E+05	8.27E+05	7.30E+05	7.56E+05	6.15E+05	6.07E+05	5.46E+05
sLLEDIR	gi9955968	Used	Unique	1.45E+05	2.46E+05	1.91E+05	2.33E+05	1.70E+05	1.71E+05	1.72E+05
sLLEDIR	gi9955968	Used	Unique	8.90E+04	1.34E+05	9.83E+04	1.04E+05	9.02E+04	7.98E+04	8.98E+04
sLLEDIR	gi9955968	Used	Unique	2.37E+05	3.98E+05	3.12E+05	3.51E+05	2.75E+05	2.50E+05	2.88E+05
eDRAPLLk	gi9845507	Used	Unique	8.49E+02	1.12E+03	7.88E+02	1.18E+03	9.99E+02	8.03E+02	1.03E+03
eDRAPLLk	gi9845507	Used	Unique	8.87E+02	1.50E+03	1.12E+03	1.45E+03	1.33E+03	1.20E+03	1.33E+03
eDRAPLLk	gi9845507	Not Used	Redundant	7.68E+03	1.27E+04	8.73E+03	1.03E+04	1.07E+04	1.11E+04	1.02E+04
eDRAPLLk	gi9845507	Not Used	Redundant	1.83E+05	2.80E+05	4.94E+05	7.43E+05	5.79E+05	4.24E+05	6.12E+05
eDRAPLLk	gi9845507	Not Used	Redundant	4.80E+05	7.70E+05	5.98E+05	8.11E+05	6.47E+05	3.57E+05	6.62E+05
eDRAPLLk	gi9845507	Not Used	Redundant	3.58E+04	6.77E+04	4.05E+04	6.62E+04	4.86E+04	3.51E+04	4.07E+04
eDRAPLLk	gi9845507	Not Used	Redundant	2.87E+05	5.07E+05	3.44E+05	5.16E+05	4.40E+05	2.97E+05	4.69E+05
eDRAPLLk	gi9845507	Not Used	Redundant	4.65E+04	9.80E+04	1.23E+05	2.27E+05	1.55E+05	1.36E+05	1.53E+05
eDRAPLLk	gi9845507	Not Used	Redundant	9.58E+06	1.65E+07	1.28E+07	1.90E+07	1.54E+07	1.13E+07	1.49E+07
eDRAPLLk	gi9845507	Not Used	Redundant	3.40E+06	6.61E+06	4.78E+06	8.34E+06	6.98E+06	5.14E+06	5.75E+06
eDRAPLLk	gi9845507	Not Used	Redundant	4.14E+05	1.04E+06	7.72E+05	1.05E+06	9.09E+05	4.87E+05	1.02E+06
eDRAPLLk	gi9845507	Not Used	Redundant	2.03E+04	9.02E+04	6.78E+03	9.07E+04	3.93E+04	2.01E+04	3.99E+04
eDRAPLLk	gi9845507	Not Used	Redundant	6.29E+03	7.02E+03	6.31E+03	1.06E+04	7.39E+03	5.58E+03	1.13E+04
eDRAPLLk	gi9845507	Not Used	Redundant	1.46E+04	5.12E+04	4.34E+03	3.37E+04	1.94E+04	1.31E+04	2.21E+04
eDRAPLLk	gi9845507	Not Used	Redundant	1.92E+04	3.60E+04	2.34E+04	3.45E+04	3.02E+04	2.14E+04	2.93E+04
fkDFLLYLR	gi9845238	Used	Unique	2.29E+03	2.15E+03	3.00E+03	3.07E+03	3.05E+03	3.37E+03	3.12E+03
fkDFLLYLR	gi9845238	Used	Unique	2.32E+03	2.41E+03	3.39E+03	3.63E+03	3.71E+03	3.82E+03	3.44E+03
iLGYTLk	gi9789987	Not Used	Redundant	2.29E+05	3.42E+05	3.08E+05	3.57E+05	2.53E+05	3.28E+05	2.39E+05
iDVLk	gi94818781	Not Used	Redundant	3.88E+05	6.67E+05	5.50E+05	6.14E+05	4.80E+05	4.82E+05	4.92E+05
gQDFEk	gi94818781	Not Used	Redundant	3.35E+04	5.34E+04	3.43E+04	4.38E+04	3.59E+04	3.48E+04	4.15E+04
eGQAVAVPSSk	gi94721327	Used	Unique	6.69E+04	8.07E+04	7.04E+04	8.35E+04	7.15E+04	7.95E+04	7.99E+04
eGQAVAVPSSk	gi94721327	Used	Unique	1.73E+05	2.15E+05	1.98E+05	2.09E+05	1.83E+05	1.82E+05	2.03E+05
eGQAVAVPSSk	gi94721327	Used	Unique	2.74E+05	3.55E+05	2.81E+05	3.29E+05	2.77E+05	3.32E+05	3.32E+05

