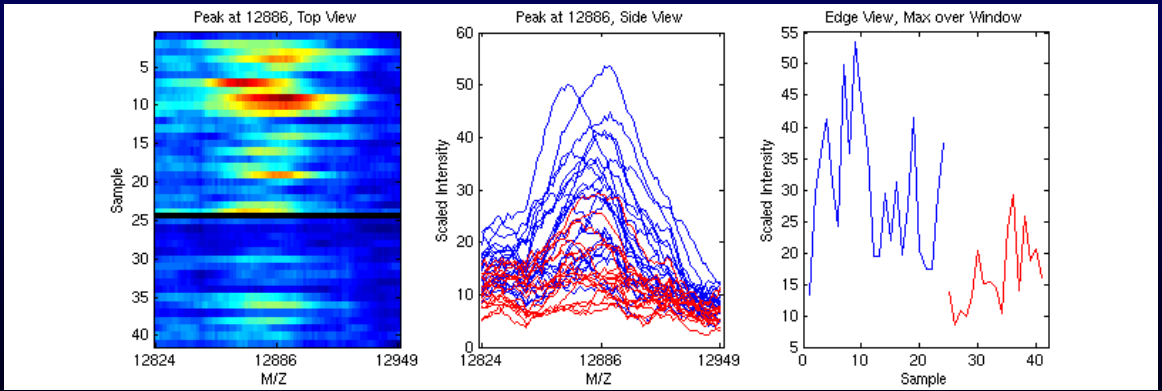
Reproducibility of SELDI/MALDI Spectra

Keith Baggerly

Biostatistics & Applied Mathematics
MD Anderson Cancer Center



Context: Why Are We Excited?

Profiles at this point are being assessed using serum and urine, not tissue biopsies

Spectra are cheaper to run on a per unit basis than microarrays

Can run samples on large numbers of patients

Why Focus on Reproducibility?

(Straw Man)

Many studies are working with (temporarily) black box “patterns”.

These are perfectly valid as long as the procedure remains stable.

More generally though, trying to track down the underlying biology is immeasurably aided if we can get pointers to the same locations over time.

What Types of Factors Can Affect Reproducibility?

preprocessing, artifacts, calibration

What Types of Factors Can Affect Reproducibility?

preprocessing, artifacts, calibration

External Factors.

What Types of Factors Can Affect Reproducibility?

preprocessing, artifacts, calibration

External Factors.

run date, run order, chip lot, sample handling

What Types of Factors Can Affect Reproducibility?

preprocessing, artifacts, calibration

External Factors.

run date, run order, chip lot, sample handling

Experimental Design

What Types of Factors Can Affect Reproducibility?

preprocessing, artifacts, calibration

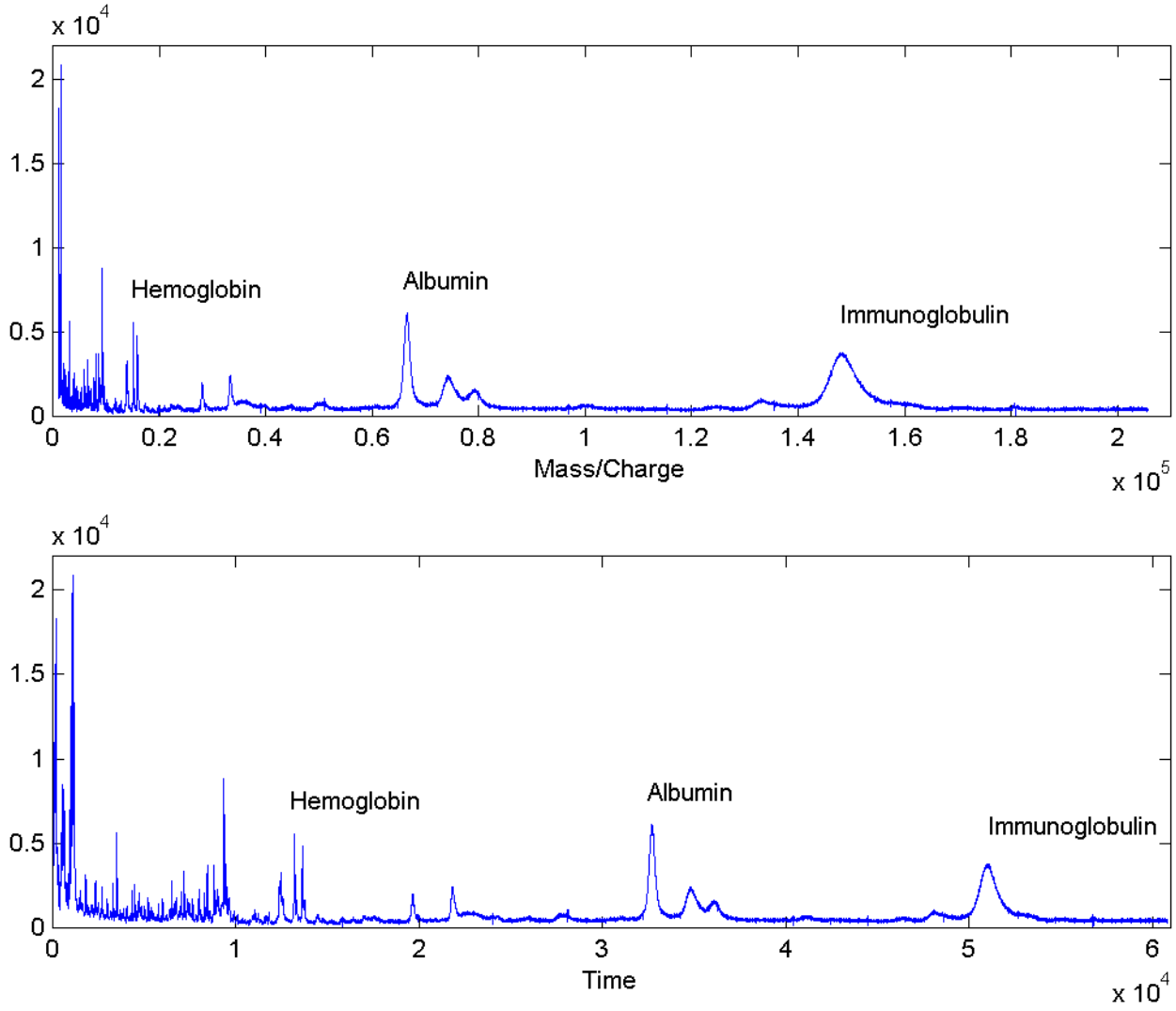
External Factors.

run date, run order, chip lot, sample handling

Experimental Design

General Rules: Be a tad suspicious of the data, be a bit aware of some of the underlying processing steps, and look at pictures.

Example: *A Spectrum*



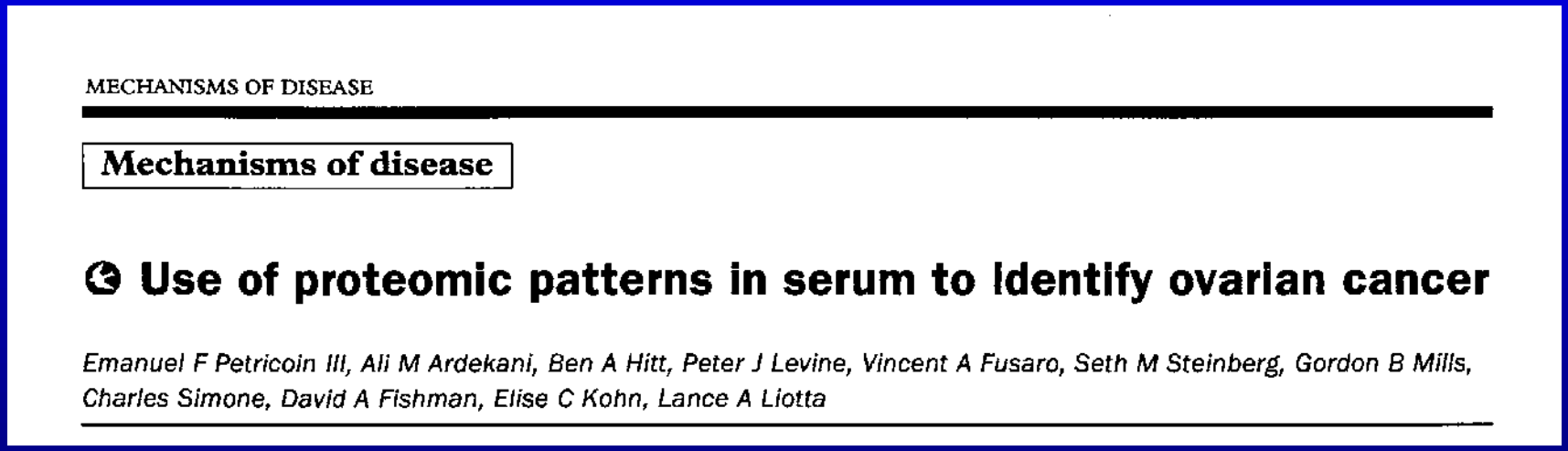
A Tale of Two Examples

Example 1 – Learning from the literature (SELDI)

Example 2 – Testing out our understanding (MALDI)

A story in pictures

A SELDI Example: Feb 16 '02 Lancet



MECHANISMS OF DISEASE

Mechanisms of disease

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

- 100 ovarian cancer patients
- 100 normal controls
- 16 patients with "benign disease"

Use 50 cancer and 50 normal spectra to train a classification method; test the algorithm on the remaining samples.

Their Results

- Correctly classified 50/50 of the ovarian cancer cases.
- Correctly classified 46/50 of the normal cases.
- Correctly classified 16/16 of the benign disease as “other”.

Data at http://www.ncifdaproteomics.com, used to be at http://clinicalproteomics.steem.com

Large sample sizes, using serum

The Data Sets

3 data sets on ovarian cancer

Data Set 1 – The initial experiment. 216 samples, baseline subtracted, H4 chip

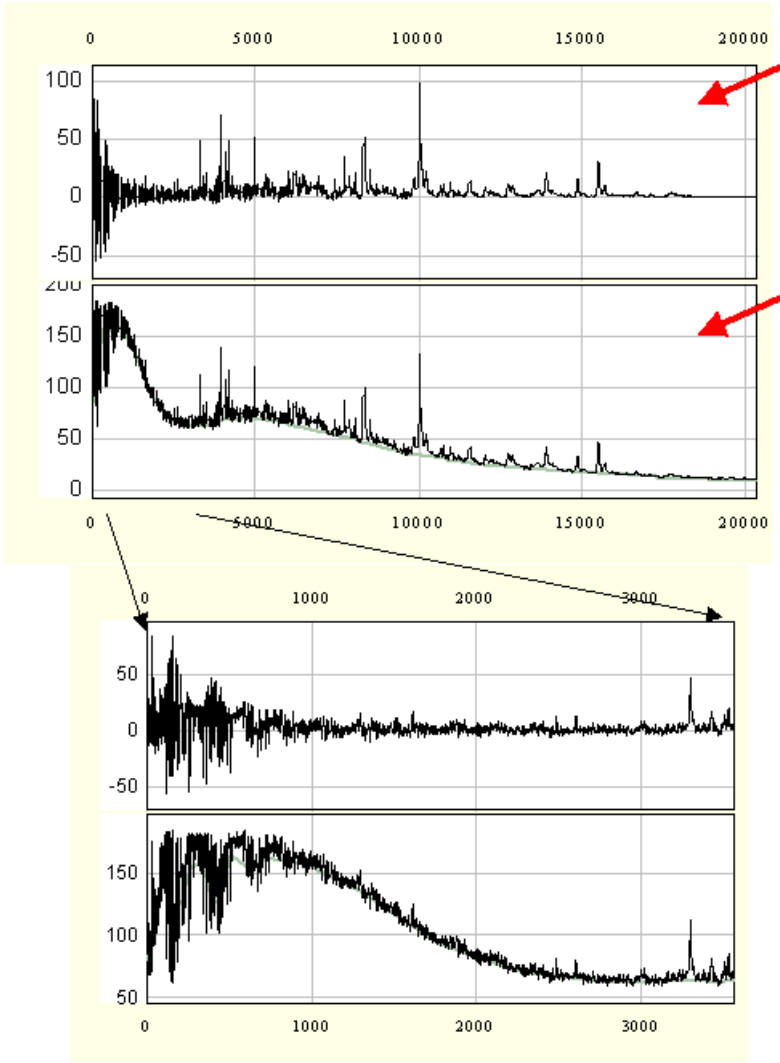
Data Set 2 – Followup: the same 216 samples, baseline subtracted, WCX2 chip

Data Set 3 – New experiment: 162 cancers, 91 normals, baseline NOT subtracted, WCX2 chip

A set of 5-7 separating peaks is supplied for each data set.

We tried to (a) replicate their results, and (b) check consistency of the proteins found

We Can't Replicate their Results (DS1 & DS2)