Reporter ion intensities



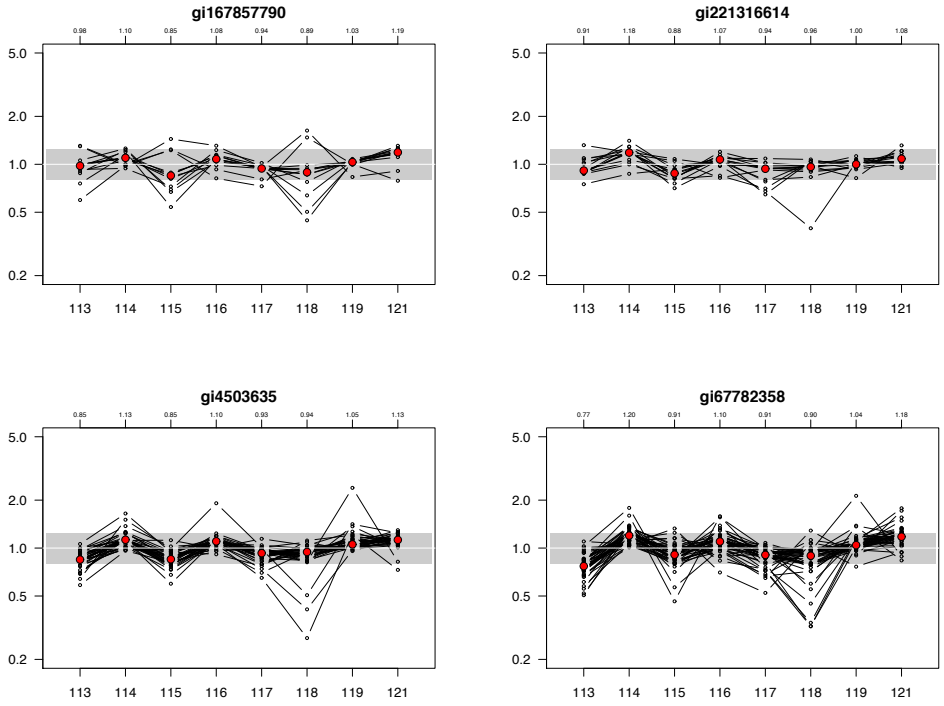
Run with 8 technical replicates of a master pool

Log reporter ion intensities (median polish)



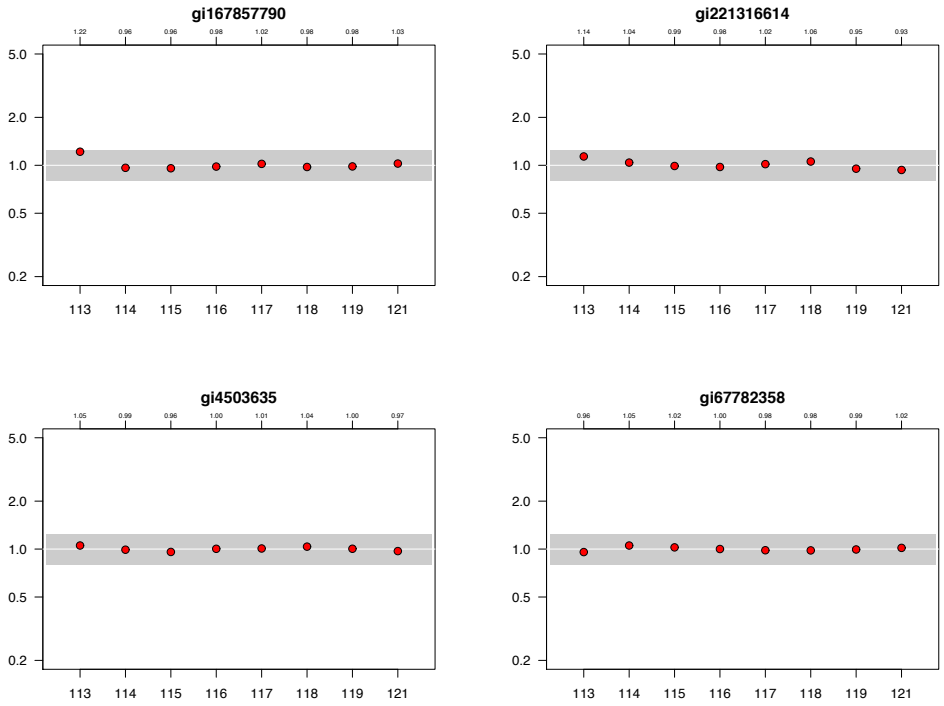
Run with 8 technical replicates of a master pool

Log reporter ion intensities (median polish, by peptide)



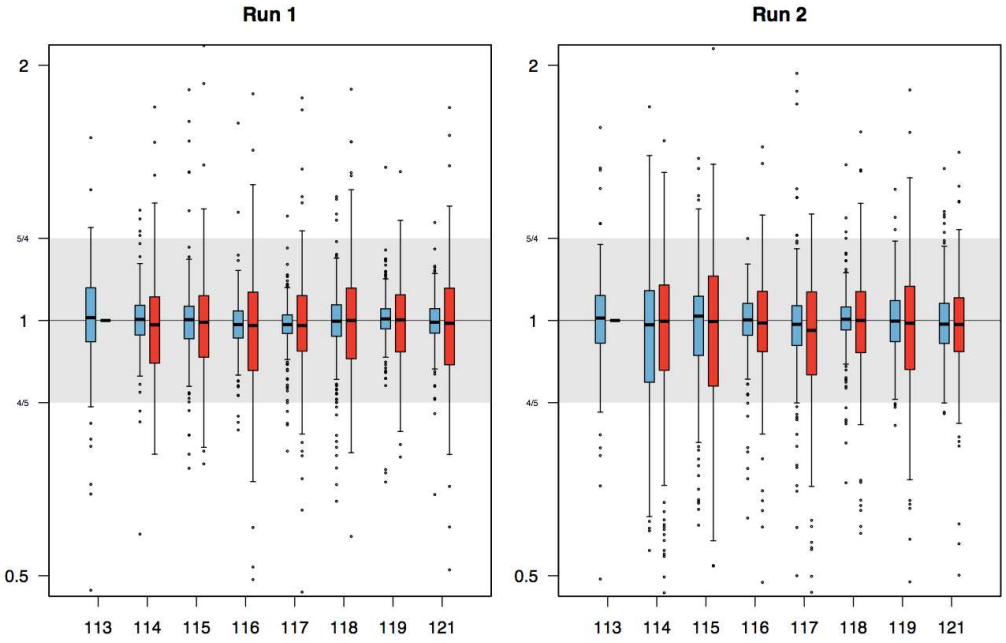
Run with 8 technical replicates of a master pool