They posted this

Baseline Subtraction ON

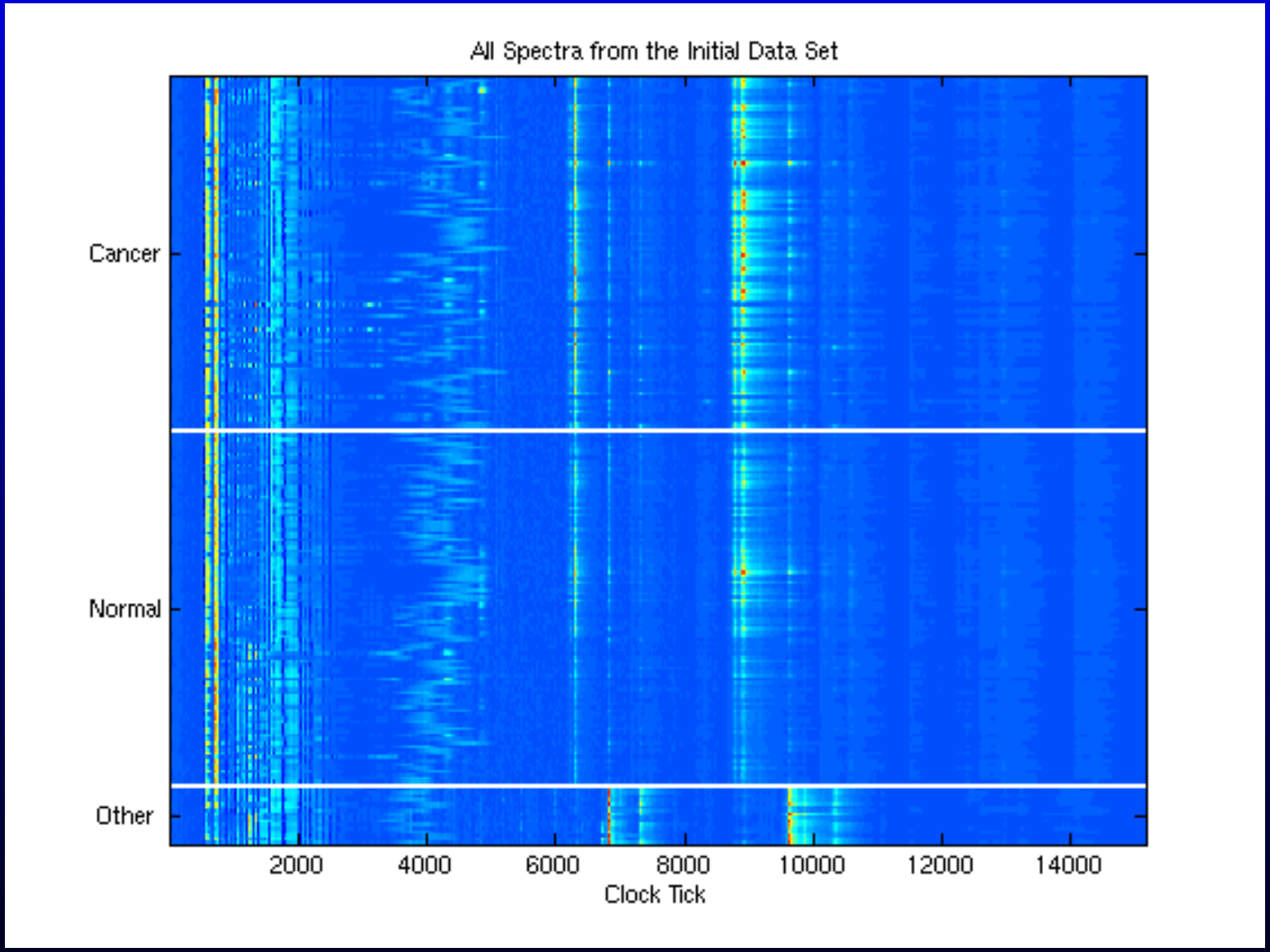
They analyzed this

Baseline Subtraction OFF

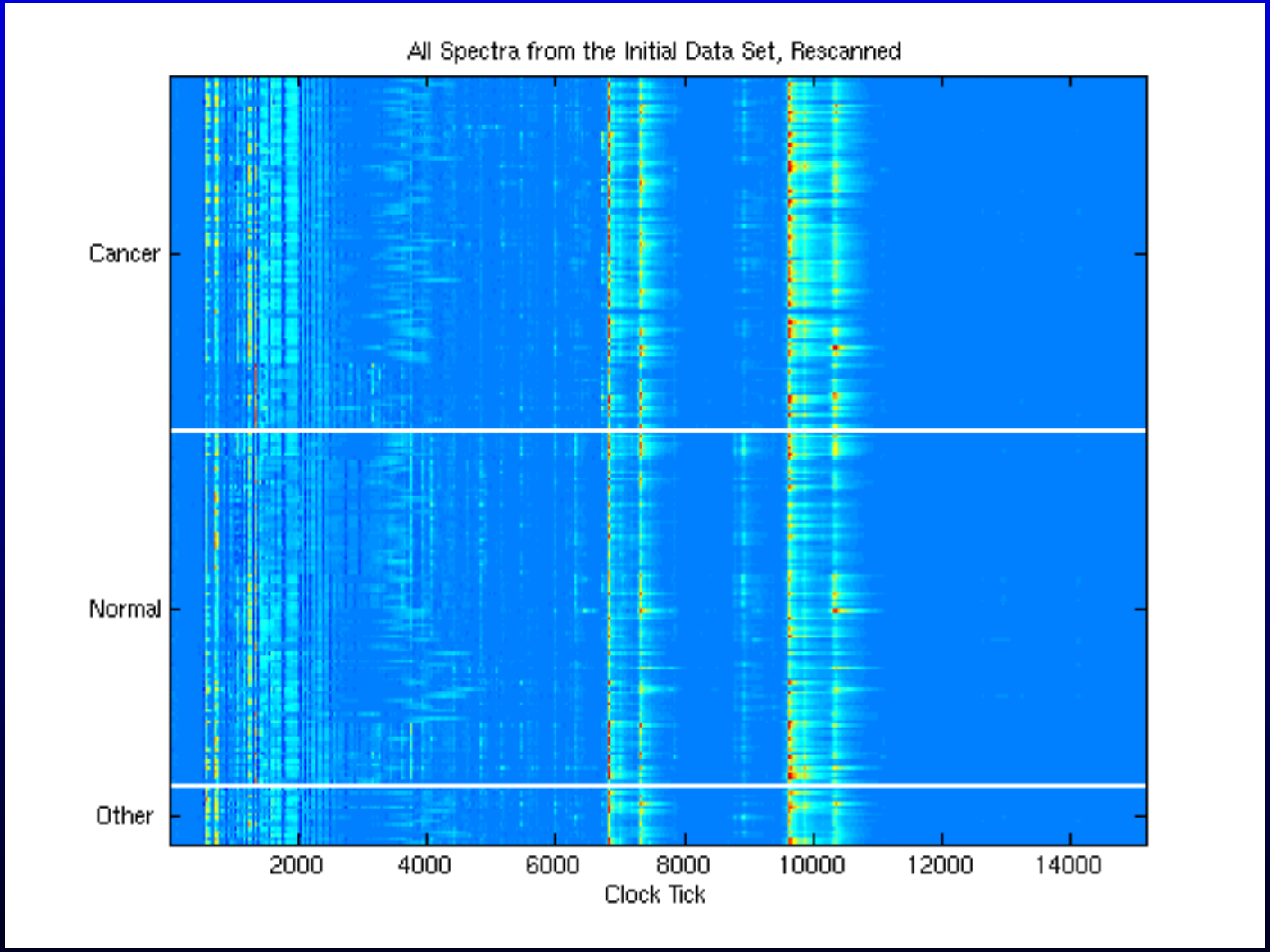
Zoomed in with Baseline Sub ON

Zoomed in with Baseline Sub OFF

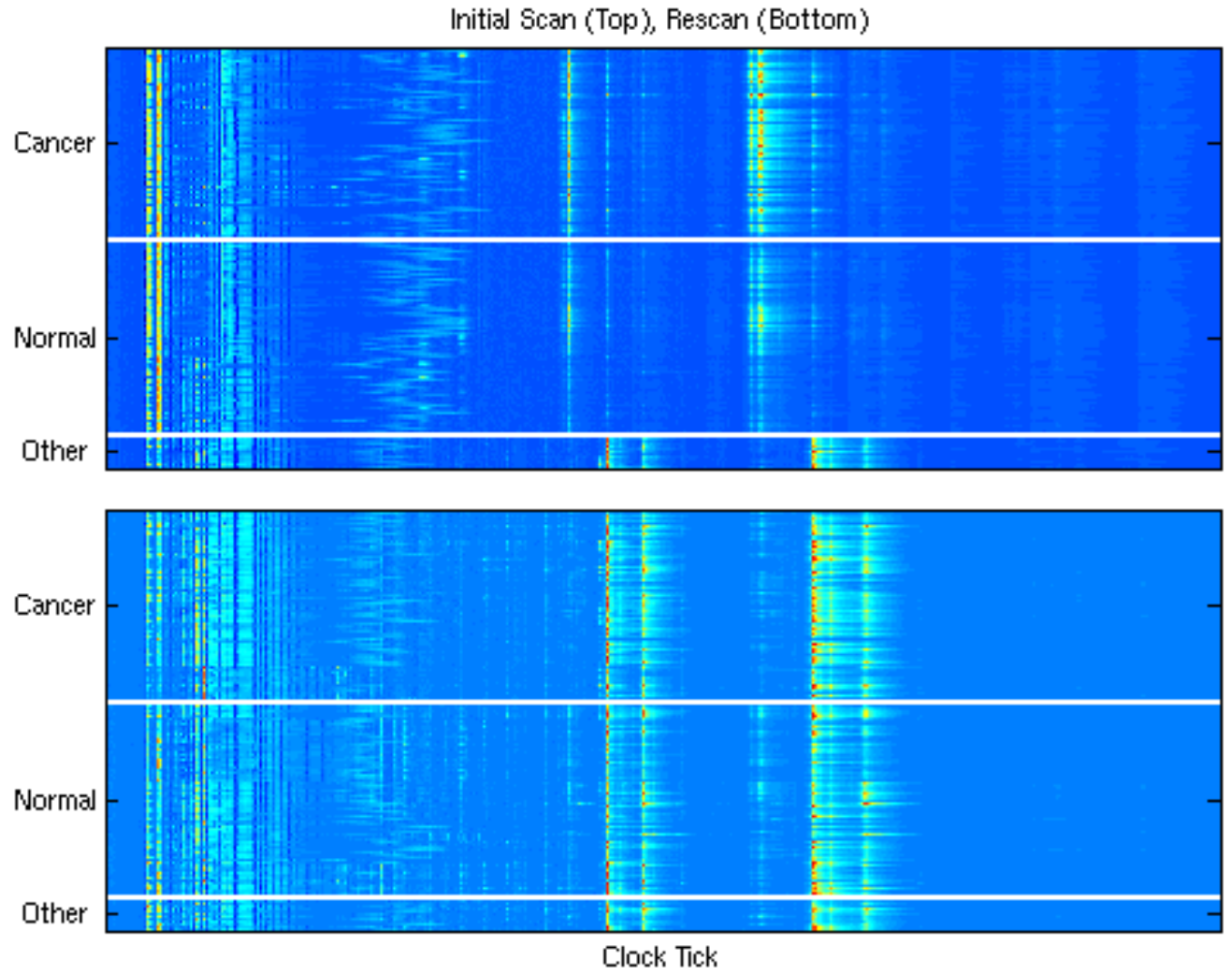
Some Structure is Visible in DS1



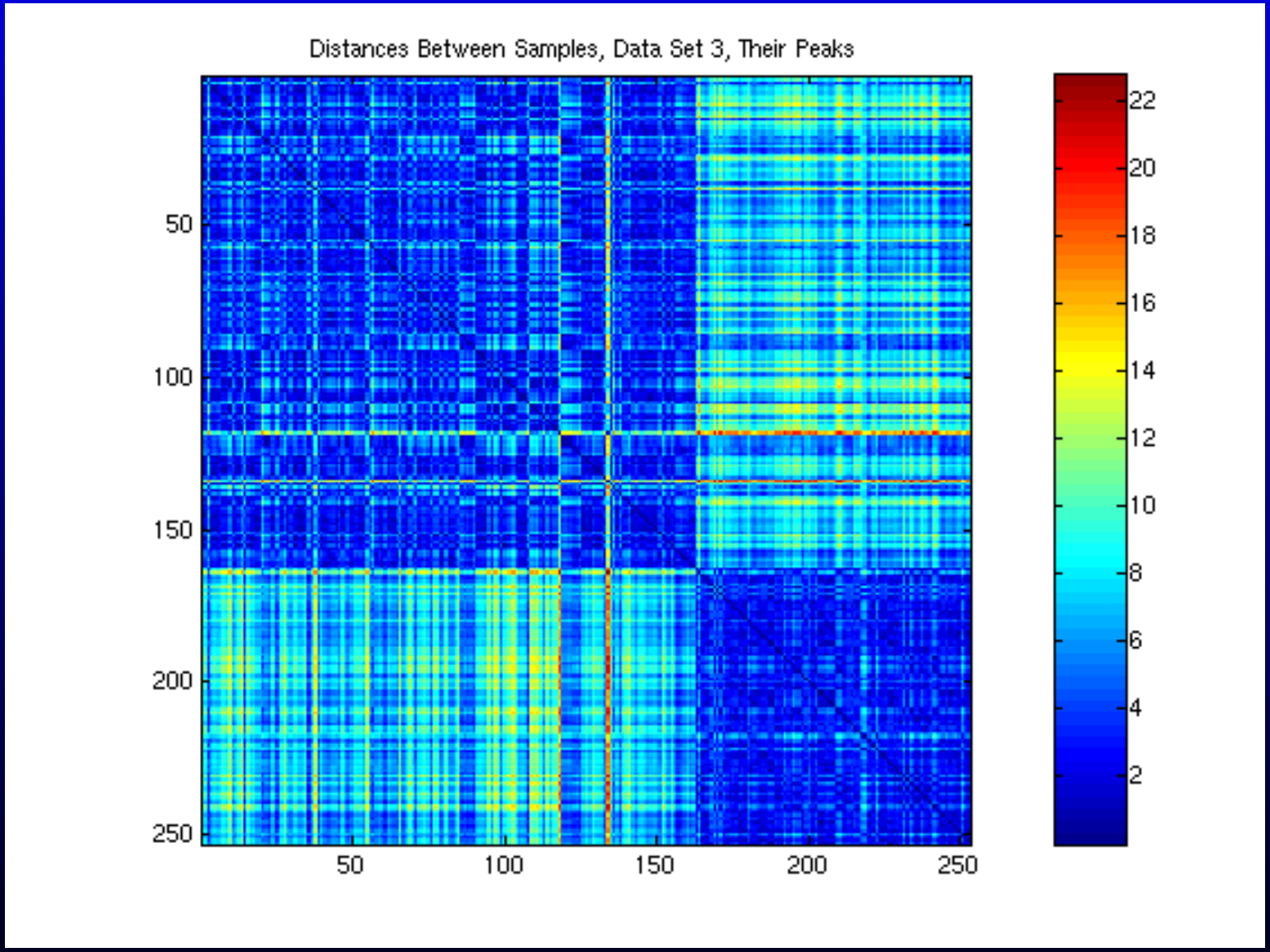
Or is it? Not in DS2



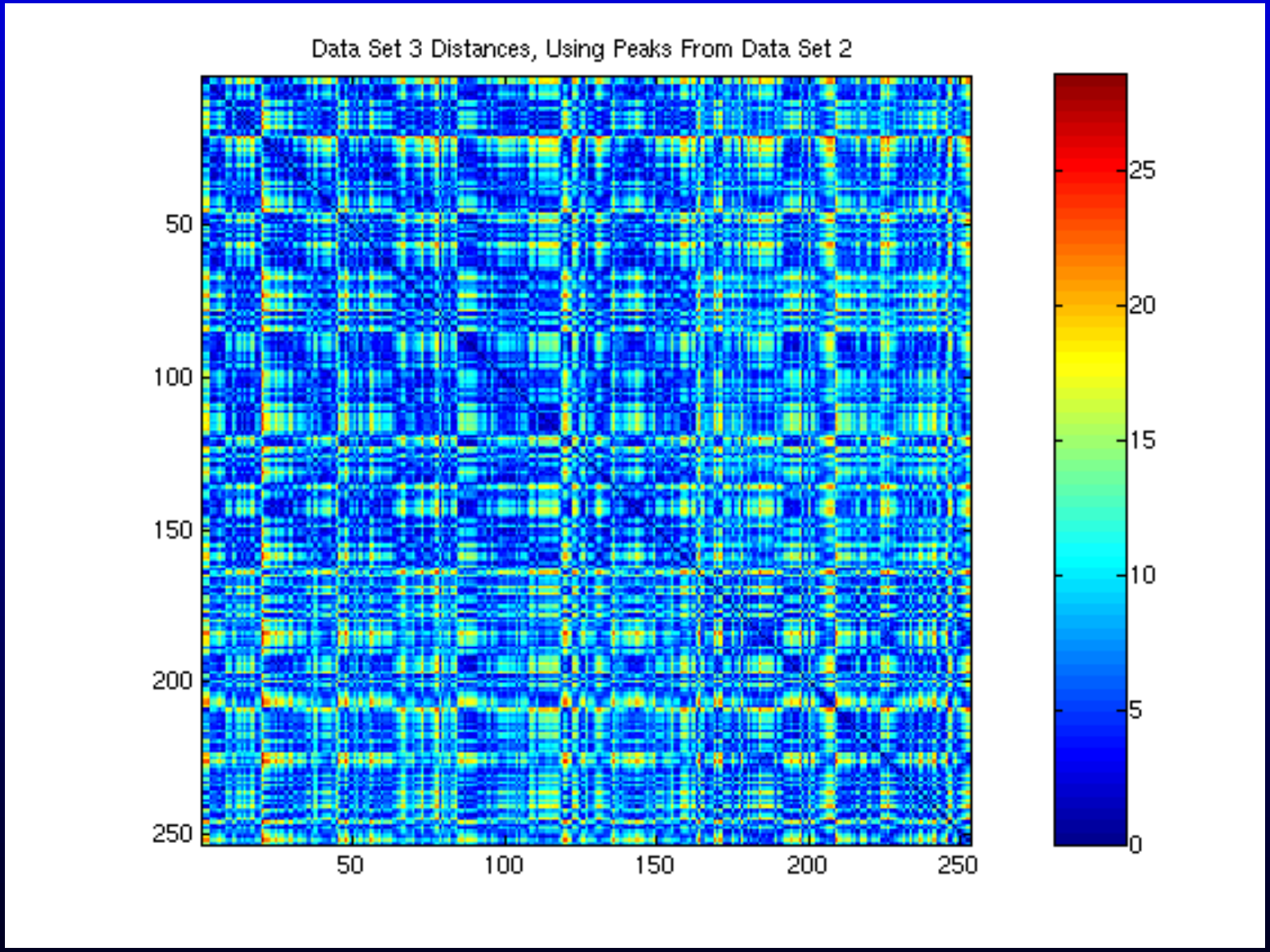
Processing Can Trump Biology (DS1 & DS2)



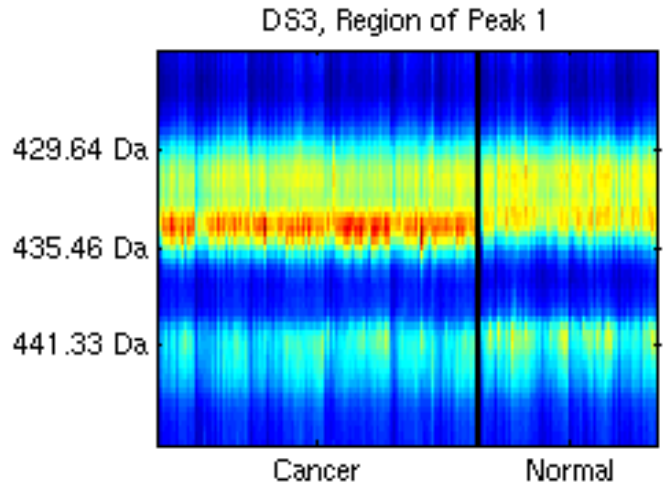
We Can Analyze Data Set 3!

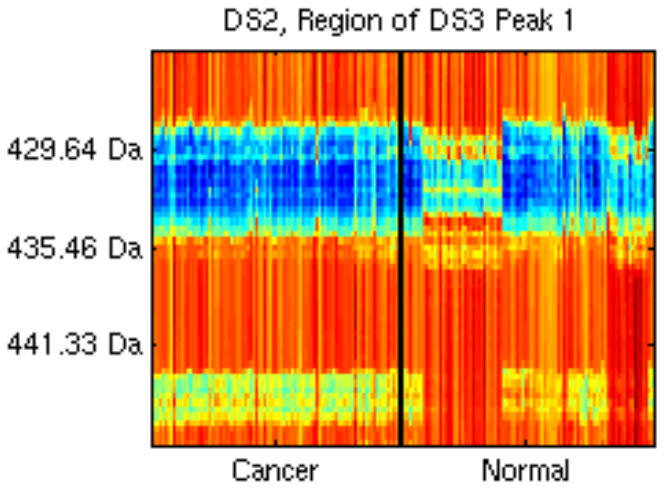


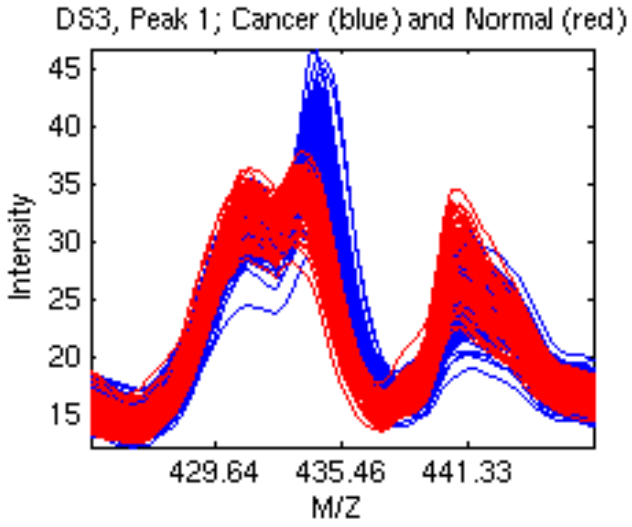
Do the DS2 Peaks Work for DS3?

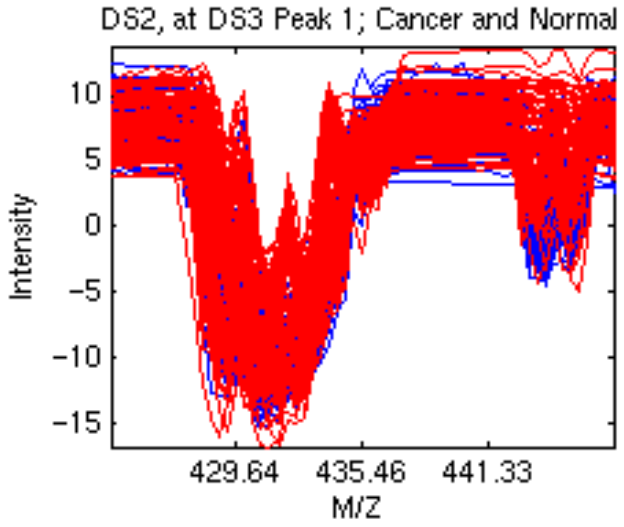


Do the DS3 Peaks Work for DS2?

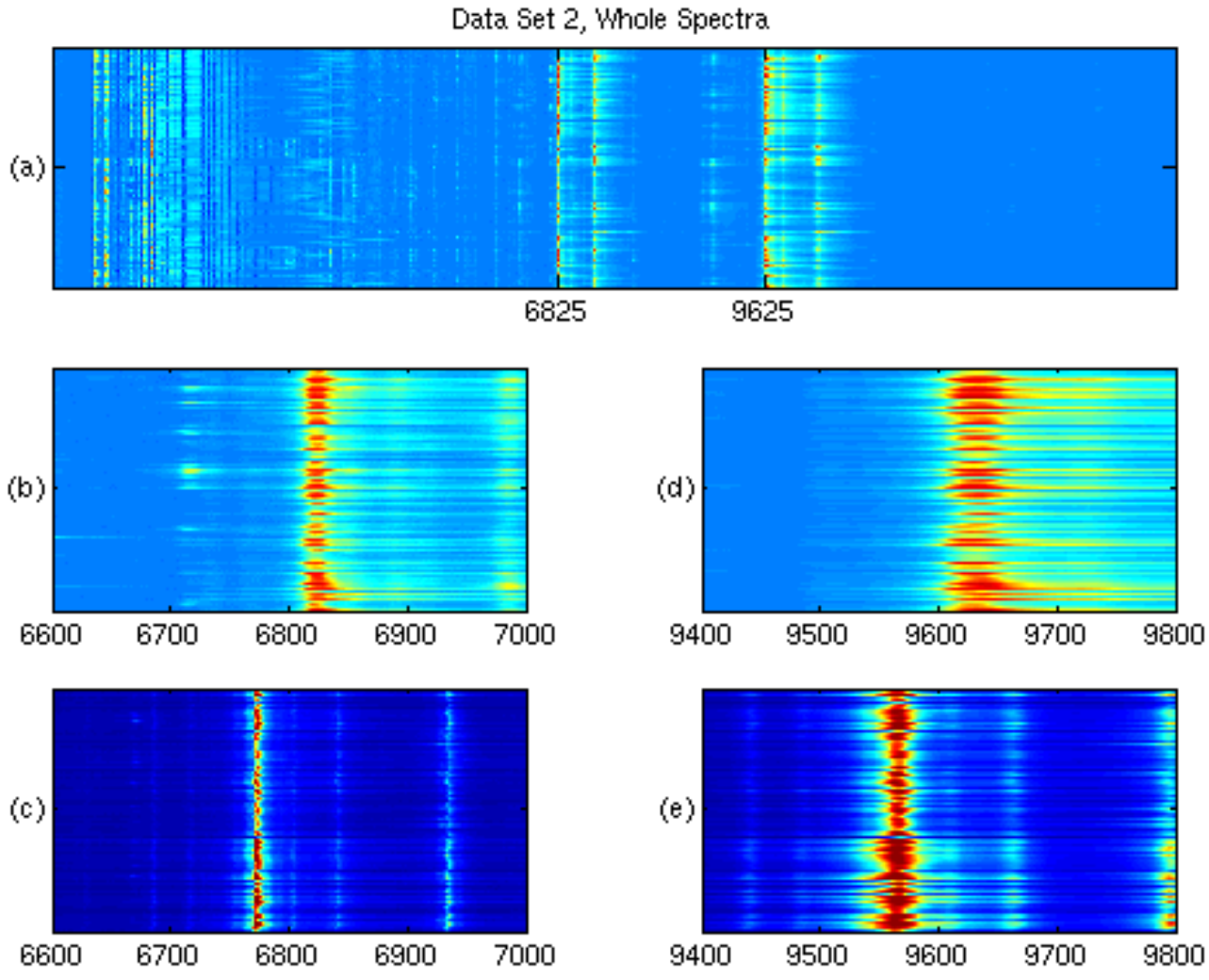




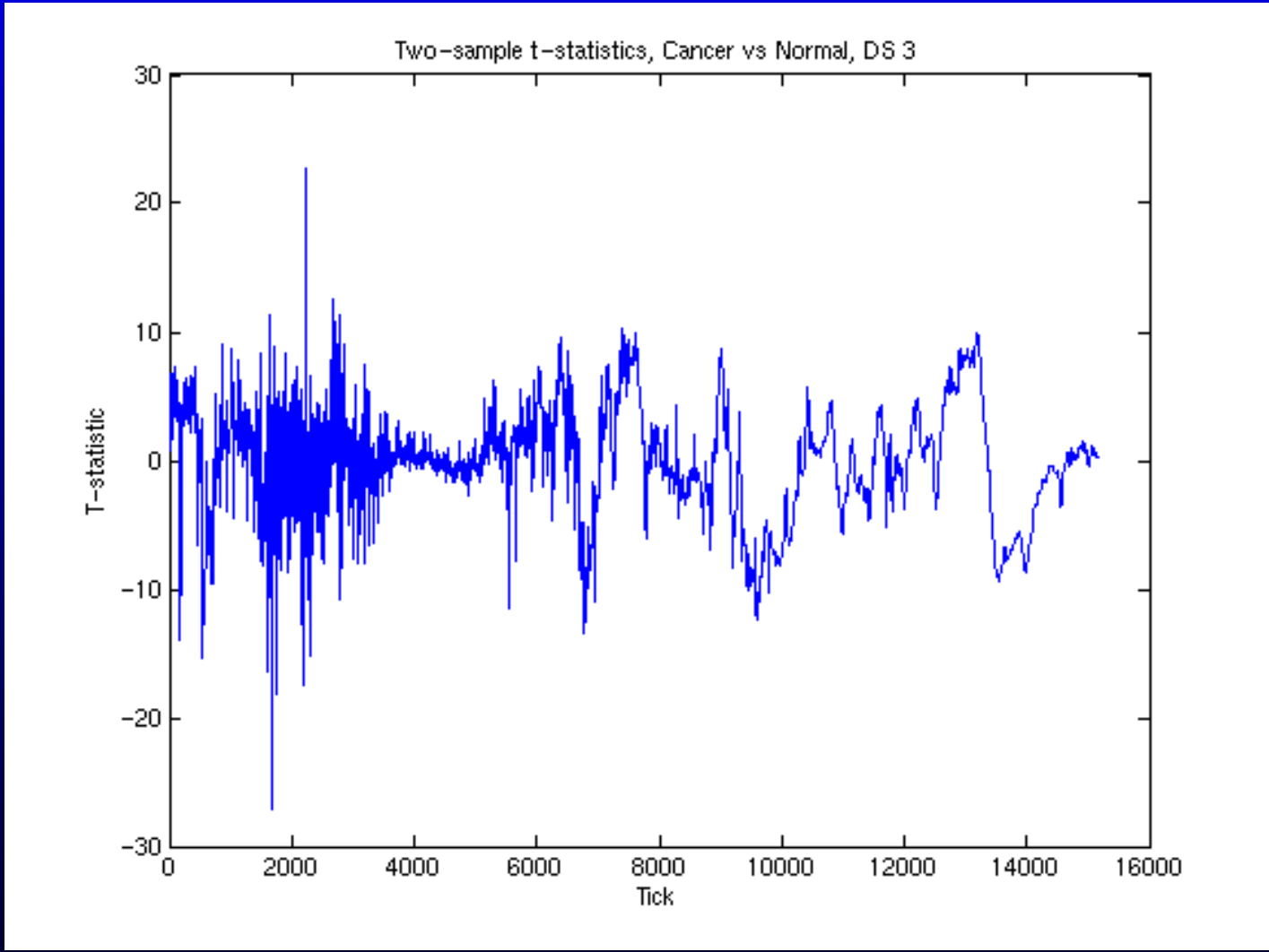




Peaks are Offset

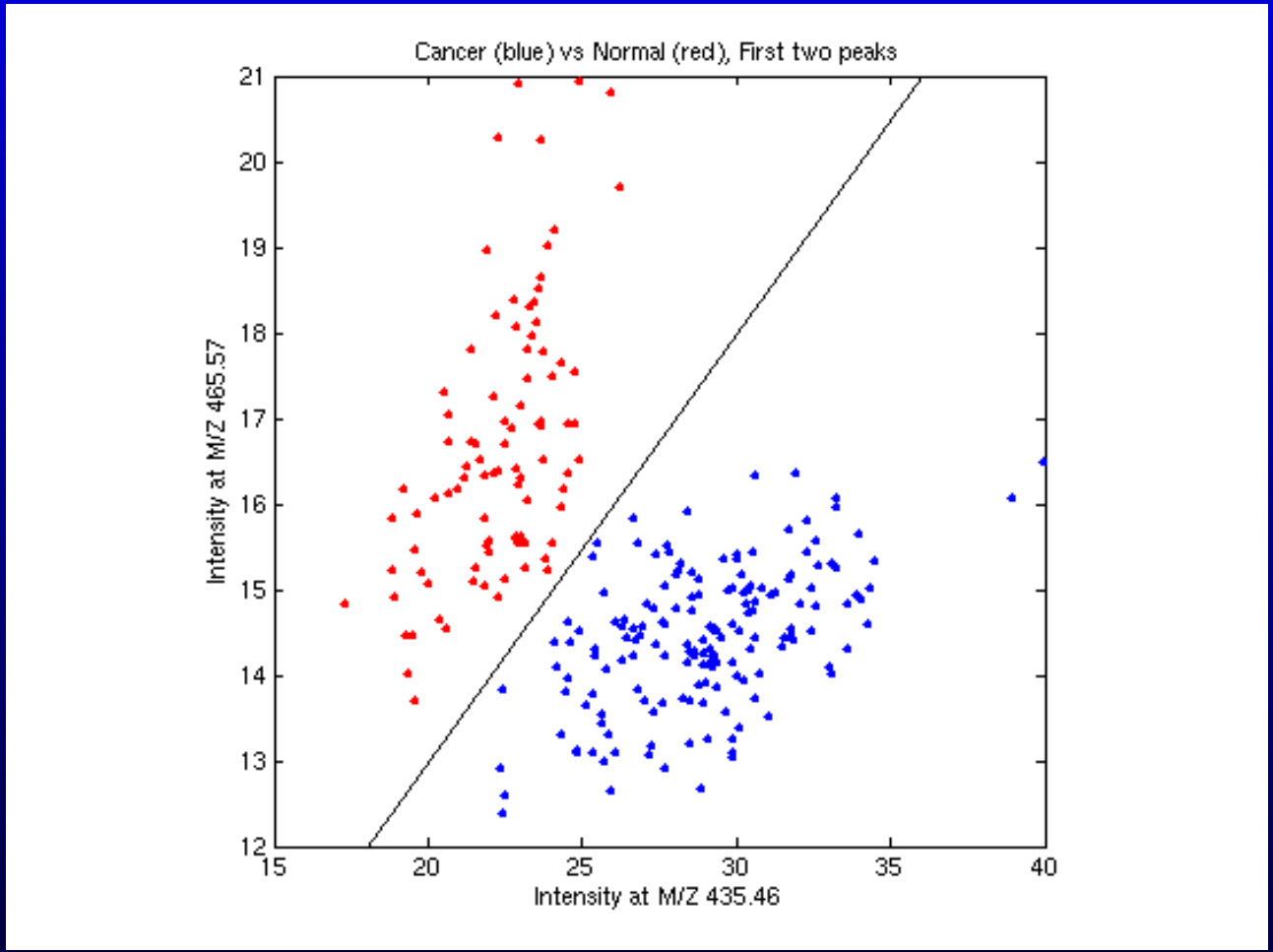


Which Peaks are Best? T-statistics



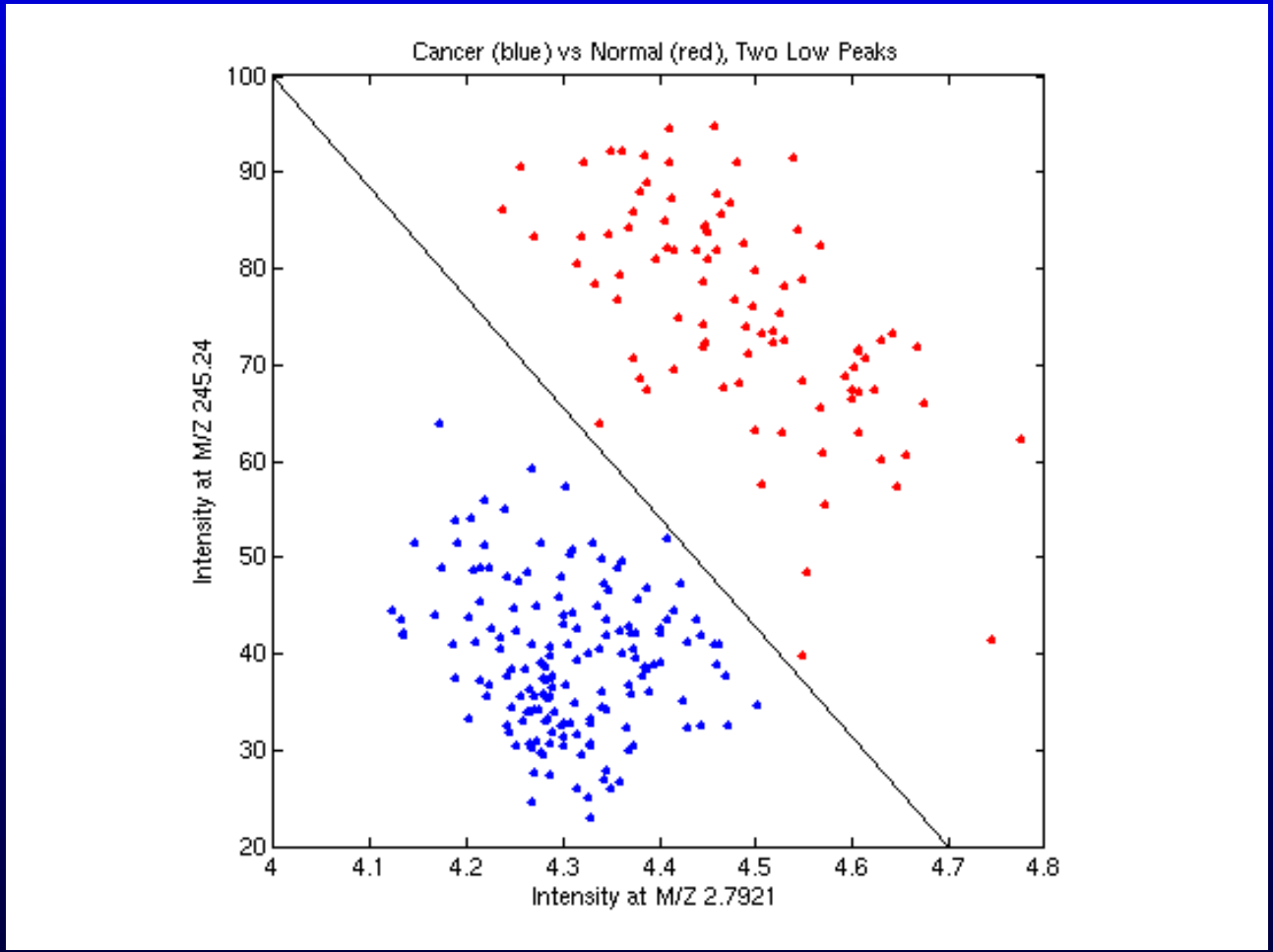
Note the magnitudes: t-values in excess of 20 (absolute value)!

One Bivariate Plot: M/Z = (435.46,465.57)



Perfect Separation. These are the first 2 peaks in their list, and ones we checked against DS2.

Another Bivariate Plot: M/Z = (2.79,245.2)



Perfect Separation, using a completely different pair. Further, look at the masses: this is the noise region.

Perfect Classification with Noise?

This is a problem, in that it suggests a qualitative difference in how the samples were processed, not just a difference in the biology.

This type of separation reminds us of what we saw with benign disease.

(Sorace and Zhan, BMC Bioinformatics, 2003)

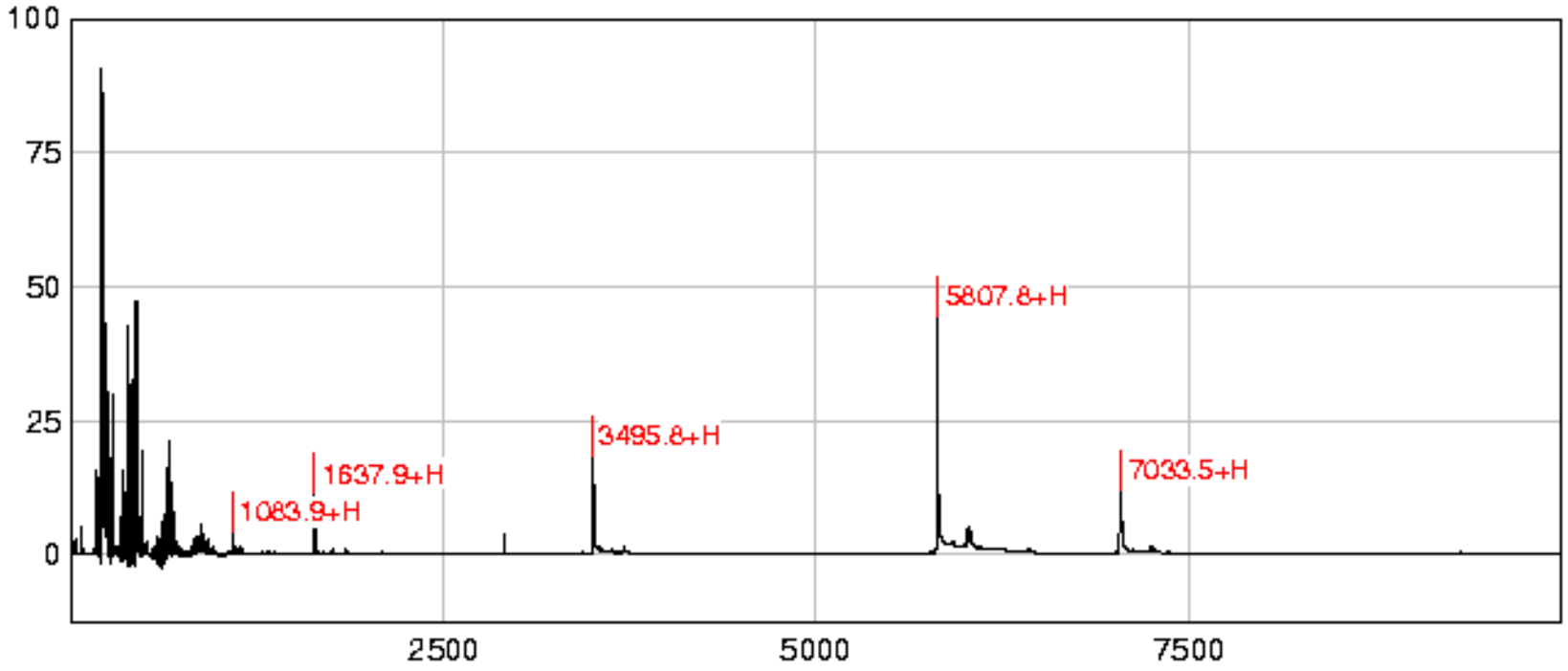
Mass Accuracy is Poor?

A tale of 5 masses...

Feb ’02 DS1	Apr ’02 DS2	Jun ’02 DS3
-7.86E-05	-7.86E-05	-7.86E-05
2.18E-07	2.18E-07	2.18E-07
9.60E-05	9.60E-05	9.60E-05
0.000366014	0.000366014	0.000366014
0.000810195	0.000810195	0.000810195

How are masses determined?

Calibrating known proteins



Calibration is the Same?

M/Z vectors the same for all three data sets.

Machine calibration the same for 4+ months?

What is the Calibration Equation?

The Ciphergen equation

$$\frac{m/z}{U} = a(t - t_0)^2 + b, \quad U = 20K, t = (0, 1, ...) * 0.004$$

Fitting it here

$$a = 0.2721697 * 10^{-3}, \quad b = 0, \quad t_0 = 0.0038$$

What is the Calibration Equation?

The Ciphergen equation

$$\frac{m/z}{U} = a(t - t_0)^2 + b, \quad U = 20K, t = (0, 1, ...) * 0.004$$

Fitting it here

$$a = 0.2721697 * 10^{-3}, \quad b = 0, \quad t_0 = 0.0038$$

These are the default settings that ship with the software!

Other issues

Prostate Cancer

Q-star data different

clinical trials?

Some teasers...

Some teasers...

We think we know the correct calibrations for DS2 and DS3.

Some teasers...

We think we know the correct calibrations for DS2 and DS3.

We think we understand why there are patterns in the electronic haze part of the DS3 spectra.

Some teasers...

We think we know the correct calibrations for DS2 and DS3.

We think we understand why there are patterns in the electronic haze part of the DS3 spectra.

We've played a bit with the Qstar spectra.

A MALDI Example: Proteomics Data Mining

41 samples, 24 with lung cancer*, 17 controls.

20 fractions per sample.

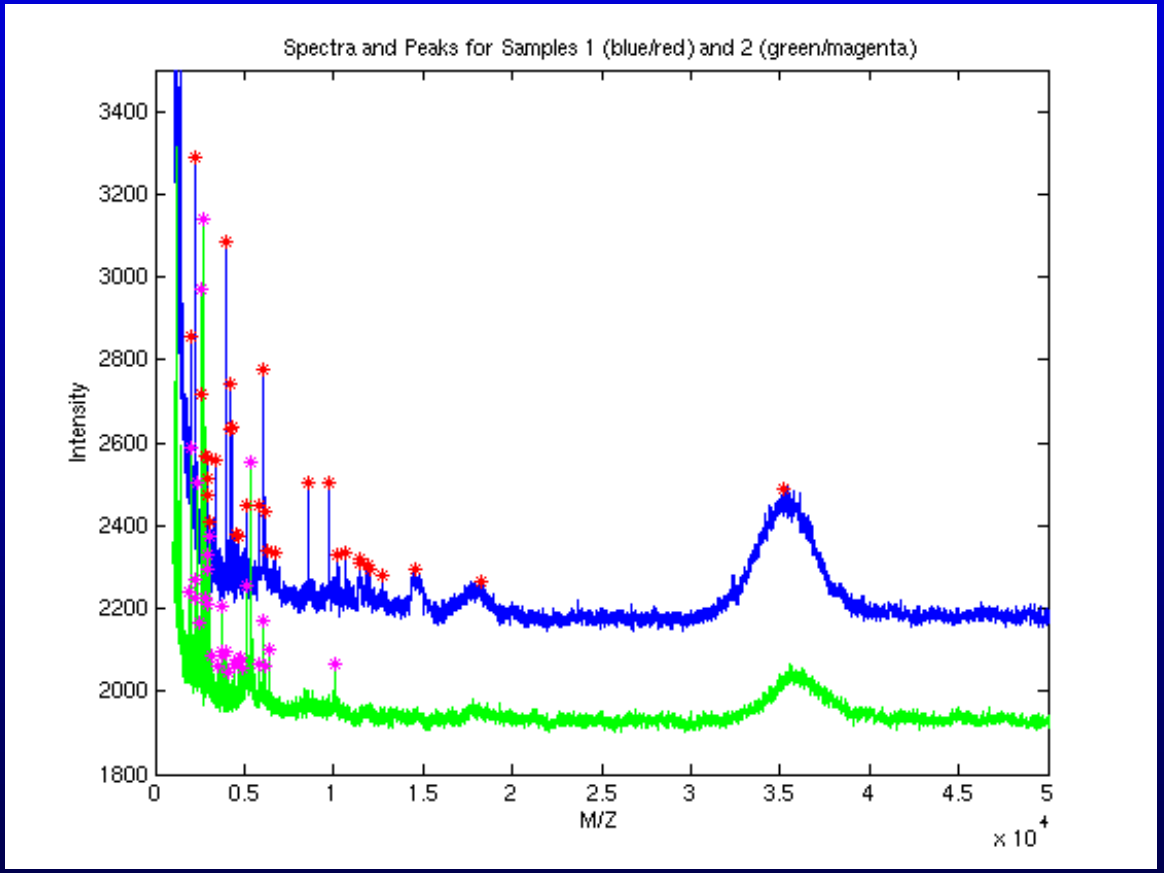
Goal: distinguish the two groups; we know this can be done due to the “zip effect”.

Data used to be at

http://www.radweb.mc.duke.edu/cme/proteomics/explain.htm

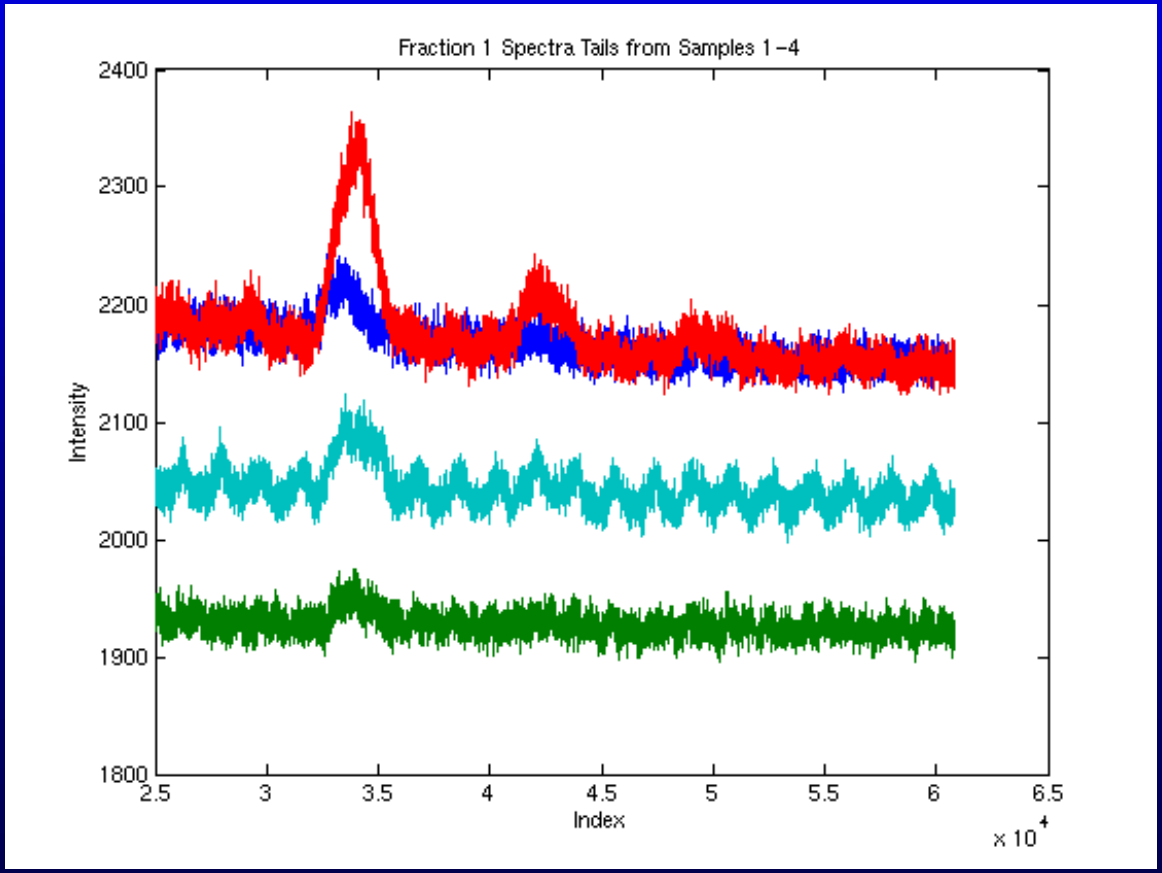
but the site has been retired. Send email to Ned Patz or Mike Campa at Duke if interested (Campa002@mc.duke.edu, patz0002@mc.duke.edu).

Raw Spectra Have Different Baselines



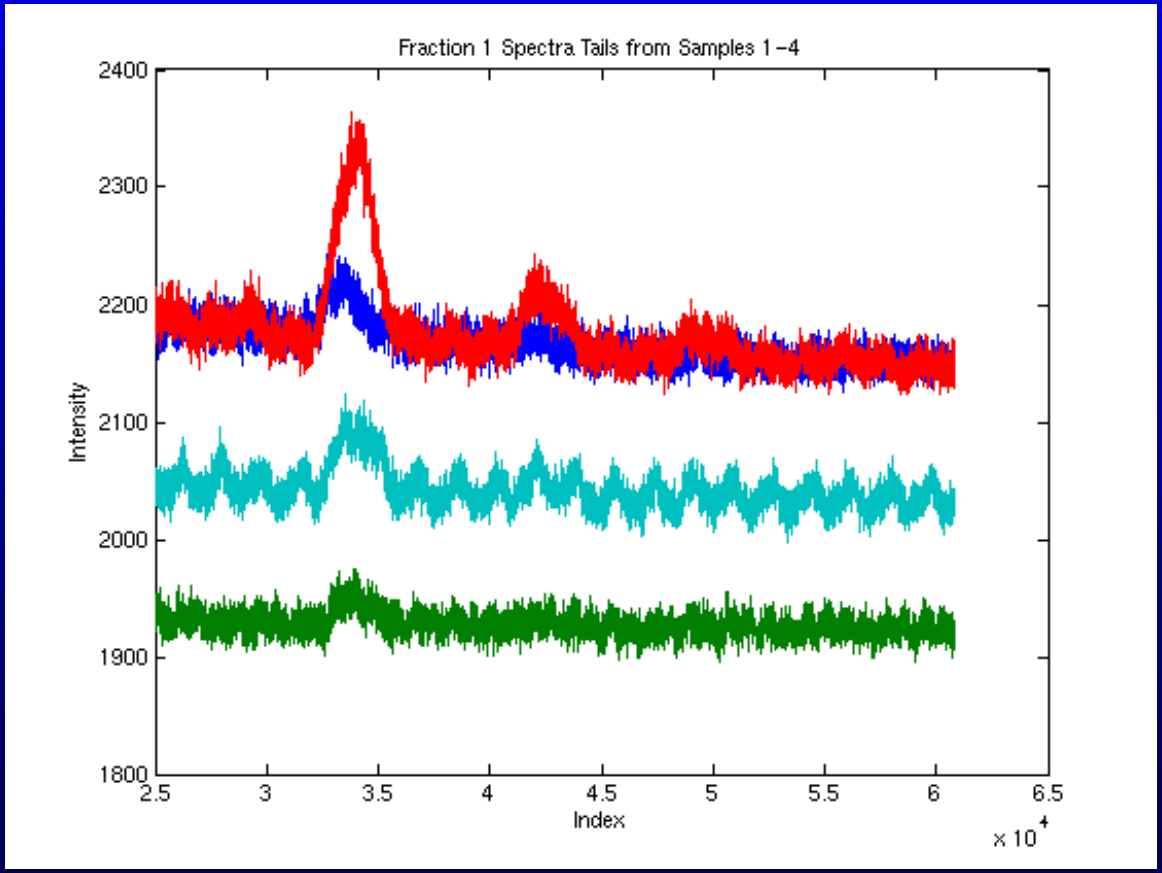
Note the need for baseline correction and normalization.

Oscillatory Behavior...



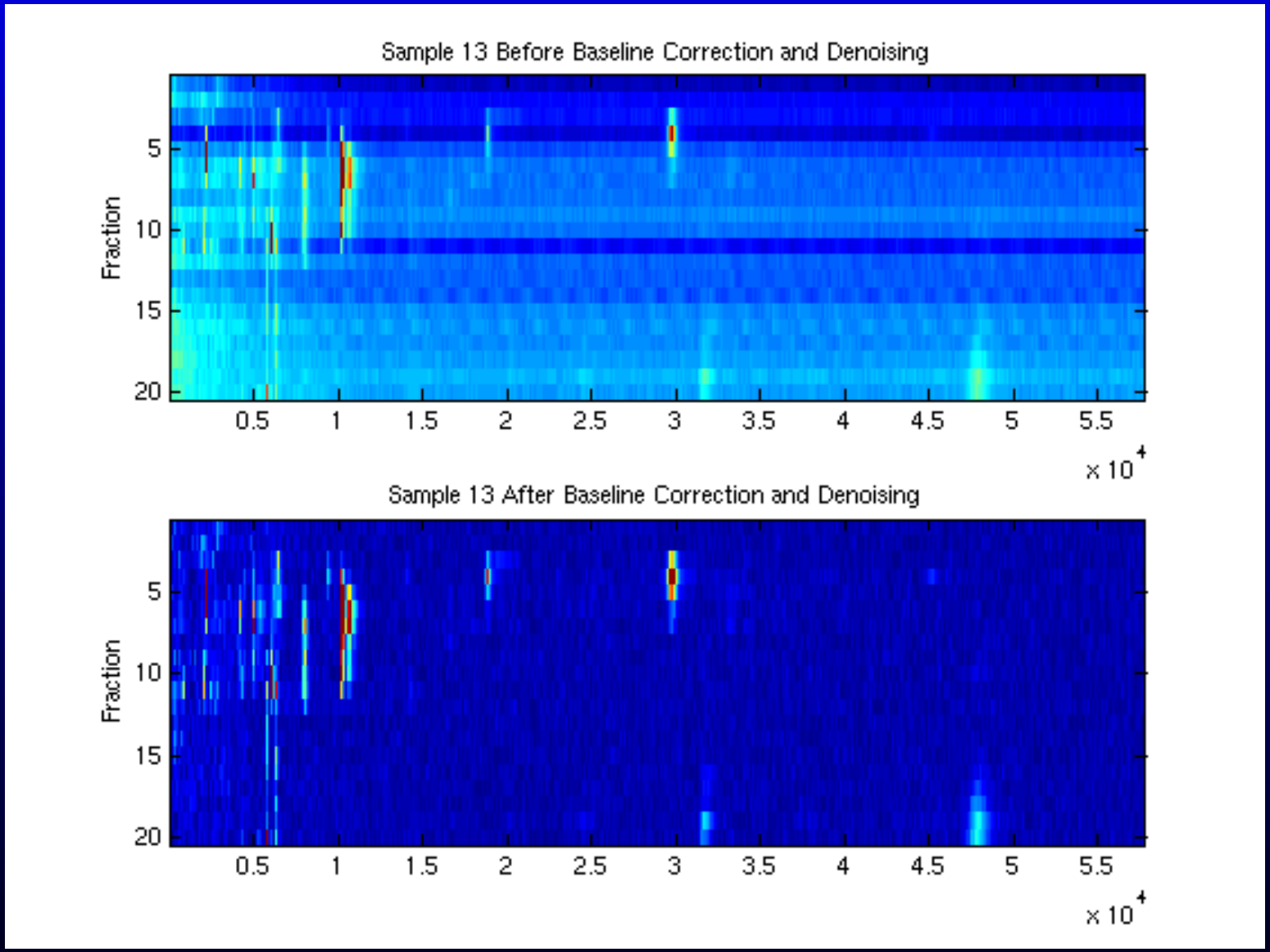
Roughly half the spectra have sinusoidal noise.