Estimated relative protein abundance



Run with 8 technical replicates of a master pool

Variability



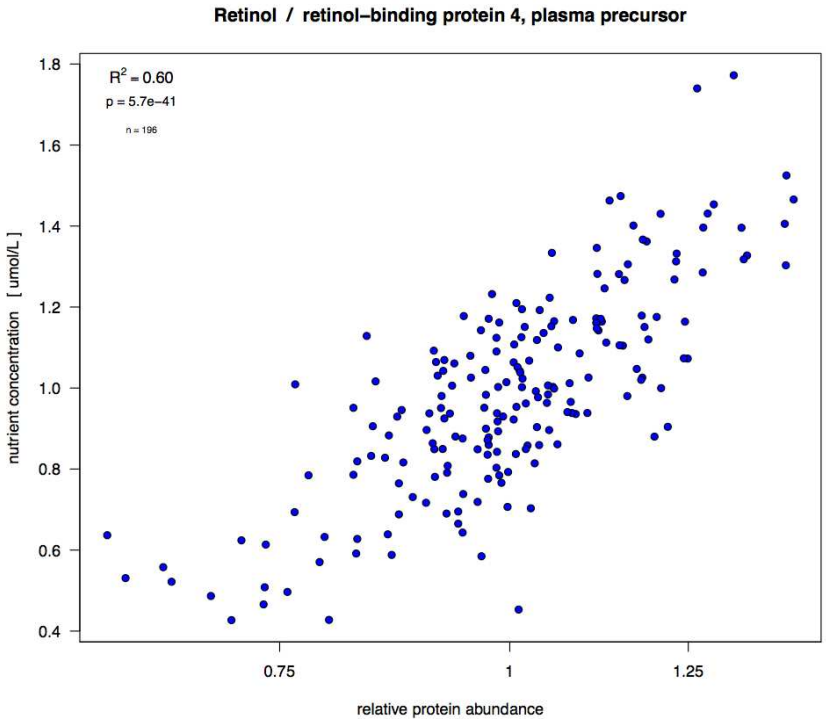
Variability in ratios

$$\mathrm{SD}\left(\frac{X}{Y}\right) \approx \left|\frac{X}{Y}\right| \times \sqrt{\left(\frac{\sigma_X}{X}\right)^2 - 2\rho_{(X,Y)}\left(\frac{\sigma_X}{X}\right)\left(\frac{\sigma_Y}{Y}\right) + \left(\frac{\sigma_Y}{Y}\right)^2}$$

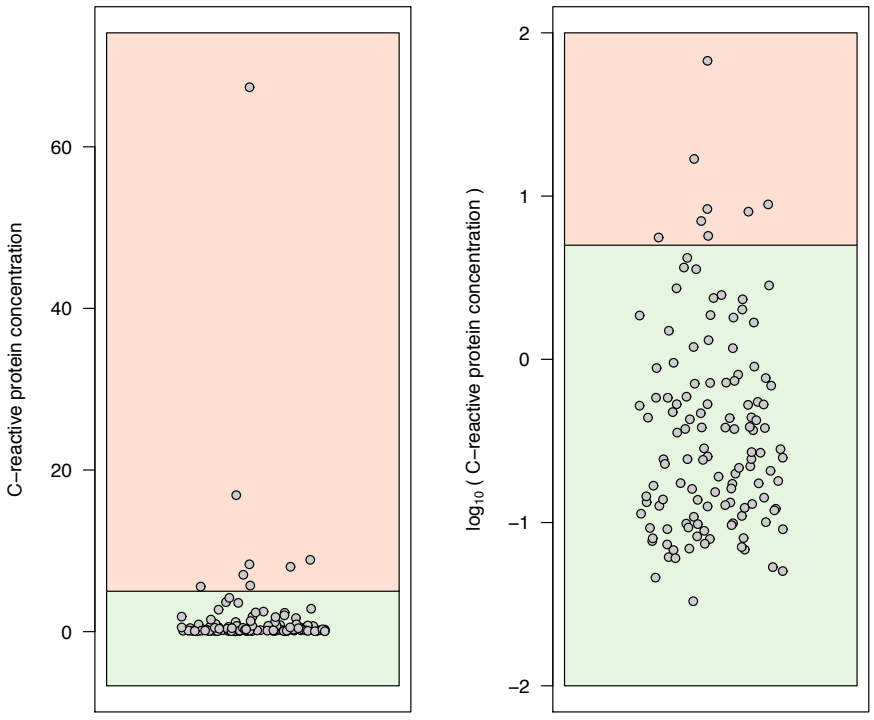
Design

	LP:113	LP:114	LP:115	LP:116	LP:117	LP:118	LP:119	LP:121
1	7	MP	5	2	1	6	4	3
2	13	9	11	10	MP	14	8	12
3	17	19	16	21	20	MP	15	18
4	27	25	26	28	23	22	24	MP
5	MP	32	30	29	31	35	33	34
6	38	42	MP	36	41	40	39	37
7	48	49	44	43	45	47	MP	46
8	52	56	50	MP	54	55	53	51
9	57	59	58	62	61	63	60	MP
10	MP	70	67	69	66	65	68	64
11	74	MP	77	75	71	72	73	76
12	80	79	MP	82	84	83	78	81
13	90	88	87	85	89	MP	86	91
14	98	97	92	MP	93	94	96	95
15	101	99	103	105	104	102	MP	100
16	106	110	107	109	MP	112	111	108