Oscillatory Behavior...

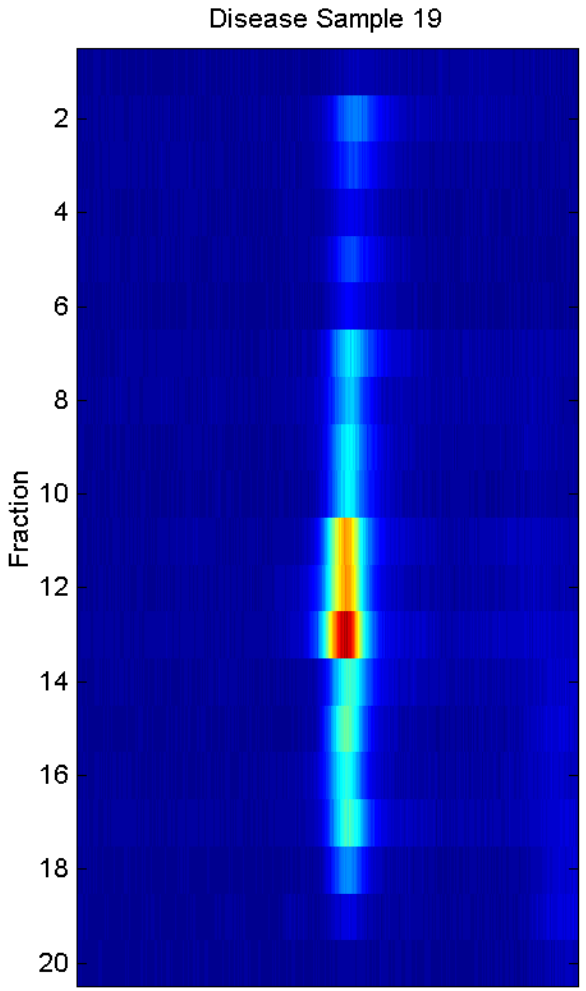


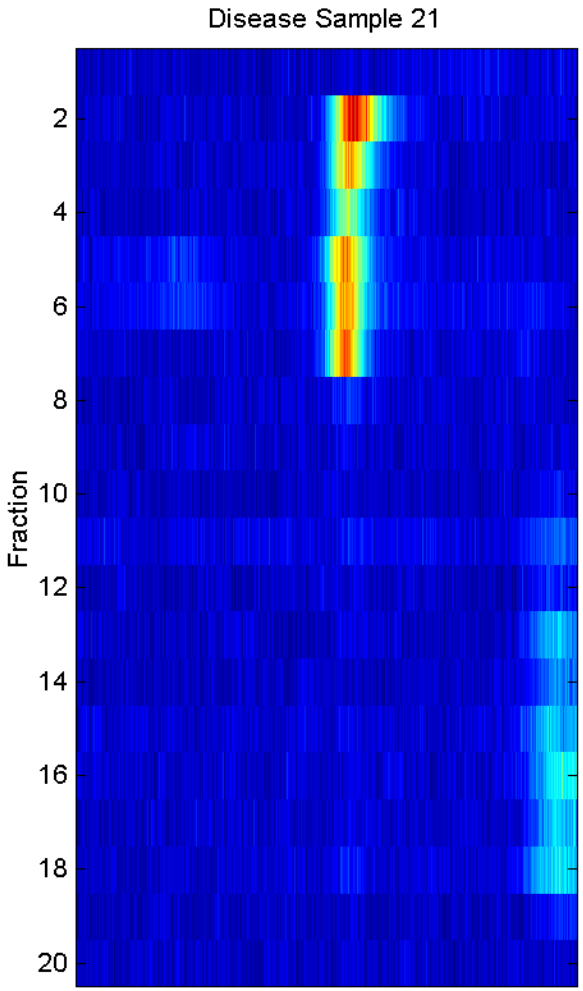
Roughly half the spectra have sinusoidal noise. We're seeing the A/C power cord.

Baseline Adj: Fraction Agreement, Before & After

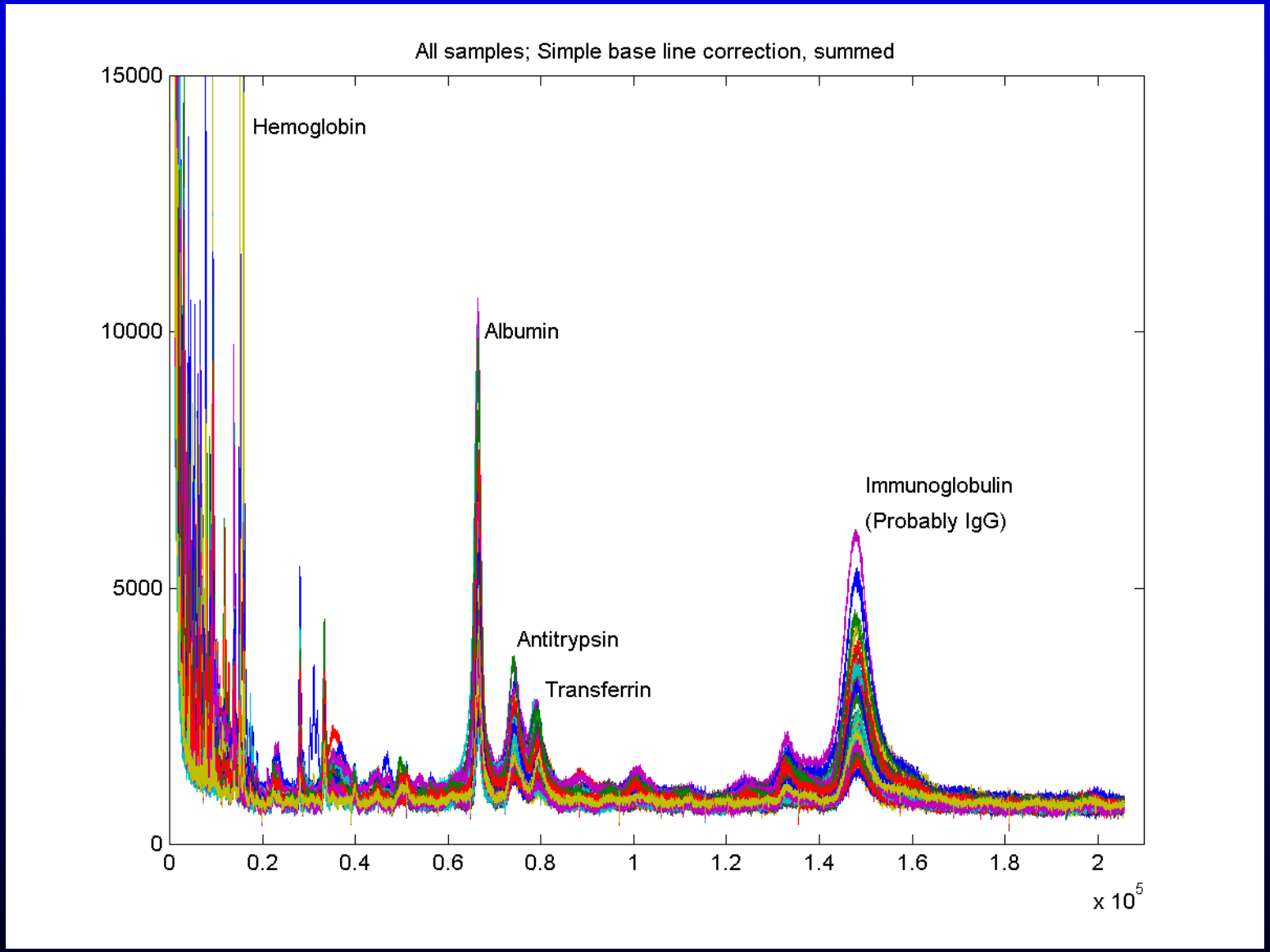


Fractionation is Unstable

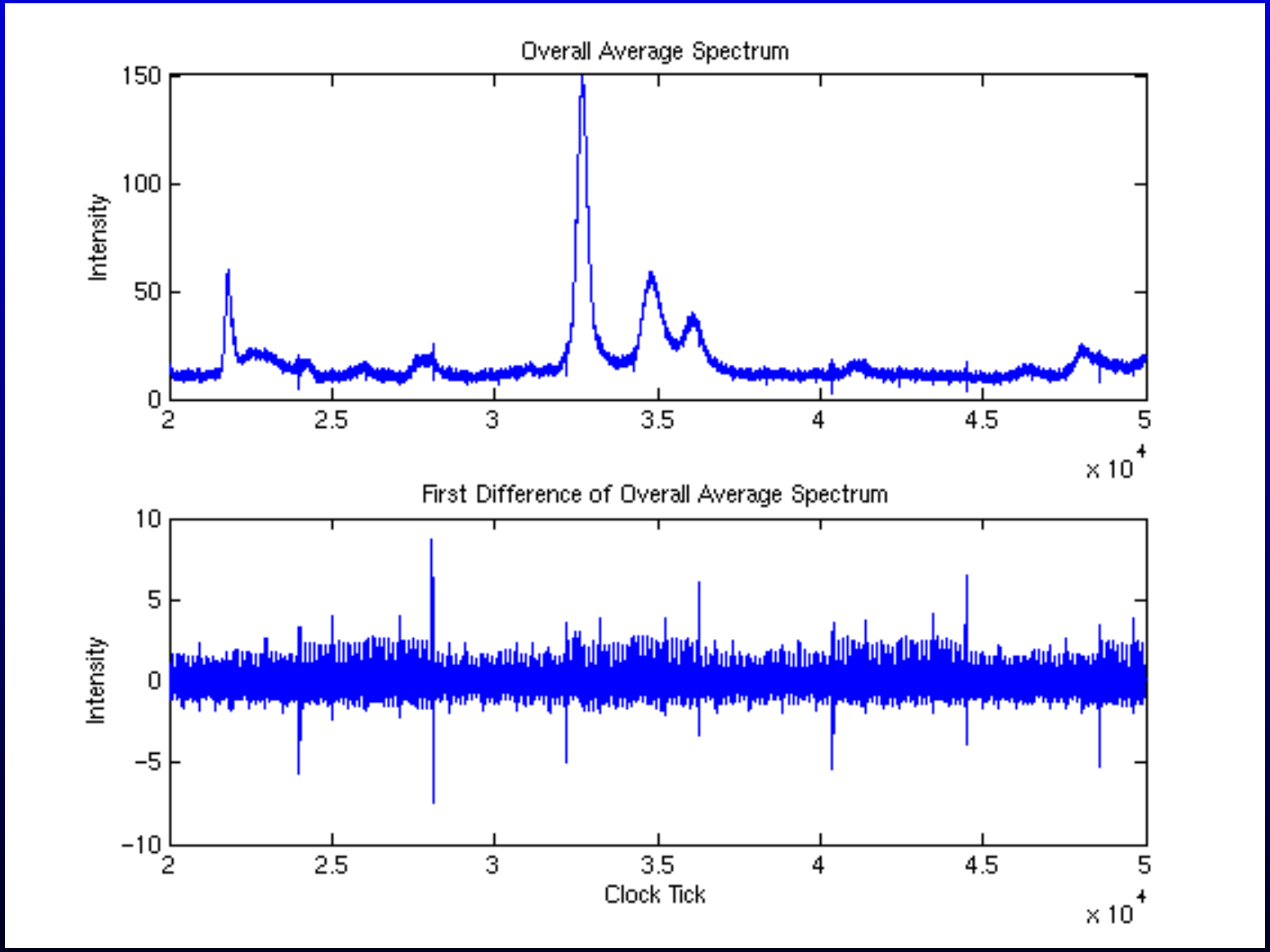




Unfractionating the Data

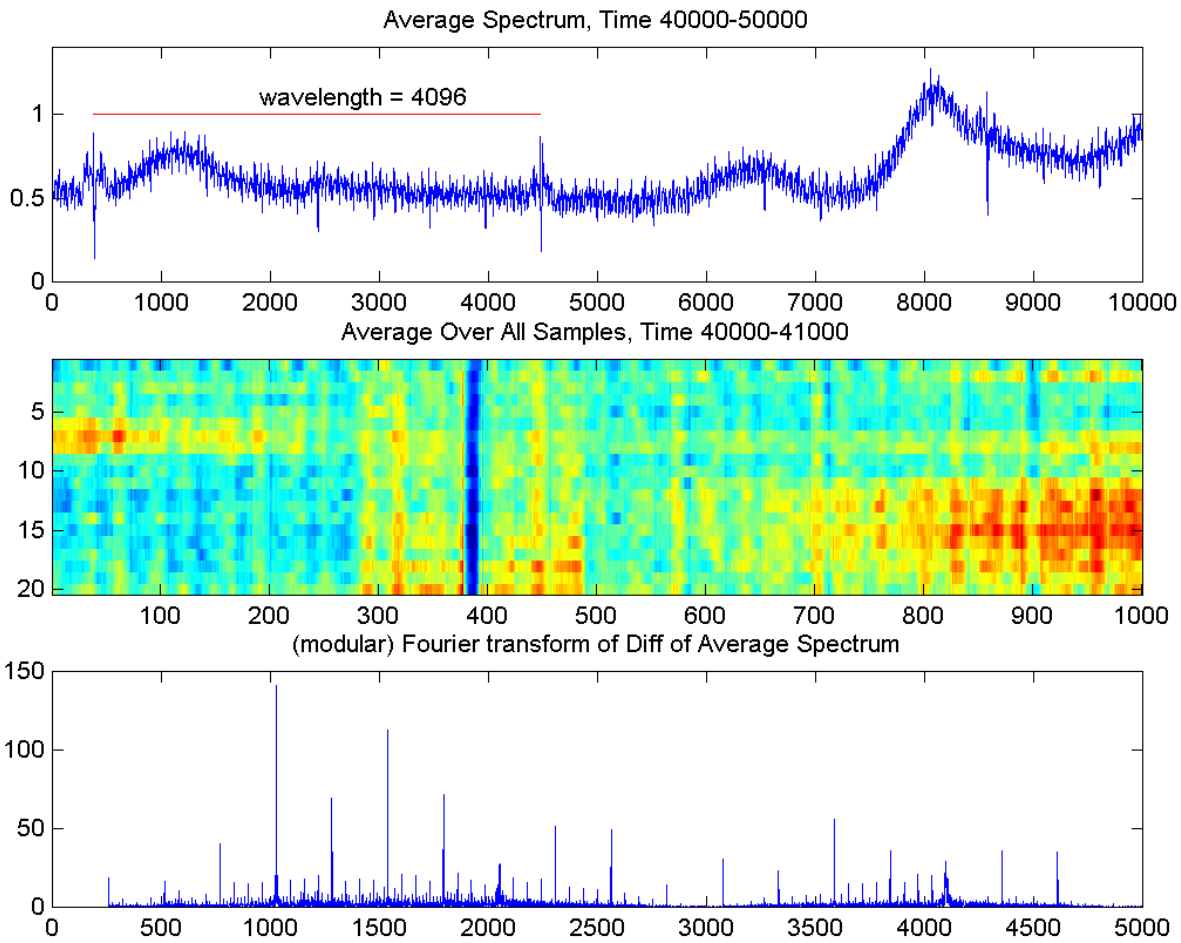


The Overall Average Shows Spikes. Difference It.



Computer Buffer?

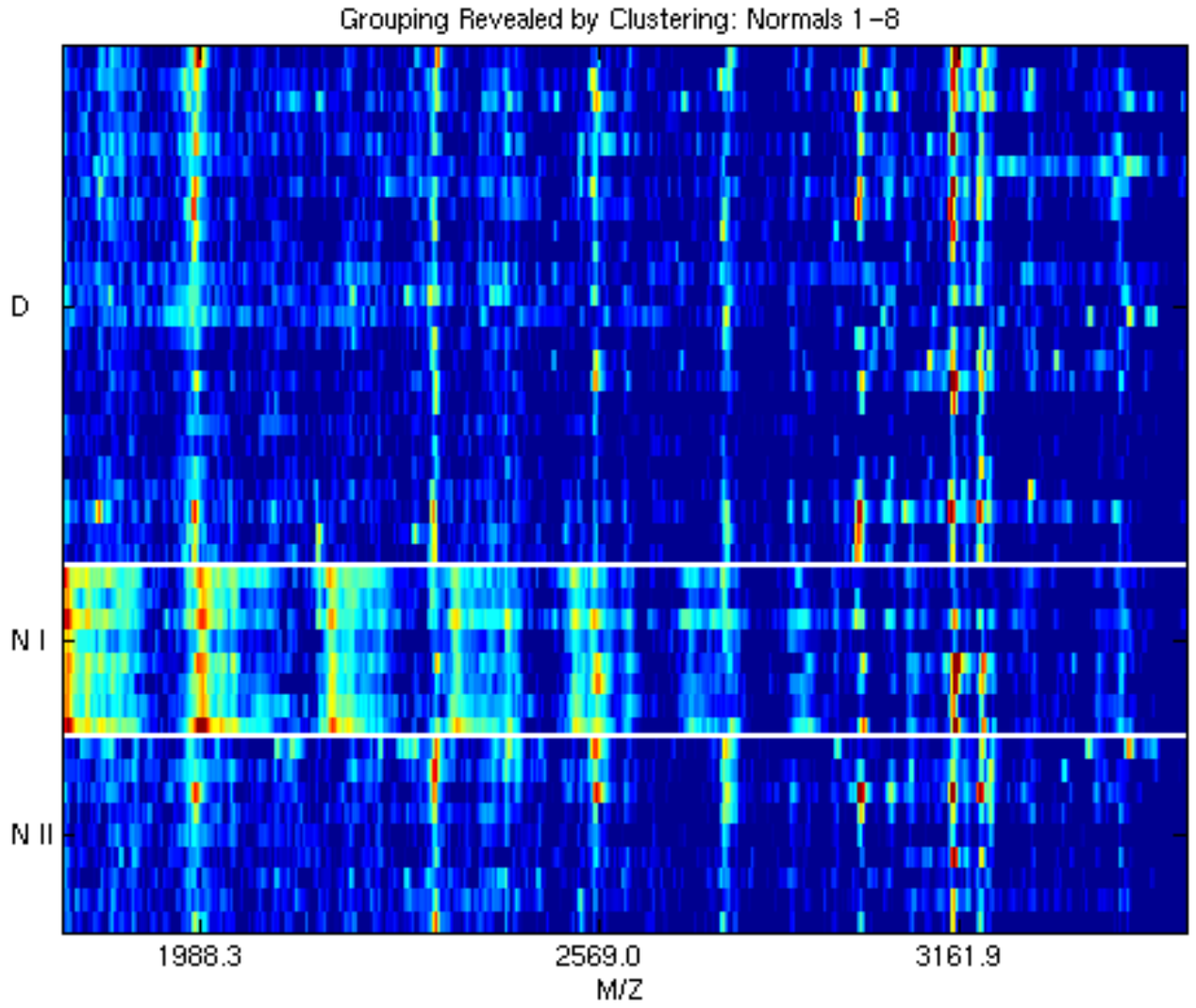
Spike spacing has a wavelength of $4096 = 2^{12}$.



Are We Done Cleaning Yet?

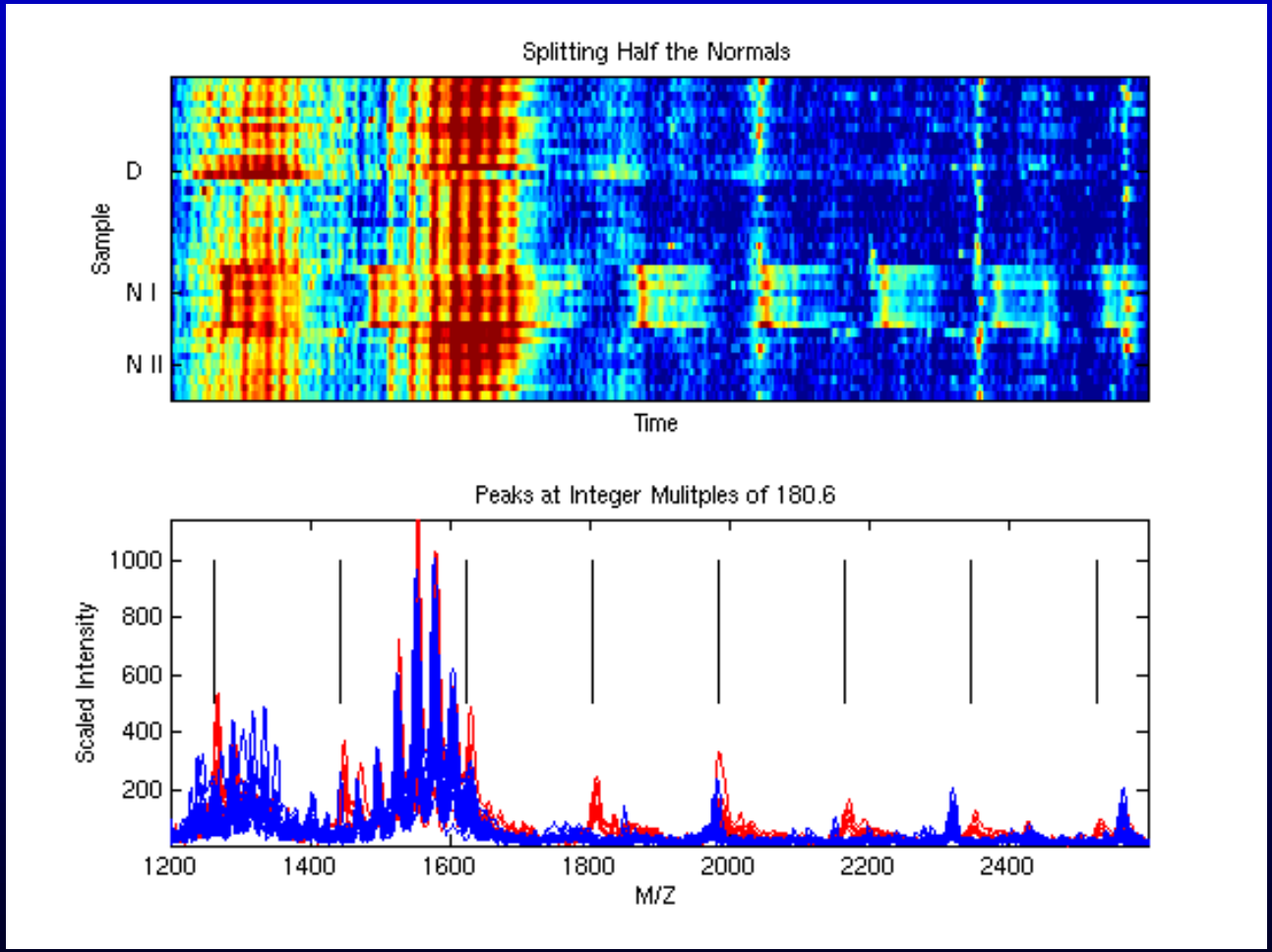
Give the problem a chance to be easy, try some simple clustering.

PCA Splits off Half the Normals



Peaks at Integer Multiples of M/Z 180.6!

This suggests a polymer. No Amino Acid dimers fit.



Cleaning Redux

- Baseline Correction and Normalization
- Inconsistent Fractionation
- Computer Buffers
- Polymers in some Normal Spectra
- Peak Finding (Use Theirs)

Data reduced to 1 spectrum/patient, with 506 peaks per spectrum.

Find the Best Separators

Peaks	MD	P-Value	Wrong	LOOCV
12886	2.547	≤ 0.005	11	11
8840, 12886	5.679	≤ 0.01	5	6
3077, 12886 74263	9.019	≤ 0.01	3	4
5863, 8143 8840, 12886	12.585	≤ 0.01	3	3
4125, 7000 9010, 12886 74263	23.108	≤ 0.01	1	1

There are 9 values that recur frequently, at masses of 3077, 4069, 5825, 6955, 8840, 12886, 17318, 61000, and 74263.

P-values are not from table lookups!

Testing Reality (Significance)

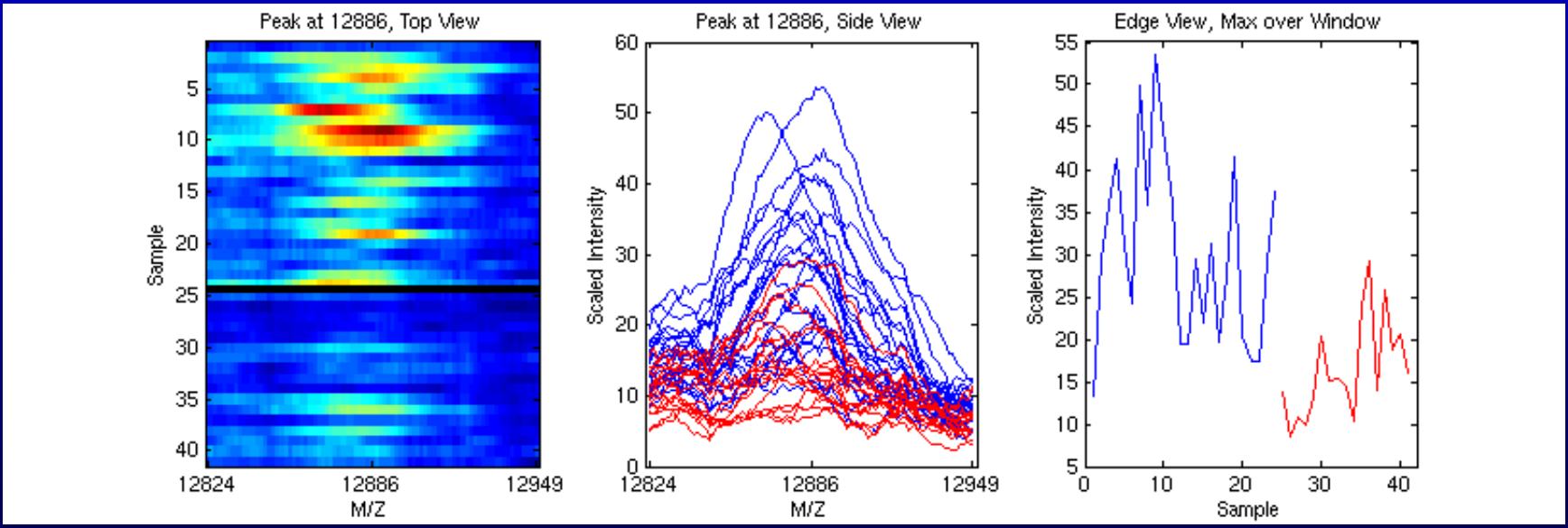
Generate a bunch of “random noise” data matrices, each 41×506 in size.

For each matrix, split the 41 noise “samples” into groups of 24 and 17.

Repeat our search procedure on the random data, and see how well we can separate things.

The Eyeball Test

We applied one last filtering step and actually *looked* at the regions identified. All 9 peaks listed above passed the eye test.



Blue lines = Cancers

Red lines = Controls

Other Stuff

We were the only ones to notice the sinusoidal noise.

Other Stuff

We were the only ones to notice the sinusoidal noise.

and the clock tick.

Other Stuff

We were the only ones to notice the sinusoidal noise.

and the clock tick.

and they were fixing the fractionation machine, and looking at the other stuff.

Other Stuff

We were the only ones to notice the sinusoidal noise.

and the clock tick.

and they were fixing the fractionation machine, and looking at the other stuff.

and they gave us a nice shiny plaque!

The Deluge at MDA

Brain Cancer

Bladder Cancer

Leukemia

Pancreatic Cancer

Breast Cancer

Several show real structure, several show processing effects.

“If you’re not working on a proteomics project, you will be soon!”
Kevin Coombes to Bioinf section at MDA

The Punchlines

- There is no magic bullet here. (Sorry.)
- Data preprocessing is extremely important with this type of data, and there is still much room for improvement.
- Use Simple Tests and Pictures
- Insist on Good Experimental Design
- There is structure in this data and it can be found!

Our Own Reports

On the *Lancet* data:

Baggerly, Morris and Coombes (2004), *Bioinformatics*, **20(5)**:777-785.

On the Proteomics Data Mining Conference Data:

Baggerly, Morris, Wang, Gold, Xiao and Coombes (2003), *Proteomics*, **3(9)**:1677-1682.

More methodology:

Coombes et al (2003), *Clinical Chemistry*, **49(10)**:1615-1623.

pdf preprints are available.

Partners in Crime

Kevin Coombes, Jeff Morris

———-

Jing Wang, David Gold, Lian-Chun Xiao

———-

Ryuji Kobayashi, David Hawke, John Koomen

Maybe I Should Keep My Day Job?

Consistent Structure in DS2/DS3?

Recently, Zhu et al. (PNAS 2003,v100:14666-71) noted that they could find a set of m/z values that separated cancers from normals in both DS2 and DS3.

Used local smoothing (Gaussian kernel) and t-statistics to identify useful peaks in DS2. Keep only those with t-stats exceeding a certain magnitude threshold, 4.22, found using random field theory.

Final list of 18 m/z values.

We Were Somewhat Surprised...

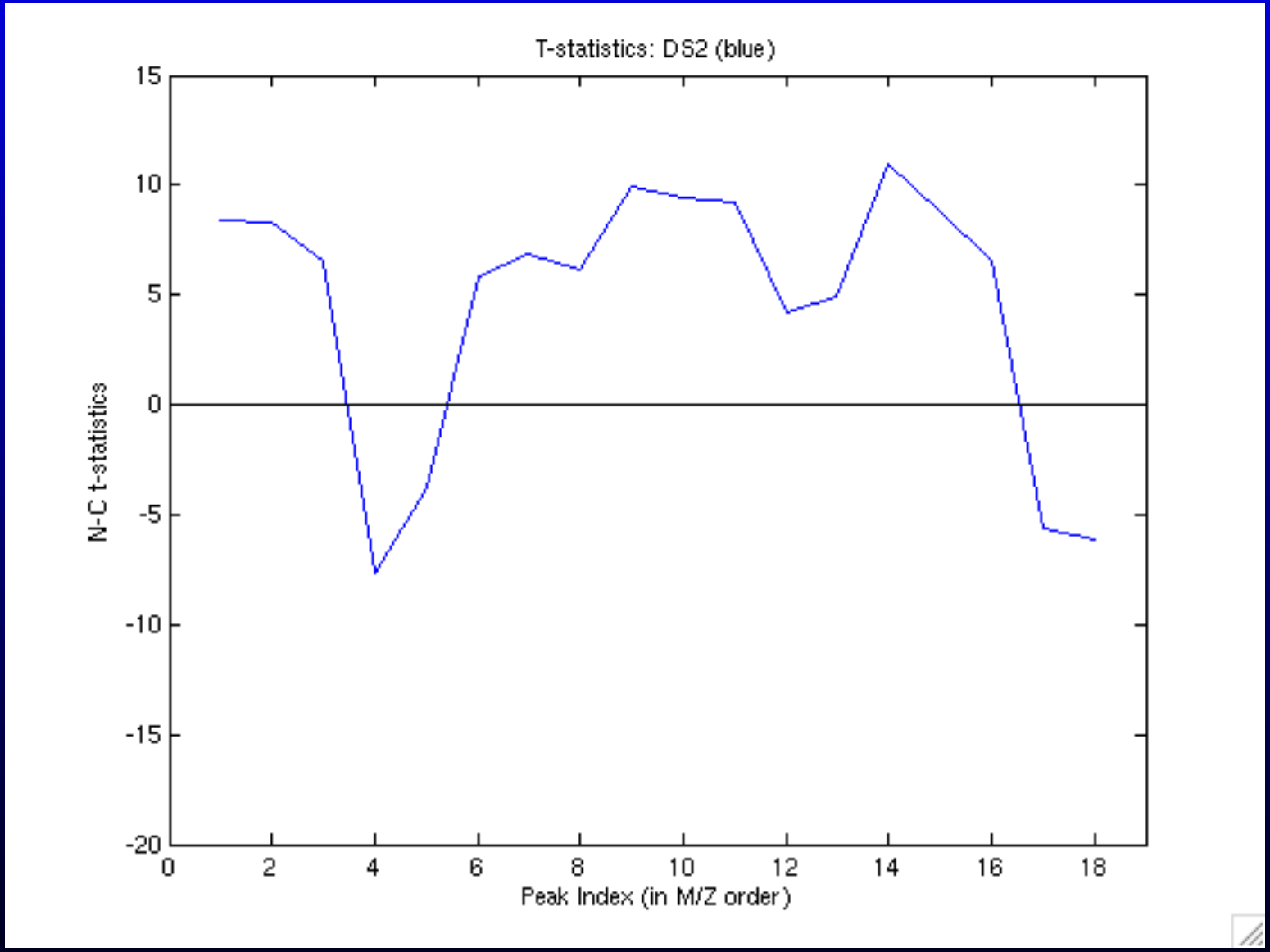
10 of the 18 m/z values are less than 500 Da.

DS2 is baseline subtracted, and DS3 is not.

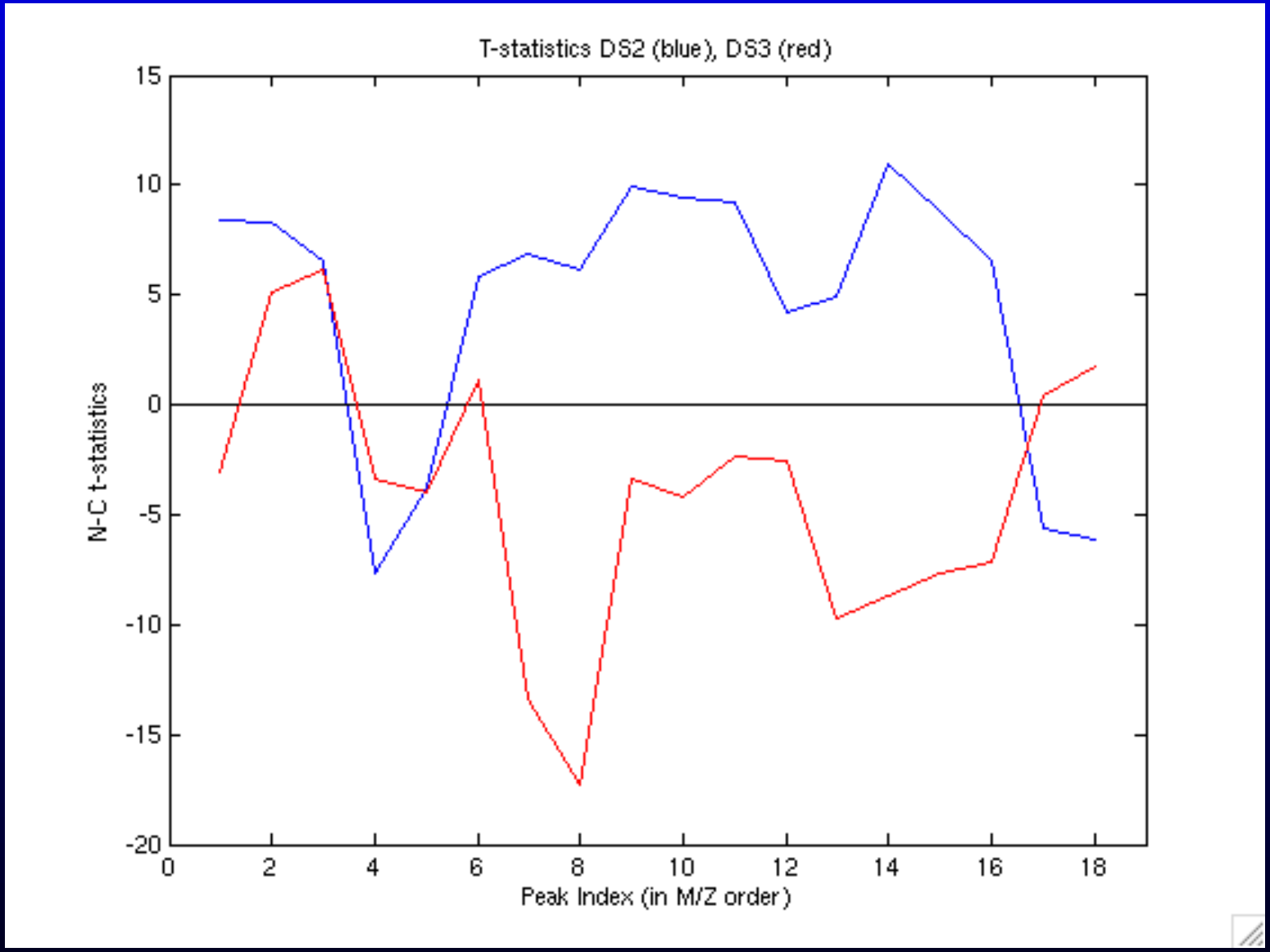
DS2 is offset relative to DS3.

If the same nominal m/z values are used, we would think they are finding different proteins.

What Are the T-Statistics?



13/18 Flip Sign?!?



What Do Sign Flips Mean?

If intensities were higher in cancer in DS2, they're higher in normals in DS3.

This type of qualitative reversal should not be present in a consistent biological signature.

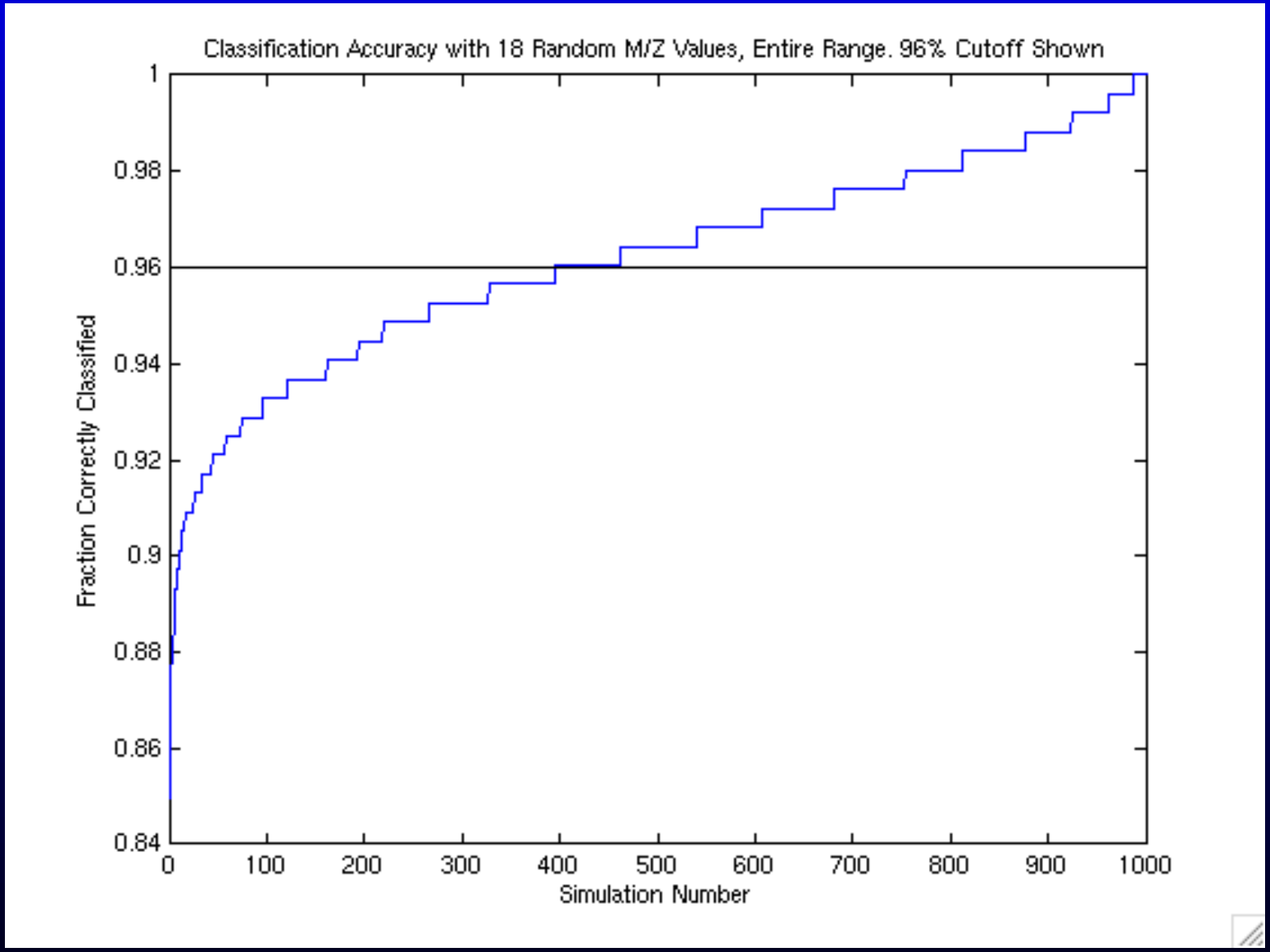
Whither Structure? A Quick Experiment

Question: Why does structure found in DS2 persist in DS3?

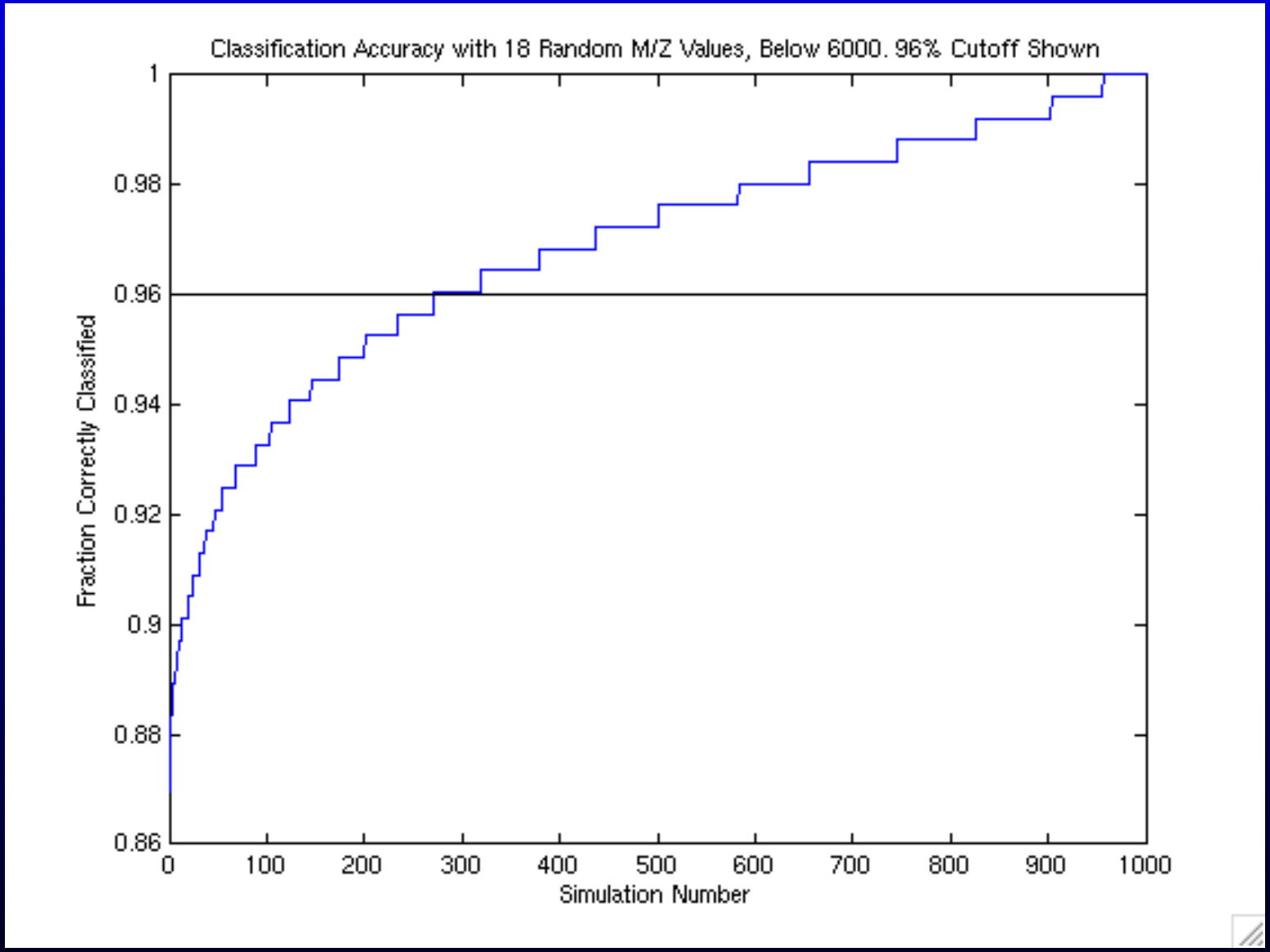
Observation: DS3 is easy to classify.

Addressable Question: is it so easy that we might expect to get good classifications without biology?