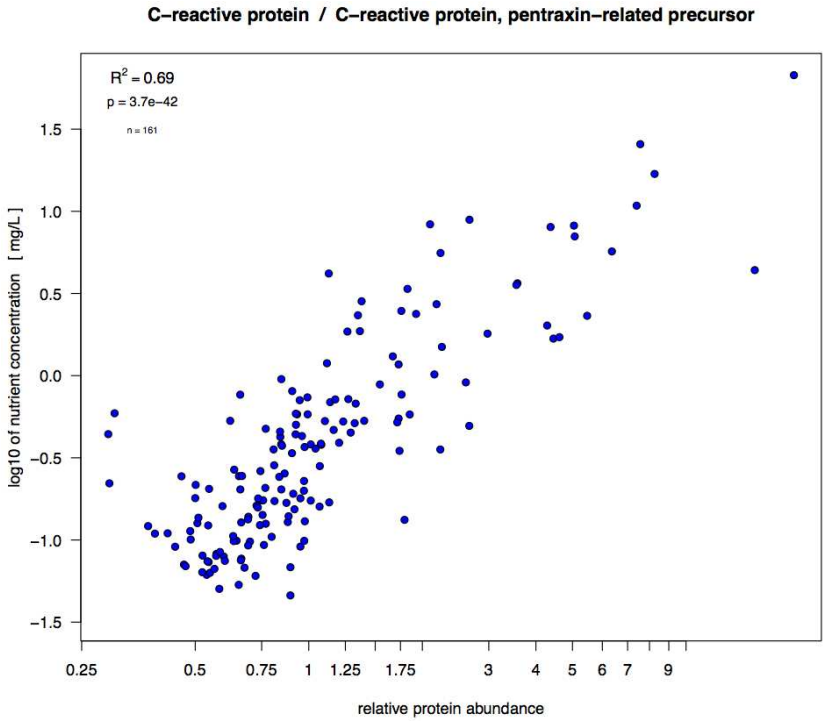
Does it work..?



C-reactive protein



C-reactive protein



Vitamin A

110823.196.ret.csv

	A	B	C	D	E	F
1		name	p	q	r2	r
2	gi55743122	retinol-binding protein 4, plasma precursor	5.69E-41	8.71E-39	0.60481557	0.77769889
3	gi4507725	transthyretin precursor	1.09E-17	8.31E-16	0.31548309	0.56167882
4	gi4557321	apolipoprotein A-I preproprotein	1.24E-12	6.31E-11	0.22940926	0.47896687
5	gi4502149	apolipoprotein A-II preproprotein	1.56E-09	5.97E-08	0.17172393	0.41439586
6	gi4826772	insulin-like growth factor binding protein,	1.75E-08	5.36E-07	0.15133691	0.38902045
7	gi4502511	complement component 9 precursor	5.91E-08	1.51E-06	0.14091446	-0.37538575
8	gi197333755	interferon-related developmental regulator 2	9.26E-08	2.03E-06	0.14691656	0.38329696
9	gi16418467	leucine-rich alpha-2-glycoprotein 1 precursor	2.81E-07	5.24E-06	0.12741449	-0.35695167
10	gi32130518	apolipoprotein C-II precursor	3.08E-07	5.24E-06	0.12662089	0.3558383
11	gi4501987	afamin precursor	5.43E-07	7.56E-06	0.12167304	0.34881663
12	gi4502161	apolipoprotein C-IV precursor	5.43E-07	7.56E-06	0.12166952	0.34881158
13	gi4557323	apolipoprotein C-III precursor	9.40E-07	1.20E-05	0.11685782	0.34184473
14	gi31652249	lipopolysaccharide-binding protein precursor	4.74E-06	5.58E-05	0.10257137	-0.32026766
15	gi4503635	prothrombin preproprotein	7.53E-06	8.24E-05	0.09843884	0.31374965
16	gi167614506	plastin-2	1.29E-05	0.00013163	0.09364461	-0.30601407
17	gi194018472	plasma serine protease inhibitor	1.59E-05	0.00015192	0.09178512	0.3029606
18	gi62243068	insulin-like growth factor binding protein	1.78E-05	0.0001606	0.09074546	0.30123986
19	gi67782358	complement factor B preproprotein	2.03E-05	0.00017299	0.08956746	-0.29927823
20	gi21361302	serine (or cysteine) proteinase inhibitor,	2.38E-05	0.00019172	0.08816145	0.29691994
21	gi133925809	inter-alpha (globulin) inhibitor H3 preproprotein	3.44E-05	0.00024956	0.08485491	-0.29129866
22	gi68299759	ecotropic viral integration site 5	3.51E-05	0.00024956	0.10236301	-0.3199422
23	gi22091452	apolipoprotein M	3.59E-05	0.00024956	0.08447905	0.2906528
24	gi119392081	complement factor I preproprotein	4.03E-05	0.00026812	0.0834349	-0.28885101
25	gi21071039	carnosinase 1 precursor	7.82E-05	0.00049907	0.07745823	0.27831319
26	gi4507509	tissue inhibitor of metalloproteinase 1	0.0001205	0.00073809	0.09297576	-0.30491926
27	gi71773110	apolipoprotein A-IV precursor	0.00013722	0.00080816	0.07238645	0.2690473
28	gi211938442	apolipoprotein L1 isoform c precursor	0.00030907	0.00175281	0.06504589	0.25504095
29	gi41393602	complement component 1, s subcomponent	0.00032142	0.00175777	0.06469123	-0.2543447
30	gi62000702	diacylglycerol kinase kappa	0.00034506	0.00182198	0.0889422	0.29823179
31	gi50659080	serpin peptidase inhibitor, clade A,	0.00047557	0.00241308	0.06114577	-0.24727671
32	gi167857790	orosomucoid 1 precursor	0.00048853	0.00241308	0.06090243	-0.24678418
33	gi4502157	apolipoprotein C-I precursor	0.00066472	0.00310527	0.05811476	0.24107003