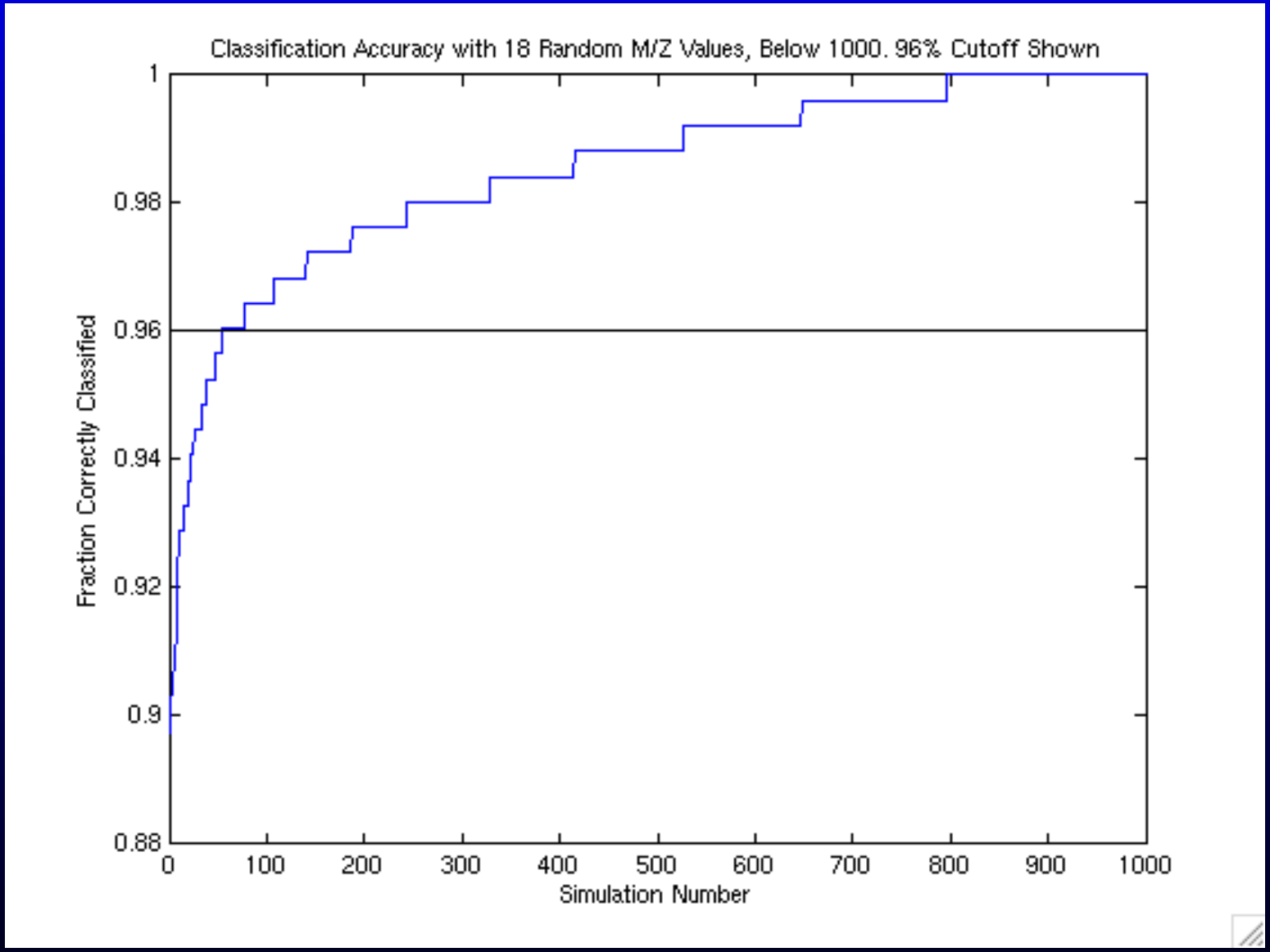
Choosing Random M/Z's... Entire Range



Choosing Random M/Z's... up to 6000 Da



Choosing Random M/Z's... up to 1000 Da



We Don’t Think It’s Biology

We shouldn’t be able to do this well with random m/z values.

If we go back to 2.79 (which we initially commented on), there’s nothing else close by and the signal is very weak.

We Don’t Think It’s Biology

We shouldn’t be able to do this well with random m/z values.

If we go back to 2.79 (which we initially commented on), there’s nothing else close by and the signal is very weak.

So, how do we explain this?

Some Sample Info About DS3

Sample	SELDI Chip ID	A
DAF-0601	454-BZ818-WCX2-D	
DAF-0602	455-BZ824-WCX2-E	
DAF-0604	456-BZ821-WCX2-B	
DAF-0605	457-BZ815-WCX2-H	
DAF-0606	454-BZ818-WCX2-G	
DAF-0608	456-BZ821-WCX2-G	
DAF-0609	457-BZ815-WCX2-F	

Ovarian Sample Info 8-7-02.xls

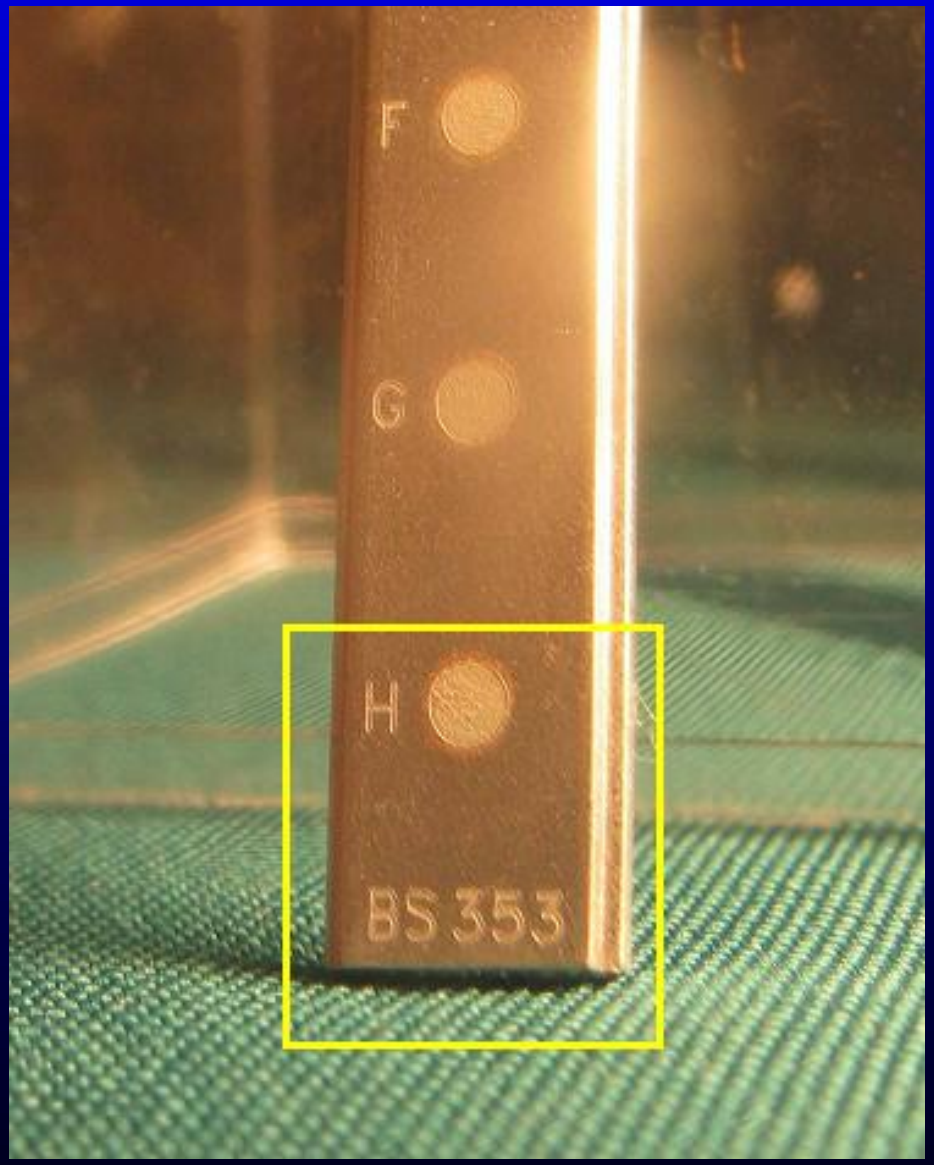
Some Sample Info About DS3 (Sorted)

Sample	SELDI Chip ID	A
DAF-0688	430-CB533-WCX2-C	
DAF-0742	430-CB533-WCX2-D	
DAF-0769	430-CB533-WCX2-F	
DAF-0770	430-CB533-WCX2-H	
DAF-0771	431-CB481-WCX2-A	
DAF-0772	431-CB481-WCX2-B	
DAF-0687	431-CB481-WCX2-D	

Ovarian Sample Info 8-7-02.xls (Sorted)

Interpreting 430-CB533-WCX2-C

Interpreting 430-CB533-WCX2-C



Interpreting 430-CB533-WCX2-C

Some Cancer File Names from June

an_Dataset_6-19-02.zip

an	e	430-cb533-wcx2-c.csv
an	f	430-cb533-wcx2-d.csv
an	f	430-cb533-wcx2-f.csv
an	f	430-cb533-wcx2-h.csv
an	f	431-cb481-wcx2-a.csv
an	f	431-cb481-wcx2-b.csv
an	e	431-cb481-wcx2-d.csv

Prefix numbers from 430 to 465

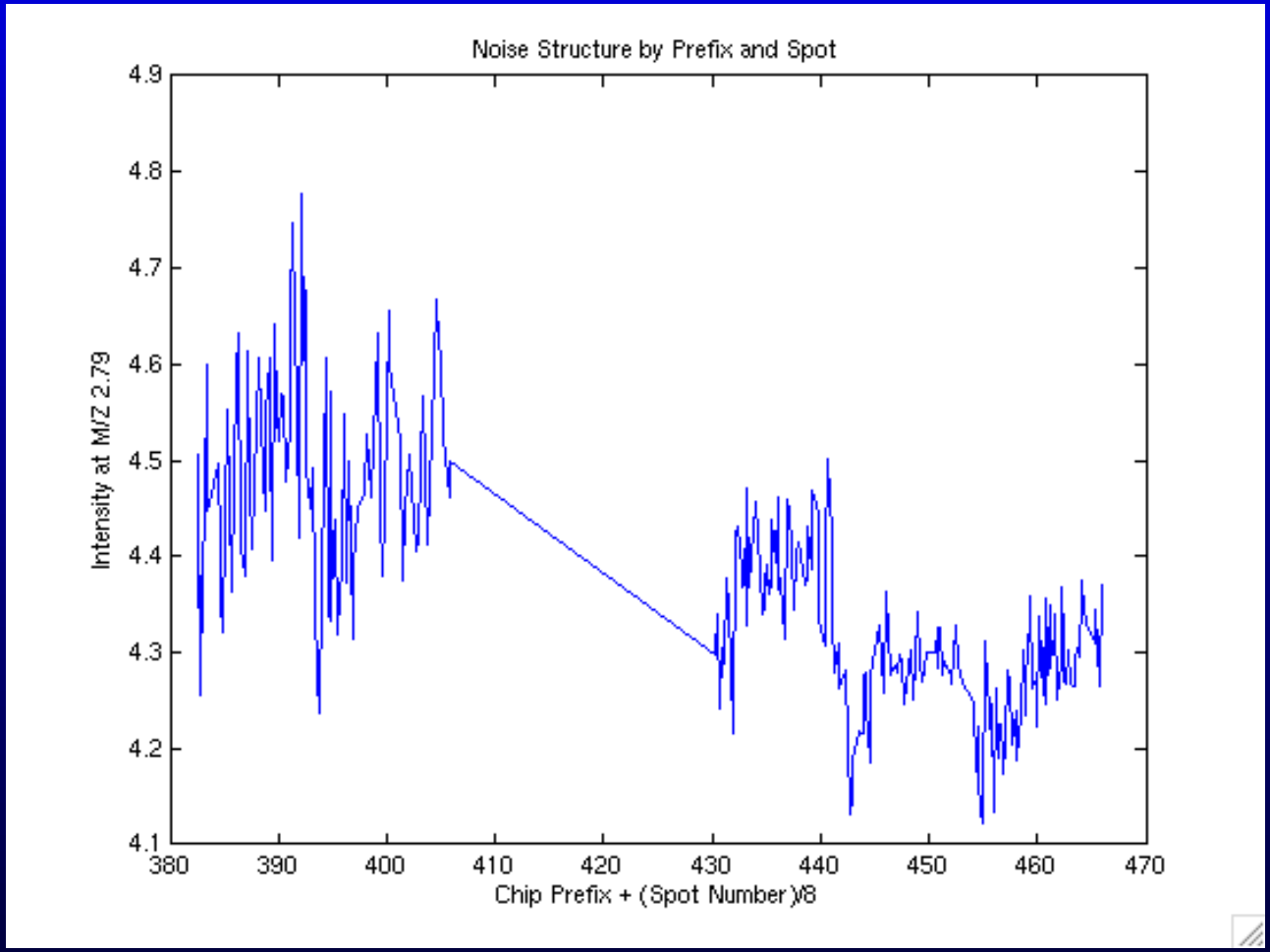
Some Normal File Names from June

arian_Dataset_6-19-02.zip

trol	d	382-ca602-wcx2-d.csv
trol	c	382-ca602-wcx2-e.csv
trol	c	382-ca602-wcx2-g.csv
trol	d	382-ca602-wcx2-h.csv
trol	d	383-ca564-wcx2-b.csv
trol	c	383-ca564-wcx2-c.csv
trol	c	383-ca564-wcx2-d.csv

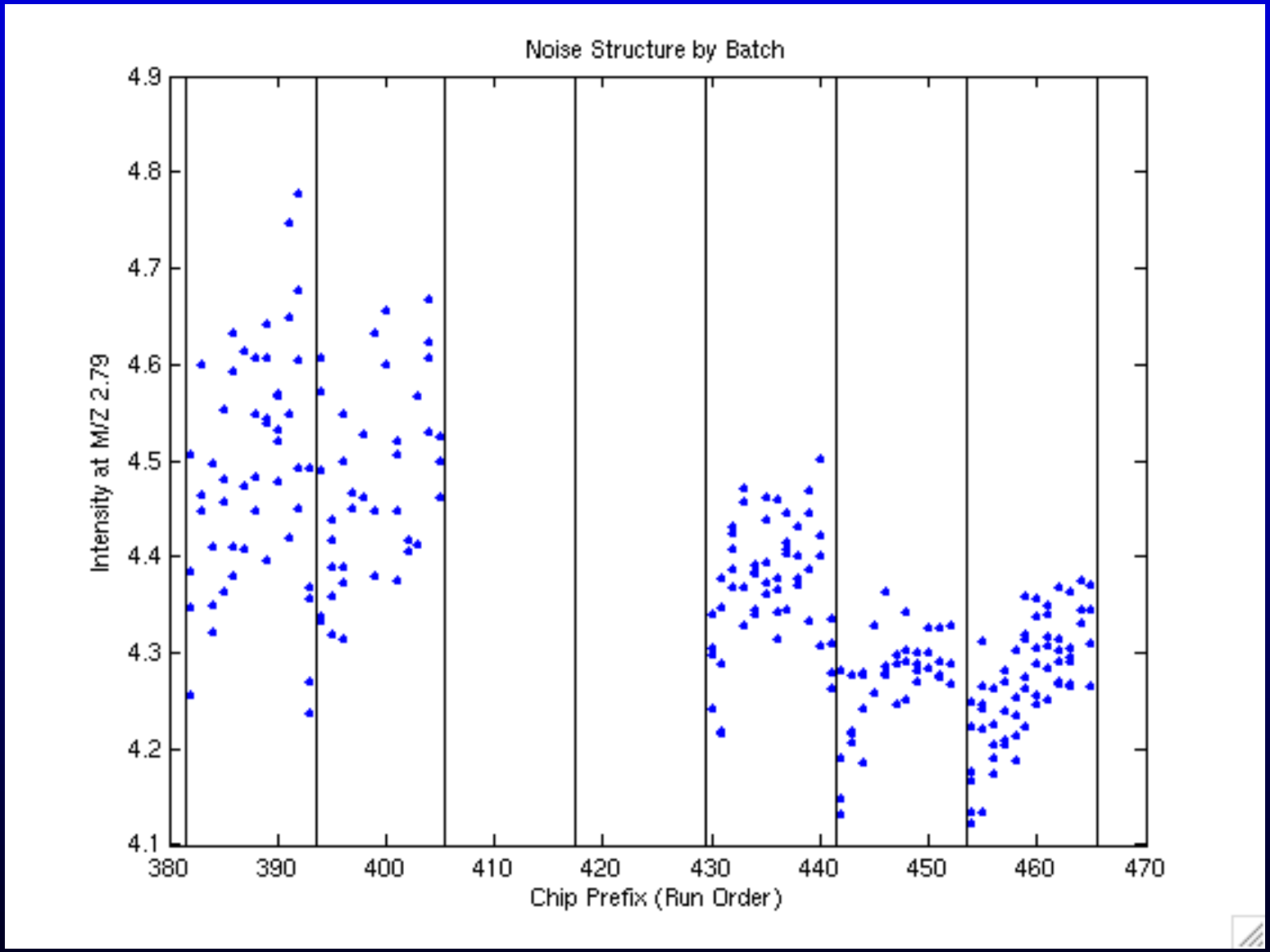
Prefix numbers from 382 to 405. No overlap!

Plotting Noise by Run Order



seems to be some “blocks” of structure

Batches of 12?



Why is 12 Important?

Why is 12 Important?

"... chip arrays should be processed in parallel using robotic liquid handling to minimize variance ... and modified to make use of a ProteinChip array bioprocessor (Ciphergen Biosystems Inc.). *The bioprocessor holds 12 ProteinChips*, each having 8 chromatographic 'spots', allowing 96 samples to be processed in parallel".

– Petricoin et al. (2004), J. Prot. Res., to appear in March/April issue on Proteomics and Disease.

Processing Spectra

We typically want to partition a spectrum into 3 components:

Low-frequency baseline, associated with matrix

A local estimate of noise

A set of peaks, representing the real signal

This requires

Denoising,
Baseline Subtraction,
Peak Indentification and Quantification

How Have Others Addressed Peak Finding?

Ciphergen: Fit a piecewise linear convex hull to the curve for baseline, and look for maxima meeting a minimum S/N threshold

Yasui et al: Give up on quantification. Fit a supersmoother to the data, and flag maxima in a 20 tick window if the maxima exceed the overall median

Problems: lots of spurious peaks

What's Hard About Peak Finding?

Peaks have different shapes depending on where they are in the spectra – higher mass peaks are wider.

This follows from two things:

1) smearing due to initial velocity distribution

2) physical spread due to isotopic differences

What New Tool do We Use?

The Undecimated Discrete Wavelet Transform (UDWT)

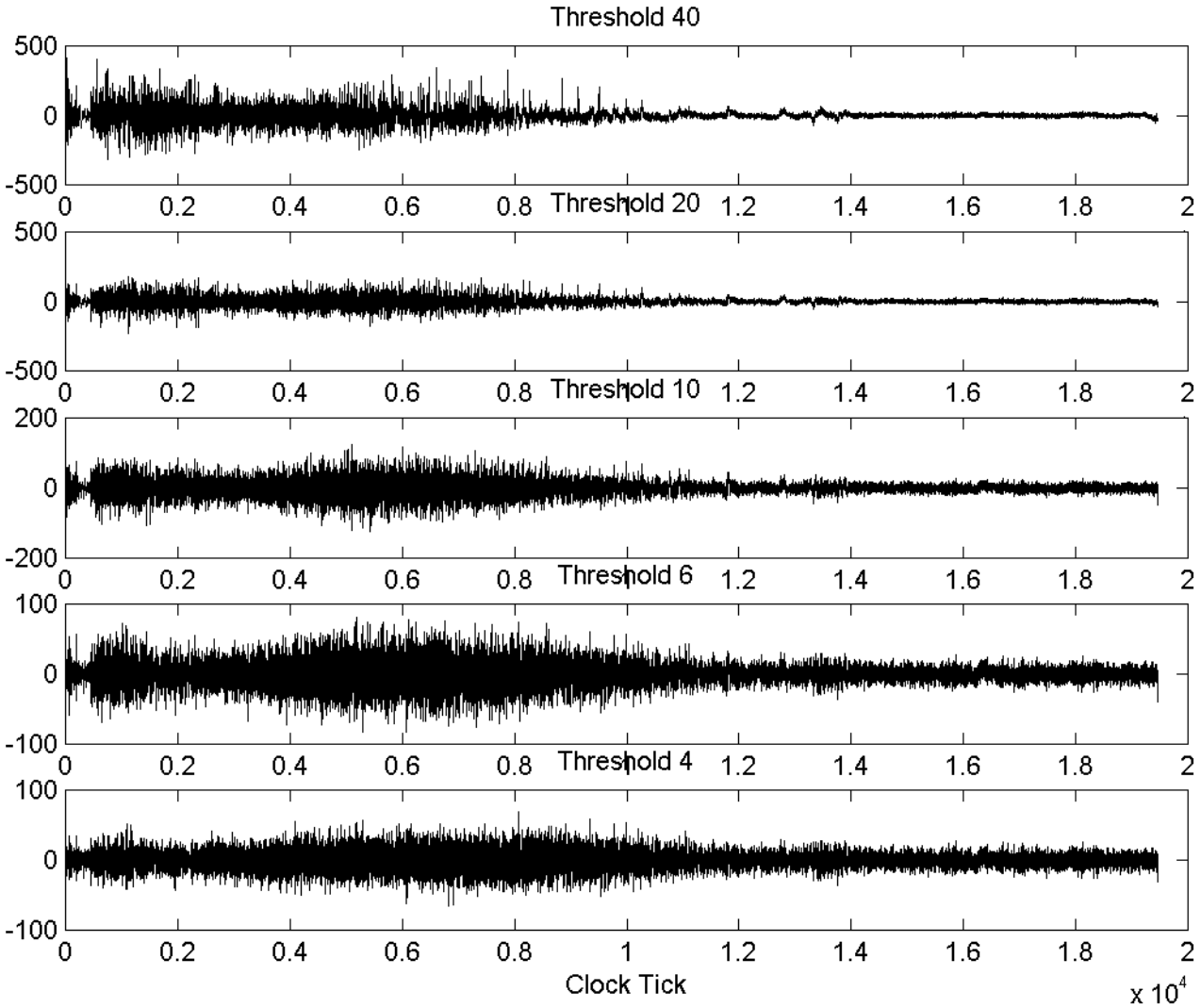
Captures the multiscale nature of the data

cycling removes boundary-driven artifacts

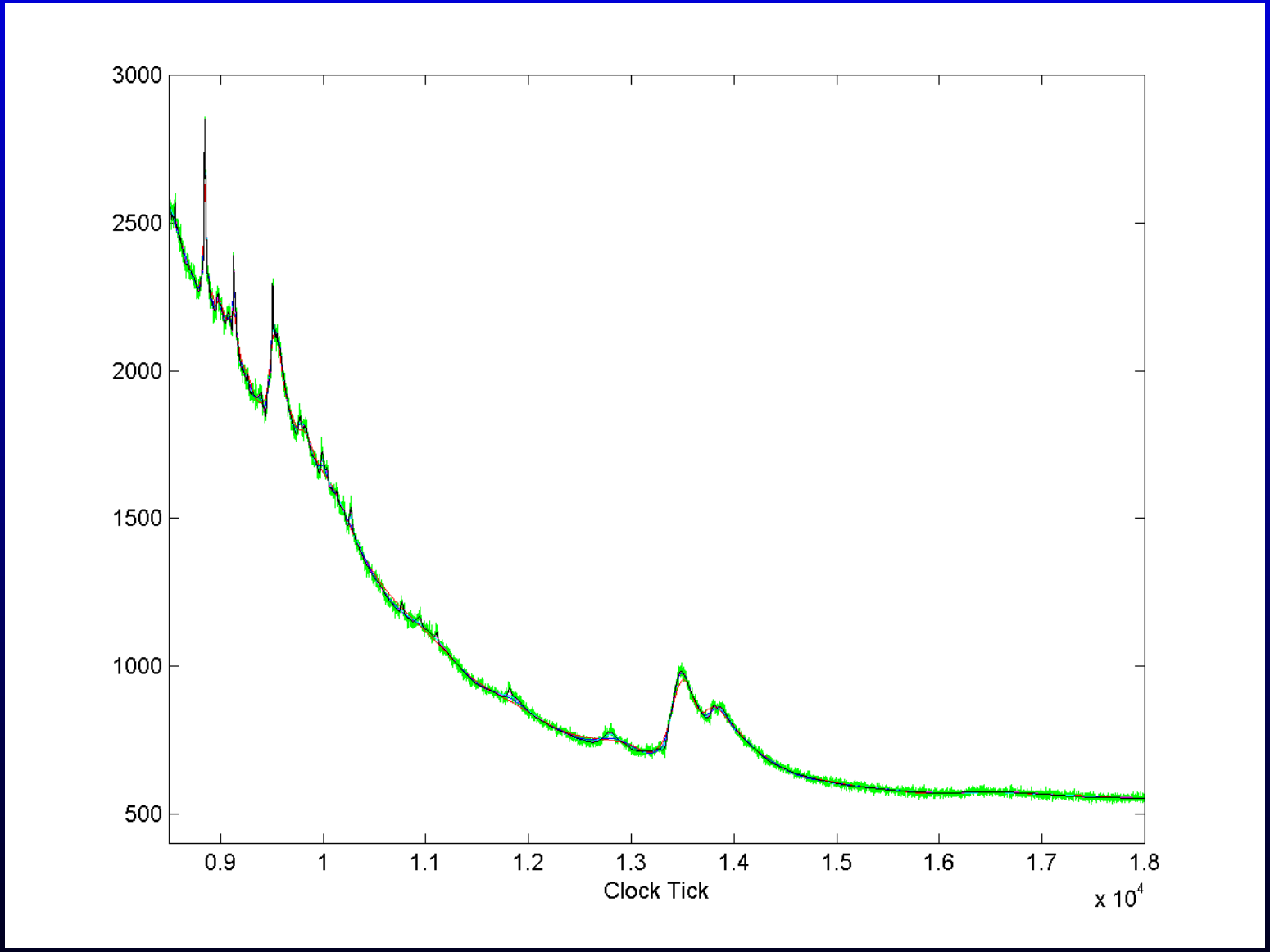
fit a local minimum curve after denoising for baseline

Need to choose the wavelet threshold

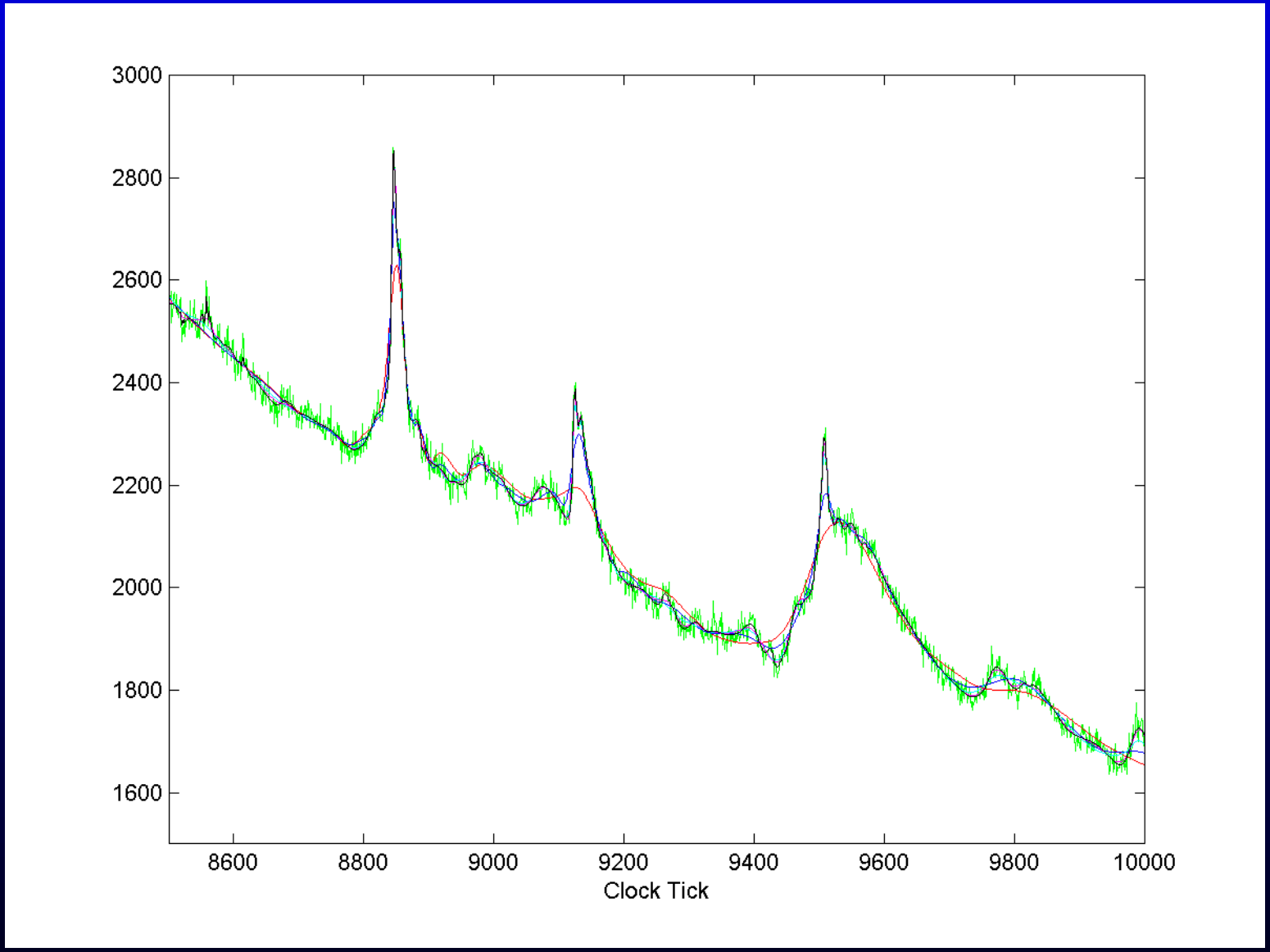
What Do the Noise Estimates Look Like?



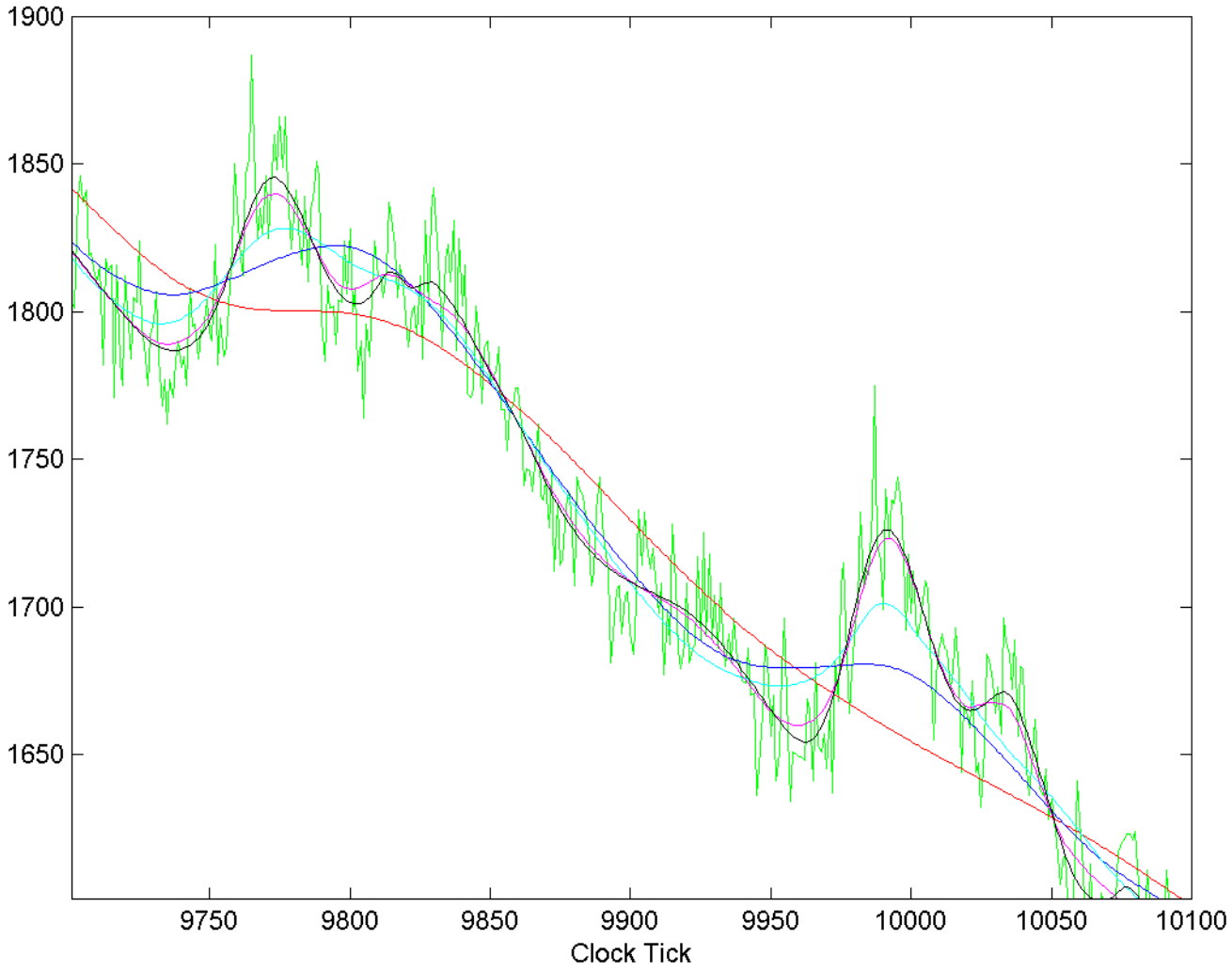
What Do the Smooth Estimates Look Like?



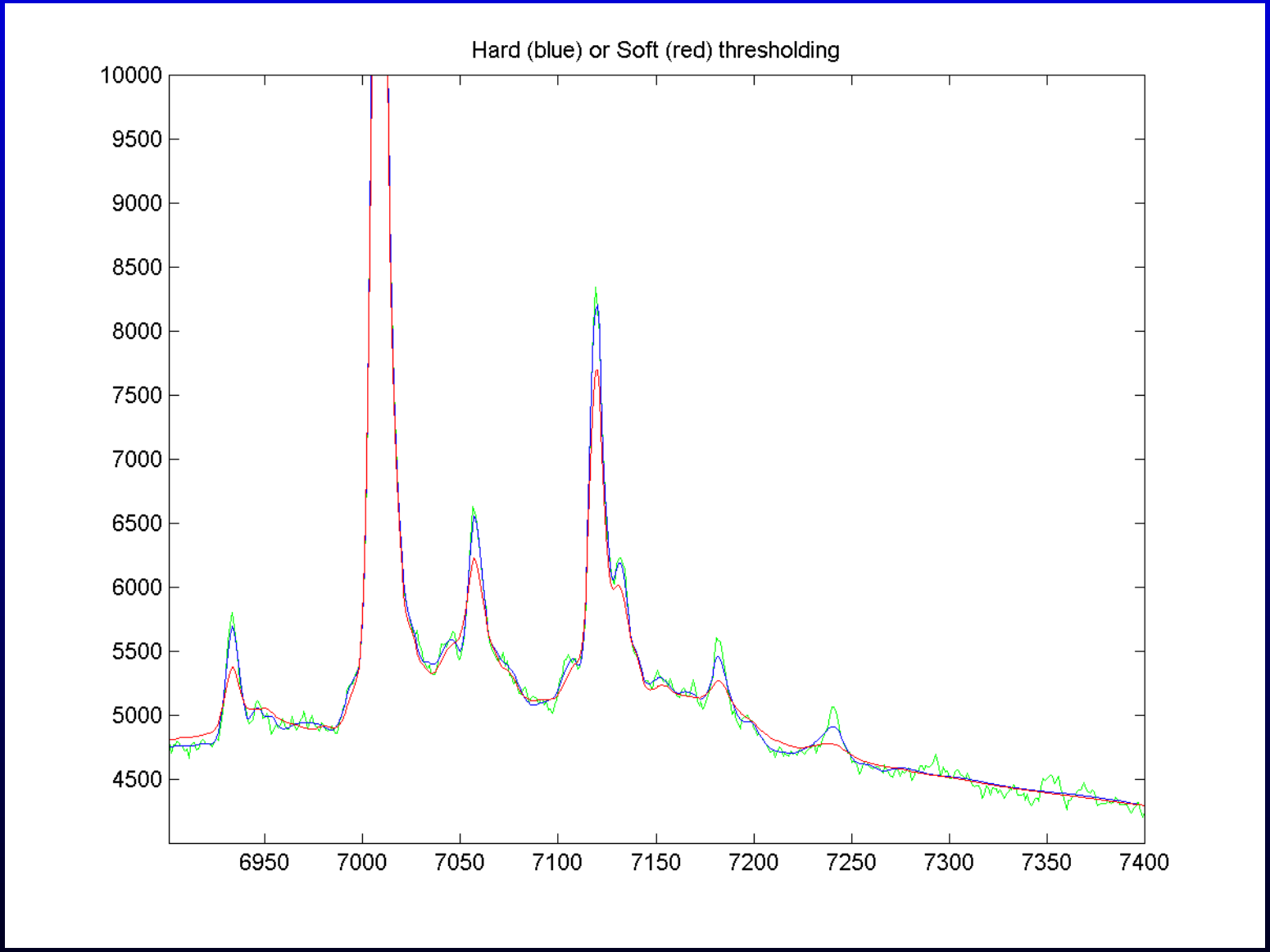
What Do the Smooth Estimates Look Like?



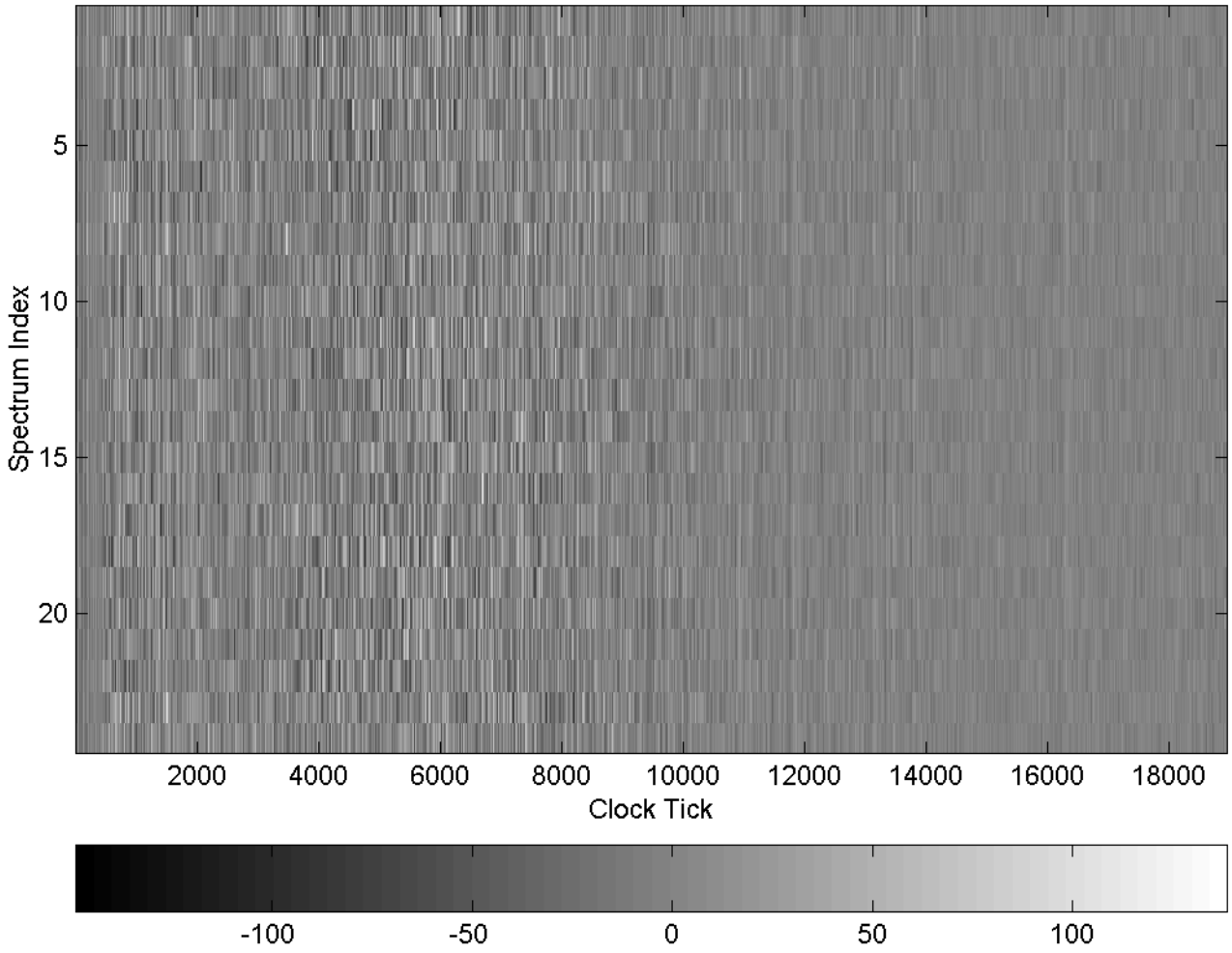
What Do the Smooth Estimates Look Like?



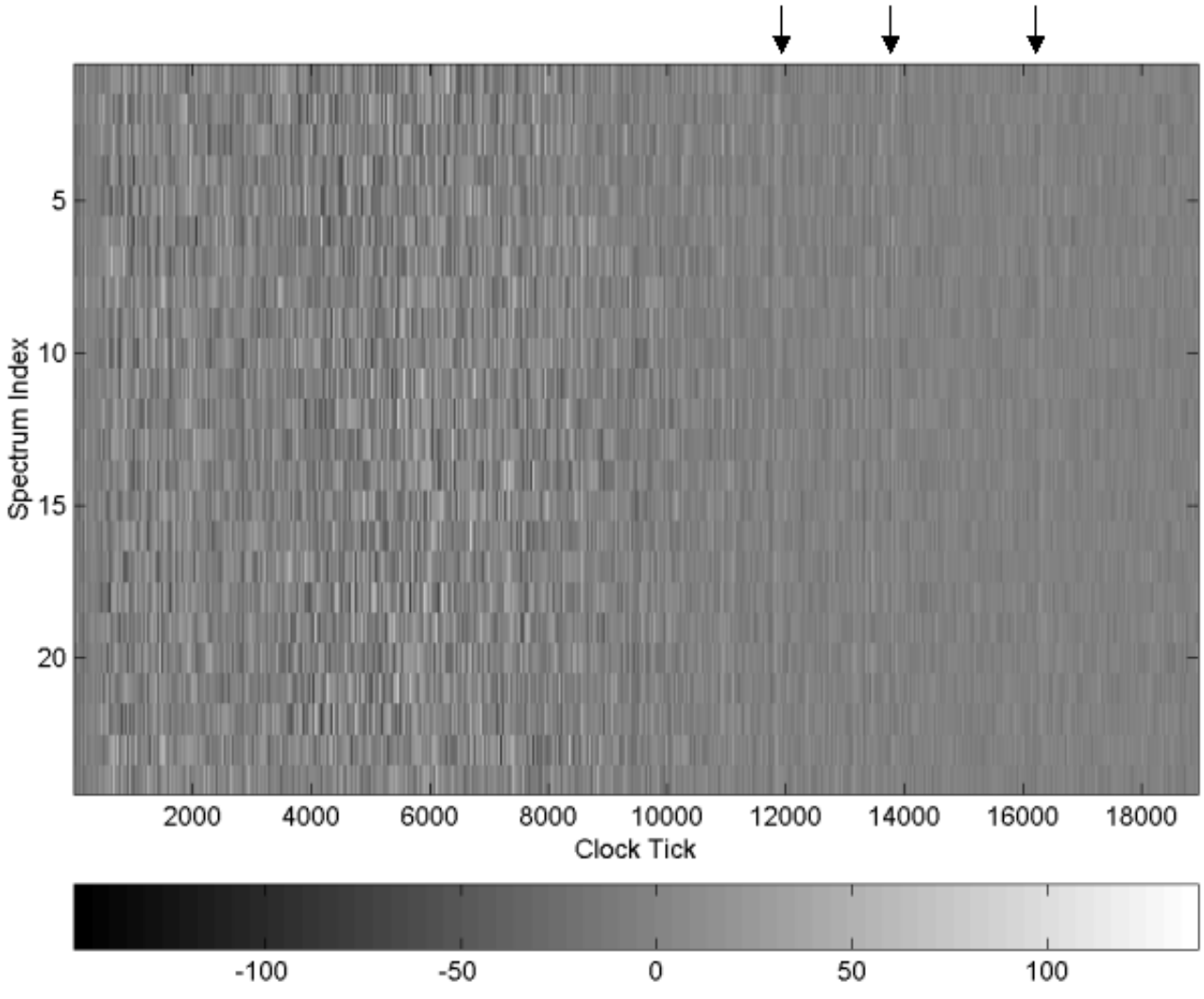
Should we Use Hard or Soft Thresholds?