C-reactive protein

110823.196.log10.crp.csv

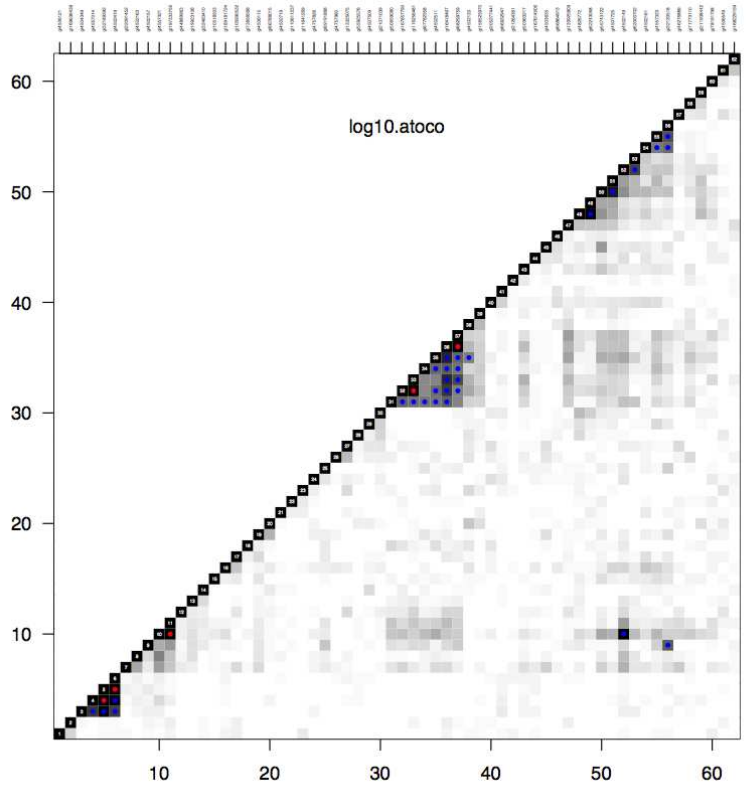
	A	B	C	D	E	F
1		name	p	q	r2	r
2	gi55770842	C-reactive protein, pentraxin-related precursor	3.70E-42	5.38E-40	0.6888836	0.82999012
3	gi16418467	leucine-rich alpha-2-glycoprotein 1 precursor	4.38E-23	3.19E-21	0.39705169	0.63012038
4	gi4502511	complement component 9 precursor	1.29E-22	6.24E-21	0.39037488	0.62479987
5	gi31652249	lipopolysaccharide-binding protein precursor	2.64E-18	9.62E-17	0.32527346	0.5703275
6	gi40316912	serum amyloid A protein preproprotein	1.32E-17	3.85E-16	0.33396889	0.57790041
7	gi167857790	orosomucoid 1 precursor	3.94E-15	9.56E-14	0.2731143	0.52260338
8	gi4502133	serum amyloid P component precursor	2.86E-14	5.93E-13	0.25833089	0.50826262
9	gi21361302	serine (or cysteine) proteinase inhibitor,	1.52E-13	2.76E-12	0.24563539	-0.49561617
10	gi68299759	ecotropic viral integration site 5	1.38E-12	2.23E-11	0.27150686	0.52106321
11	gi50659080	serpin peptidase inhibitor, clade A,	2.50E-12	3.64E-11	0.22388196	0.47316166
12	gi116256481	TNFAIP3-interacting protein 1	3.01E-12	3.98E-11	0.26444131	0.51423857
13	gi67782358	complement factor B preproprotein	1.85E-09	2.24E-08	0.17029677	0.41267029
14	gi4505047	lumican precursor	5.09E-09	5.69E-08	0.1618094	-0.40225539
15	gi119392081	complement factor I preproprotein	7.97E-09	8.28E-08	0.15801793	0.39751469
16	gi4505529	orosomucoid 2 precursor	9.29E-09	9.01E-08	0.15671691	0.39587487
17	gi10835095	serum amyloid A-4 protein precursor	1.24E-08	1.12E-07	0.15430208	0.39281303
18	gi156523970	alpha-2-HS-glycoprotein	2.32E-07	1.99E-06	0.12907536	-0.3592706
19	gi38016947	complement component 5 preproprotein	5.58E-07	4.51E-06	0.12143726	0.3484785
20	gi156627579	tetranectin precursor	6.84E-07	5.24E-06	0.11965164	-0.34590698
21	gi4501987	afamin precursor	1.37E-06	9.97E-06	0.11353807	-0.33695411
22	gi4502261	serpin peptidase inhibitor, clade C,	3.01E-06	2.05E-05	0.10658821	-0.32647849
23	gi133925809	inter-alpha (globulin) inhibitor H3 preproprotein	3.09E-06	2.05E-05	0.106347	0.32610888
24	gi70778918	inter-alpha globulin inhibitor H2 polypeptide	4.49E-06	2.84E-05	0.10304765	-0.32101035
25	gi190341024	SPARC-like protein 1 precursor	6.03E-06	3.65E-05	0.10399463	-0.32248199
26	gi4557391	complement component 8, beta polypeptide	6.28E-06	3.66E-05	0.10005909	0.31632118
27	gi156616294	N-acetylmuramoyl-L-alanine amidase precursor	7.59E-06	4.10E-05	0.09837695	-0.31365099
28	gi154759291	serine (or cysteine) proteinase inhibitor,	7.61E-06	4.10E-05	0.09835044	0.31360874
29	gi71773110	apolipoprotein A-IV precursor	2.28E-05	0.00011835	0.08855056	-0.29757446
30	gi13540563	complement factor H-related 5 precursor	2.75E-05	0.00013772	0.09332612	0.30549323
31	gi7382460	sex hormone-binding globulin isoform 1	3.34E-05	0.00016205	0.08816074	-0.29691873
32	gi51896037	g protein-coupled receptor kinase 6	5.59E-05	0.00026214	0.11138431	0.33374288
33	gi4507725	transthyretin precursor	9.07E-05	0.00040745	0.07612001	-0.27589856

Alpha-tocopherol

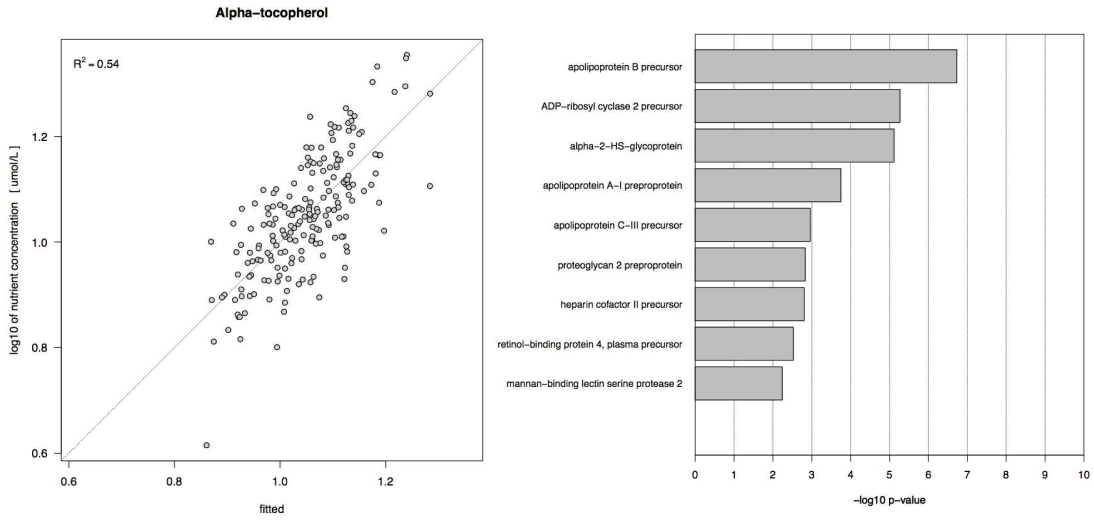
110823.196.log10.atoco.csv

	A	B	C	D	E	F
1		name	p	q	r2	r
2	gi4557321	apolipoprotein A-I preproprotein	1.51E-11	2.17E-09	0.20962777	0.45785125
3	gi4502149	apolipoprotein A-II preproprotein	1.59E-08	8.26E-07	0.15215365	0.39006878
4	gi4557323	apolipoprotein C-III precursor	1.72E-08	8.26E-07	0.15148774	0.38921426
5	gi197333755	interferon-related developmental regulator 2	2.55E-08	9.19E-07	0.1587073	0.39838085
6	gi55743122	retinol-binding protein 4, plasma precursor	2.31E-07	6.65E-06	0.12913973	0.35936017
7	gi32130518	apolipoprotein C-II precursor	2.87E-07	6.88E-06	0.1272512	0.35672286
8	gi4502161	apolipoprotein C-IV precursor	4.10E-07	8.25E-06	0.12412622	0.35231551
9	gi22091452	apolipoprotein M	4.58E-07	8.25E-06	0.12315996	0.35094154
10	gi189181724	proteoglycan 4 isoform D	7.35E-07	1.18E-05	0.12319958	0.35099798
11	gi105990532	apolipoprotein B precursor	8.26E-07	1.19E-05	0.1179936	0.34350197
12	gi46276889	proteoglycan 2 preproprotein	2.85E-05	0.00037273	0.10525685	-0.32443312
13	gi4502157	apolipoprotein C-I precursor	0.00053118	0.00637529	0.06014477	0.24524431
14	gi4506115	protein C preproprotein	0.00082365	0.00912515	0.05617441	0.23701141
15	gi167614506	plastin-2	0.00105849	0.01065496	0.05390453	-0.23217348
16	gi4506121	vitamin K-dependent protein Z precursor	0.00117451	0.01065496	0.05921668	0.24334478
17	gi16418467	leucine-rich alpha-2-glycoprotein 1 precursor	0.00118367	0.01065496	0.05289343	-0.22998572
18	gi4507725	transthyretin precursor	0.00135436	0.01147425	0.05167527	0.22732196
19	gi4507509	tissue inhibitor of metalloproteinase 1	0.00145027	0.01160419	0.06474316	-0.25444677
20	gi4826772	insulin-like growth factor binding protein,	0.00163967	0.01224974	0.04994732	0.22348898
21	gi205277441	serine (or cysteine) proteinase inhibitor,	0.00170105	0.01224974	0.04961525	-0.22274481
22	gi4757826	beta-2-microglobulin precursor	0.00349793	0.0235699	0.04469439	-0.21141049
23	gi21071039	carnosinase 1 precursor	0.00360032	0.0235699	0.04285333	0.20701046
24	gi10518503	coagulation factor VII isoform b	0.00396277	0.0248148	0.0454377	0.21316122
25	gi4757960	cadherin 1, type 1 preproprotein	0.00425644	0.02485371	0.05769314	-0.24019397
26	gi4502133	serum amyloid P component precursor	0.00431411	0.02485371	0.04122702	0.20304437
27	gi89191868	von Willebrand factor preproprotein	0.00456029	0.02526147	0.04072853	-0.2018131
28	gi4503635	prothrombin preproprotein	0.00505979	0.02636908	0.03979555	0.19948821
29	gi44889963	scavenger receptor cysteine-rich type 1	0.00512641	0.02636908	0.04465269	-0.21131183
30	gi4502511	complement component 9 precursor	0.00539909	0.02681402	0.03921341	-0.19802377
31	gi133925809	inter-alpha (globulin) inhibitor H3 preproprotein	0.00613171	0.02943744	0.0380733	-0.19512381
32	gi68299759	ecotropic viral integration site 5	0.00663732	0.03083688	0.04542465	-0.21313058
33	gi55925576	insulin-like growth factor-binding protein 2	0.0074594	0.03357327	0.03805789	-0.19508431

Alpha-tocopherol



Alpha-tocopherol



Statistical models

ANOVA models for microarray data

A microarray experiment may involve multiple arrays to compare multiple samples. Every measurement in a microarray experiment is associated with a particular combination of an array in the experiment, a dye (red or green), a variety, and a gene. Let y_{ijkg} denote the measurement from the i^{th} array, j^{th} dye, k^{th} variety, and g^{th} gene. To account for the multiple sources of variation in a microarray experiment, consider the model

$$\log(y_{ijkg}) = \mu + A_i + D_j + V_k + G_g + (AG)_{ig} + (VG)_{kg} + \epsilon_{ijkg}, \quad (1)$$

where μ is the overall average signal, A_i represents the effect of the i^{th} array, D_j represents the effect of the j^{th} dye, V_k represents the effect of the k^{th} variety, G_g represents the effect of the g^{th} gene, $(AG)_{ig}$ represents a combination of array i and gene g (i.e., a particular spot on a particular array), and $(VG)_{kg}$ represents the interaction between the k^{th} variety and the g^{th} gene. The error terms ϵ_{ijkg} are assumed to be independent and identically distributed with mean 0. The array effects A_i account for differences

[KERR ET AL | J·COMP·BIO 7: 2000]

Statistical models

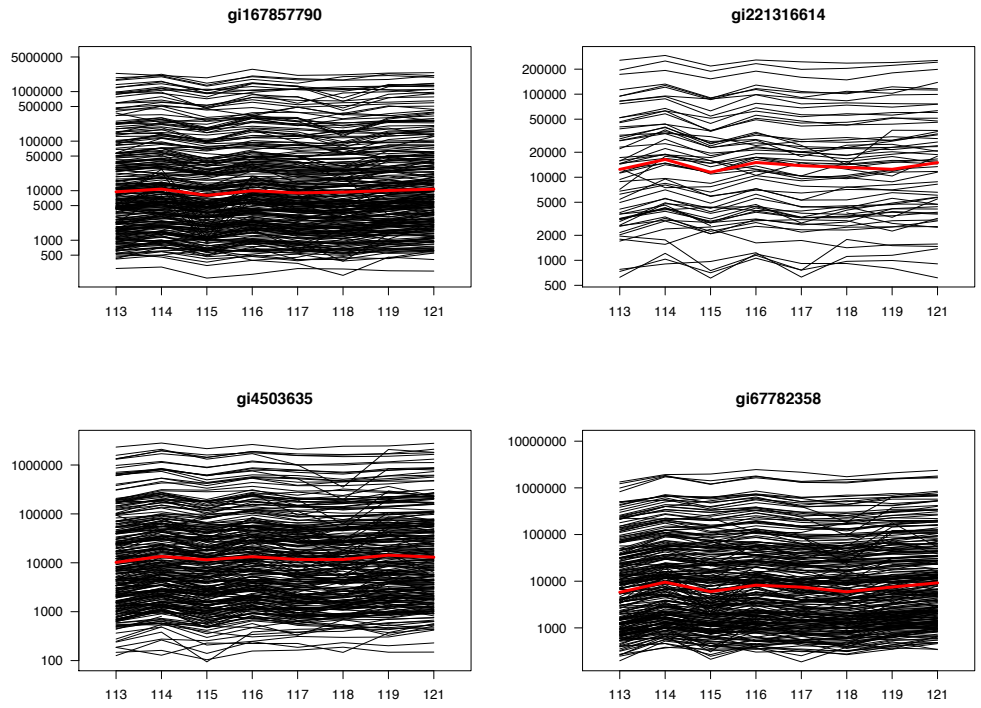
factor specified in Model 2. Furthermore, let $y_{i,j(i),c,q,s,l}$ be the log reporter ion peak area corresponding to protein i, peptide j, condition c, iTRAQ experiment q, tag l, and spectrum s. On a logarithmic scale, the collection of observed reporter ion peak areas is given by the expression

$$y_{i,j(i),c,q,s,l} = p_i + r_{i,c} + r_c + f_{j(i)} + g_{j(i),c} + v_{q,l} + b_q + h_{i,j(i),c,q,s,l} \quad (3)$$

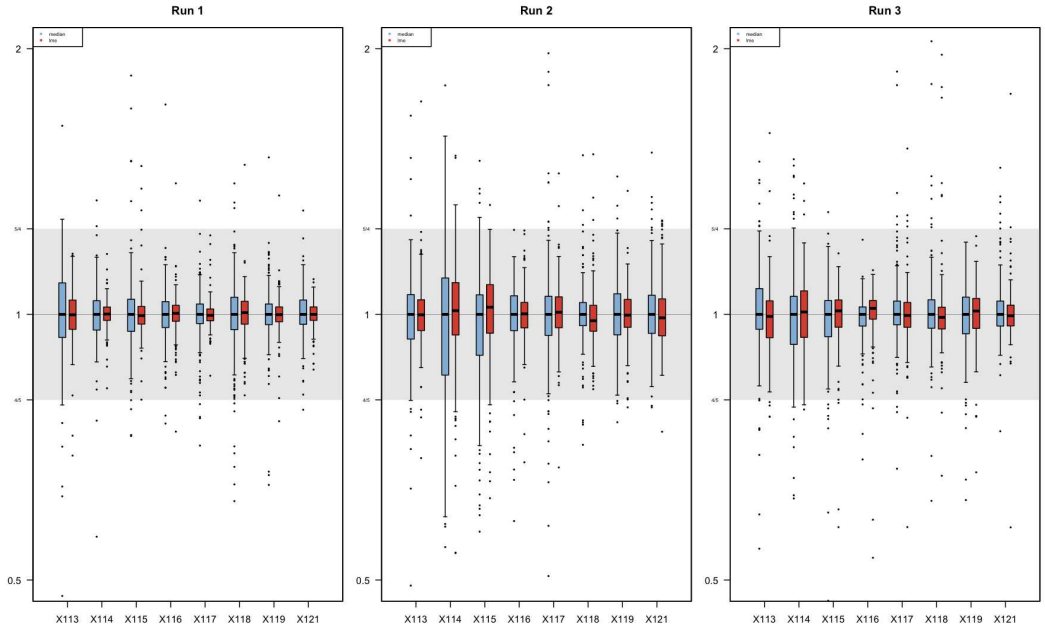
Note that we now write our model as an additive equation.

[HILL ET AL | J·PROT·RES 7: 2008]

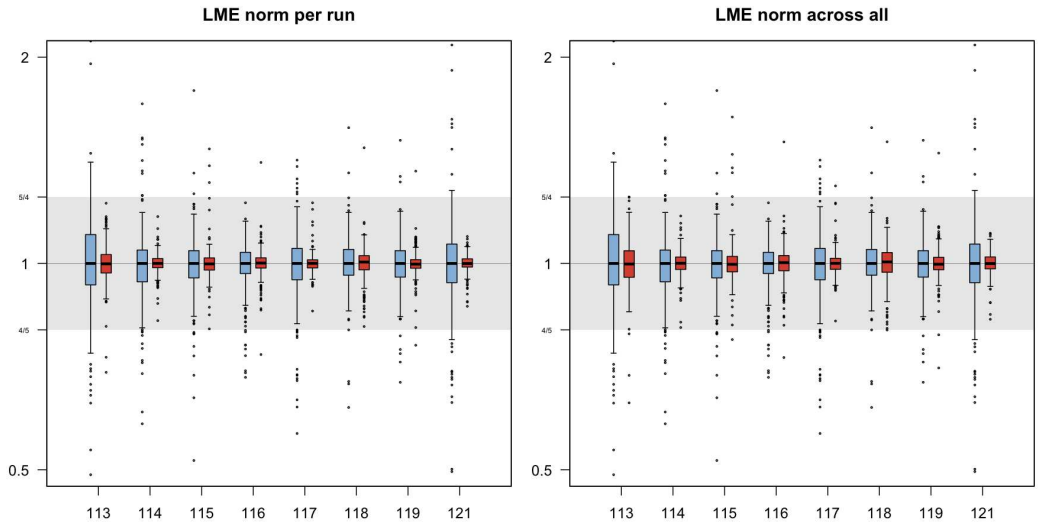
Reporter ion intensities



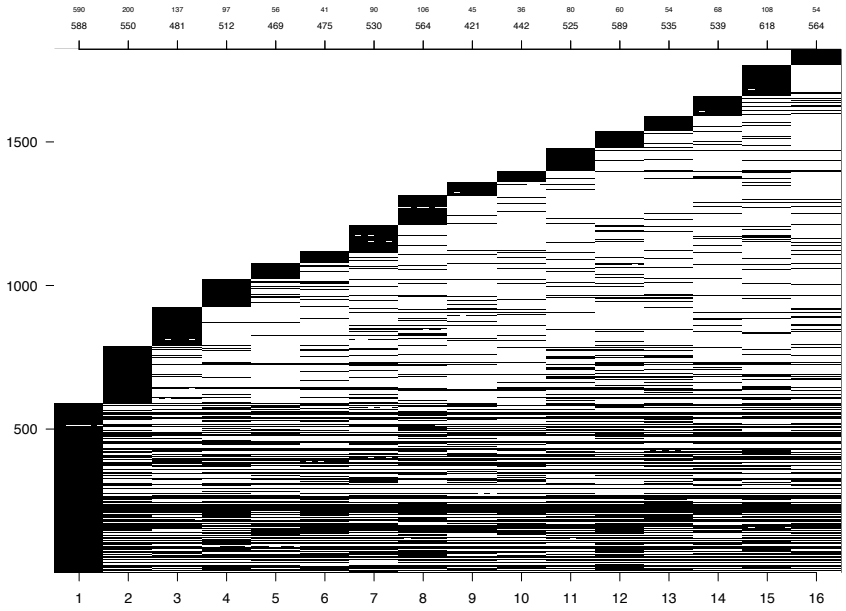
Quantitation



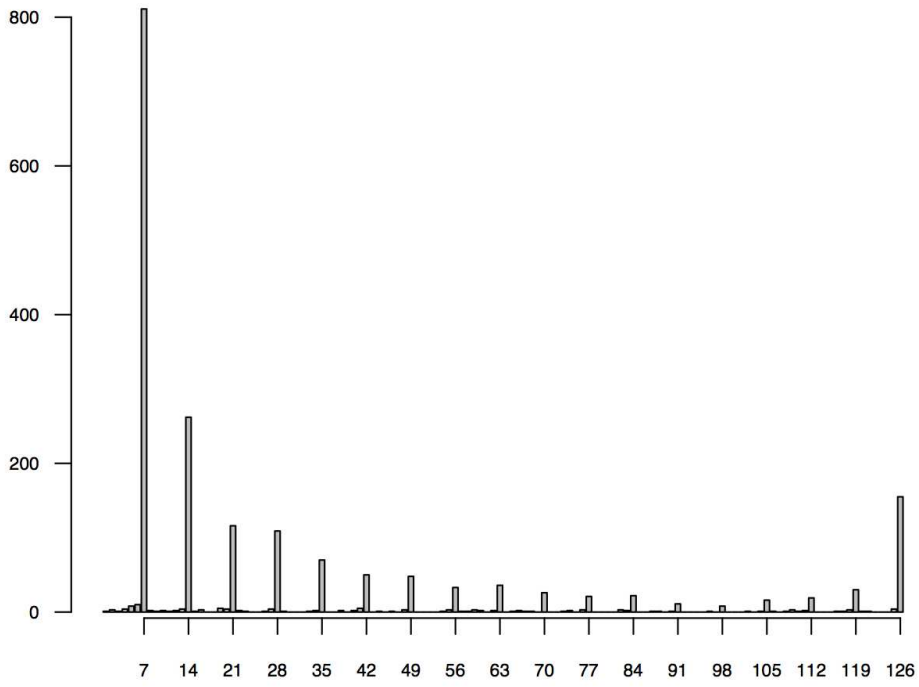
Quantitation



Missing data



Missing data



http: //biostat.jhsph.edu/∼iruczins/

ingo@jhu.edu