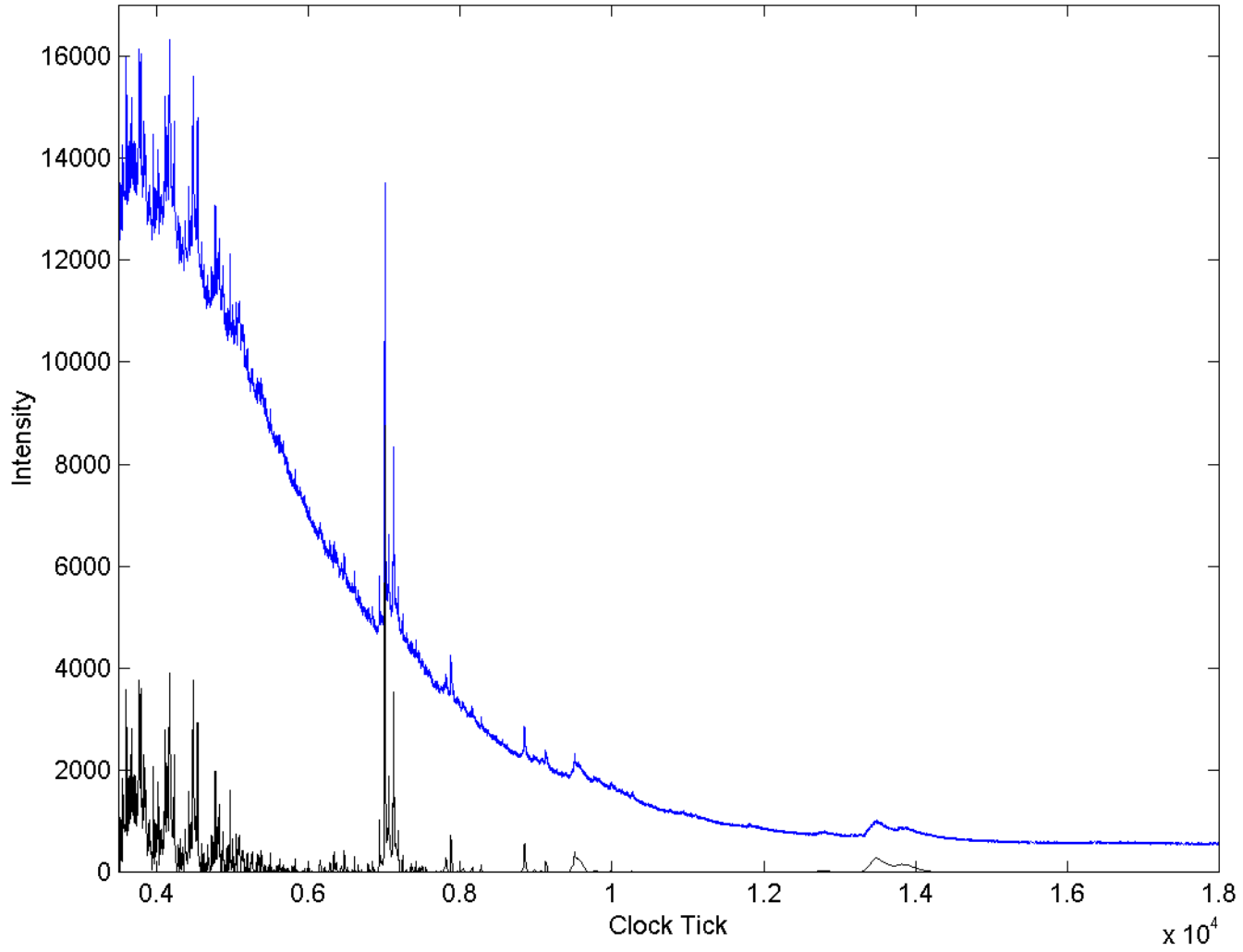
Is There Structure in the Residuals?



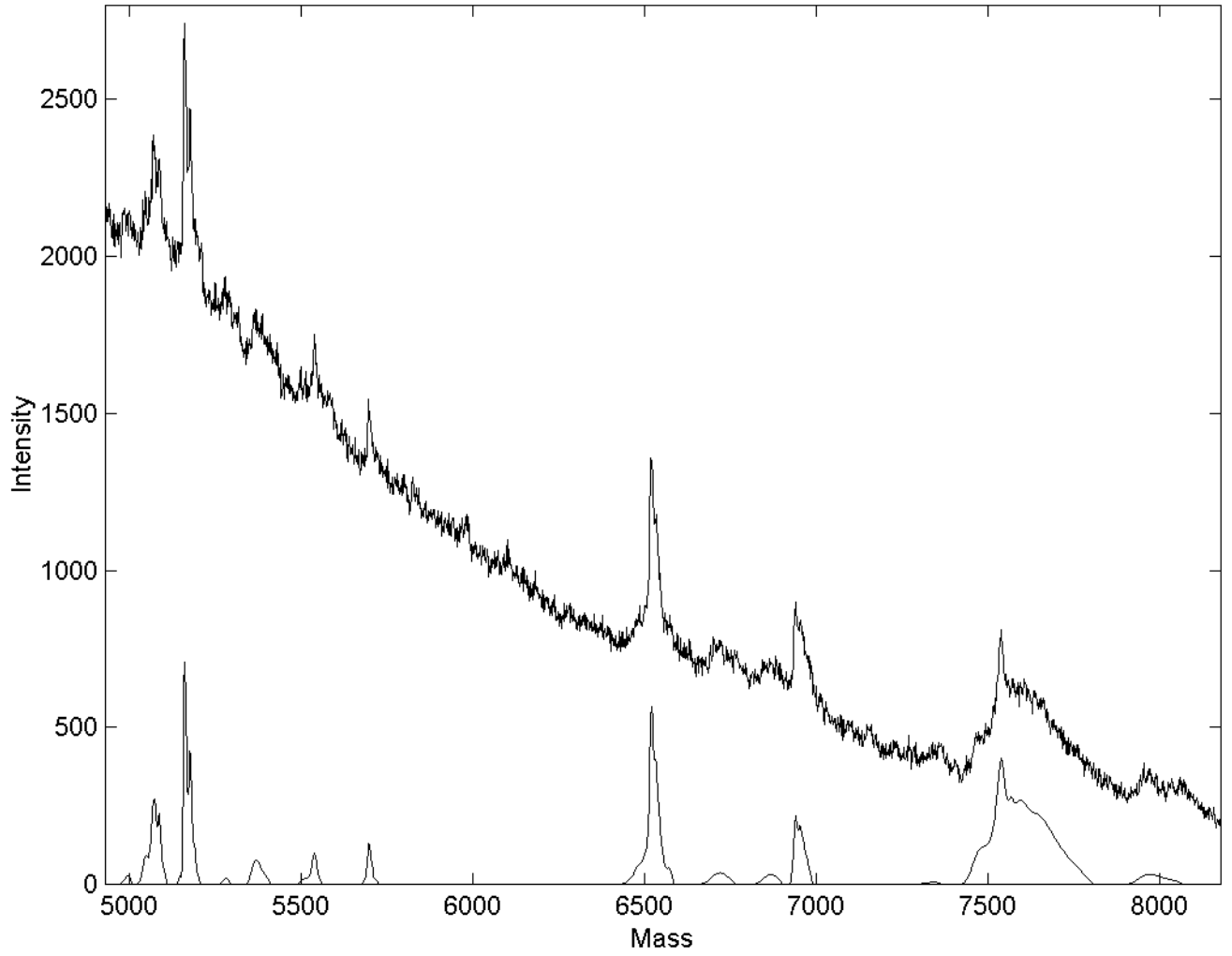
Is There Structure in the Residuals?



What Does a Corrected Spectrum Look Like?



What Does a Corrected Spectrum Look Like?

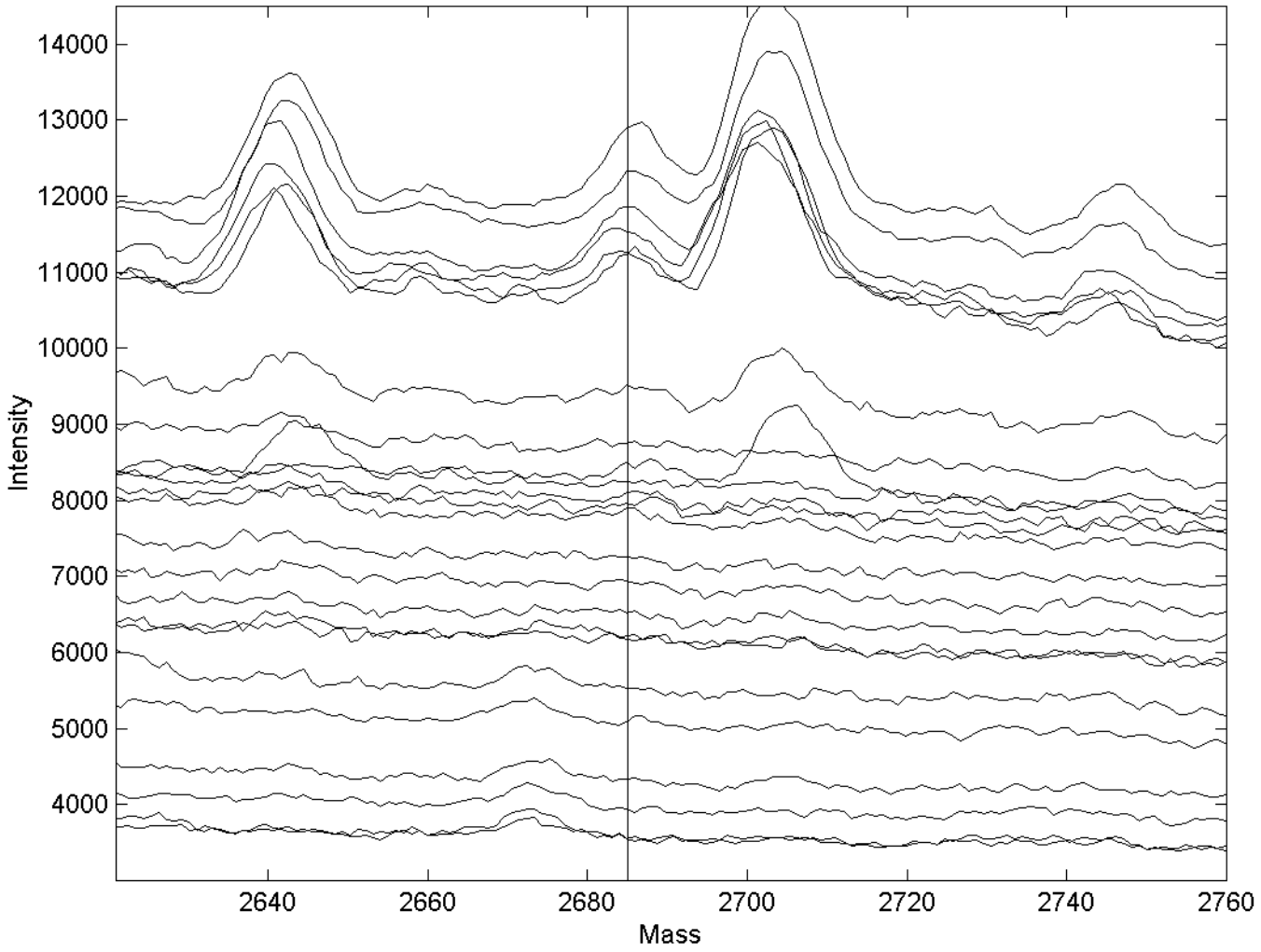


Make Two Passes to ID Peaks

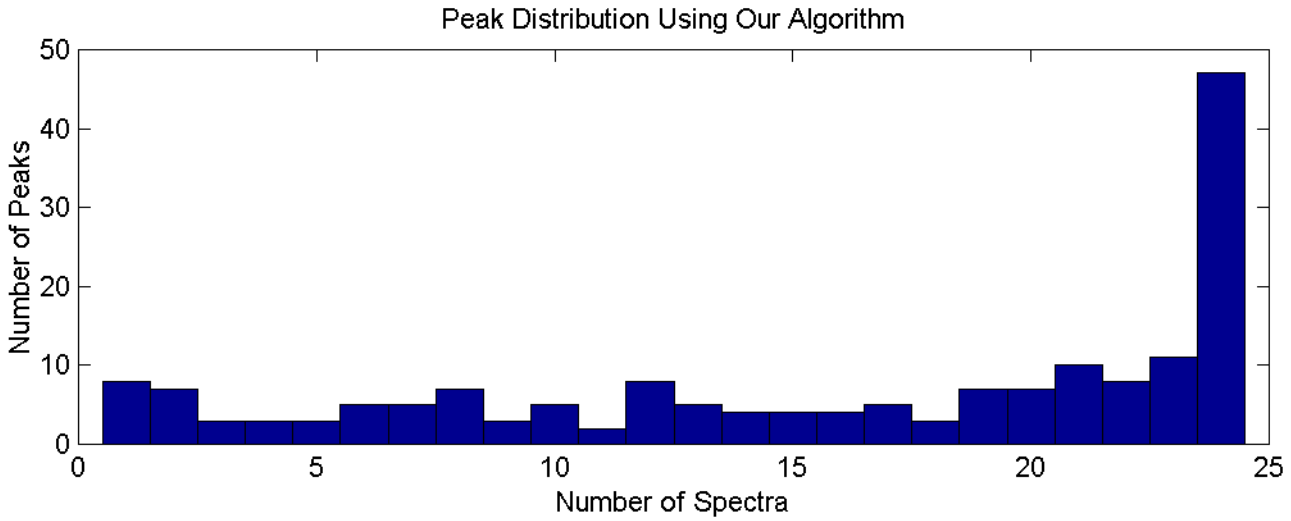
Pass 1: Keep a peak if it has S/N $>$ 10

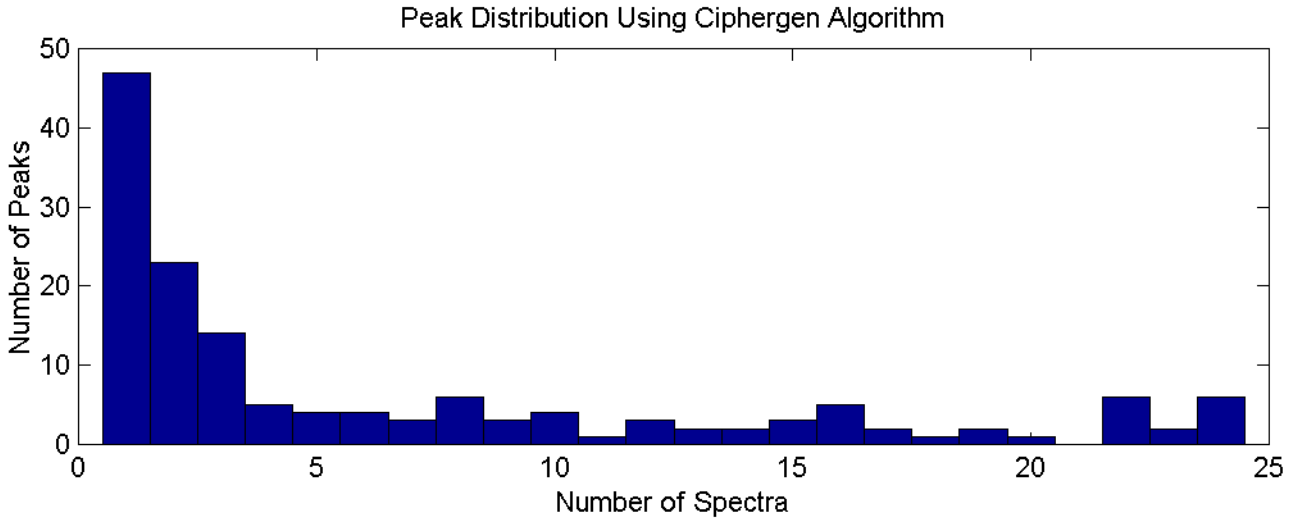
Pass 2: Keep peaks at already identified locations if they have S/N $>$ 2

Do the peaks look real?

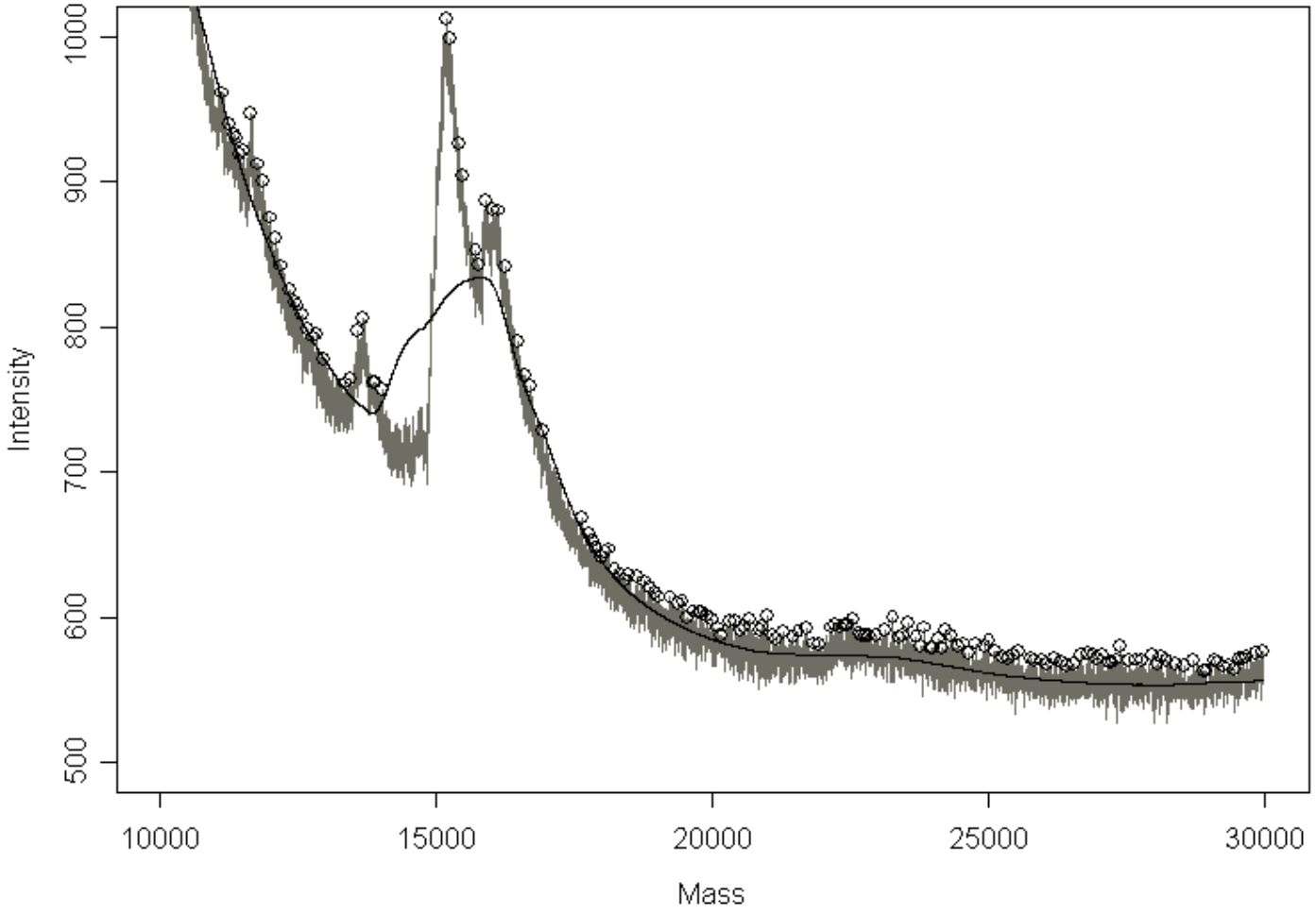


How do We Stack Up? (Ciphergen)

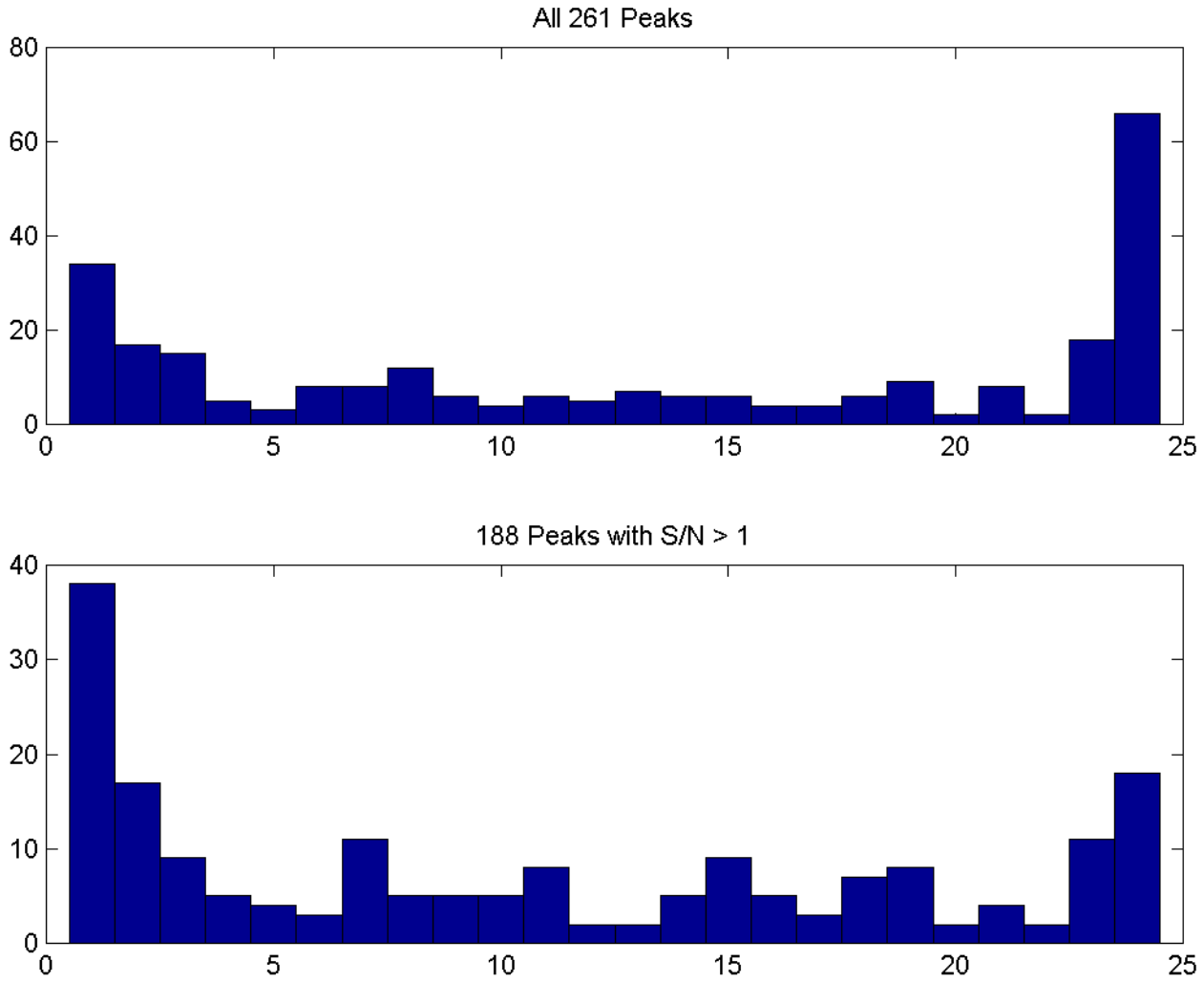




How do We Stack Up? (Yasui et al)



How do We Stack Up? (Yasui et al)



Did it Work?

We think so. We believe the peaks we find, and we find them consistently.

As a side note, this processing is extremely fast.

There are still open problems (eg normalization)

Some References

on the UDWT:

Lang, Guo, Odegard, Burrus and Wells (1996) IEEE Signal Processing Letters 3(1):10-12.

Coombes, Tsavachidis, Morris, Baggerly and Kuerer. preprint

Can We Find the Correct SELDI Calibration?

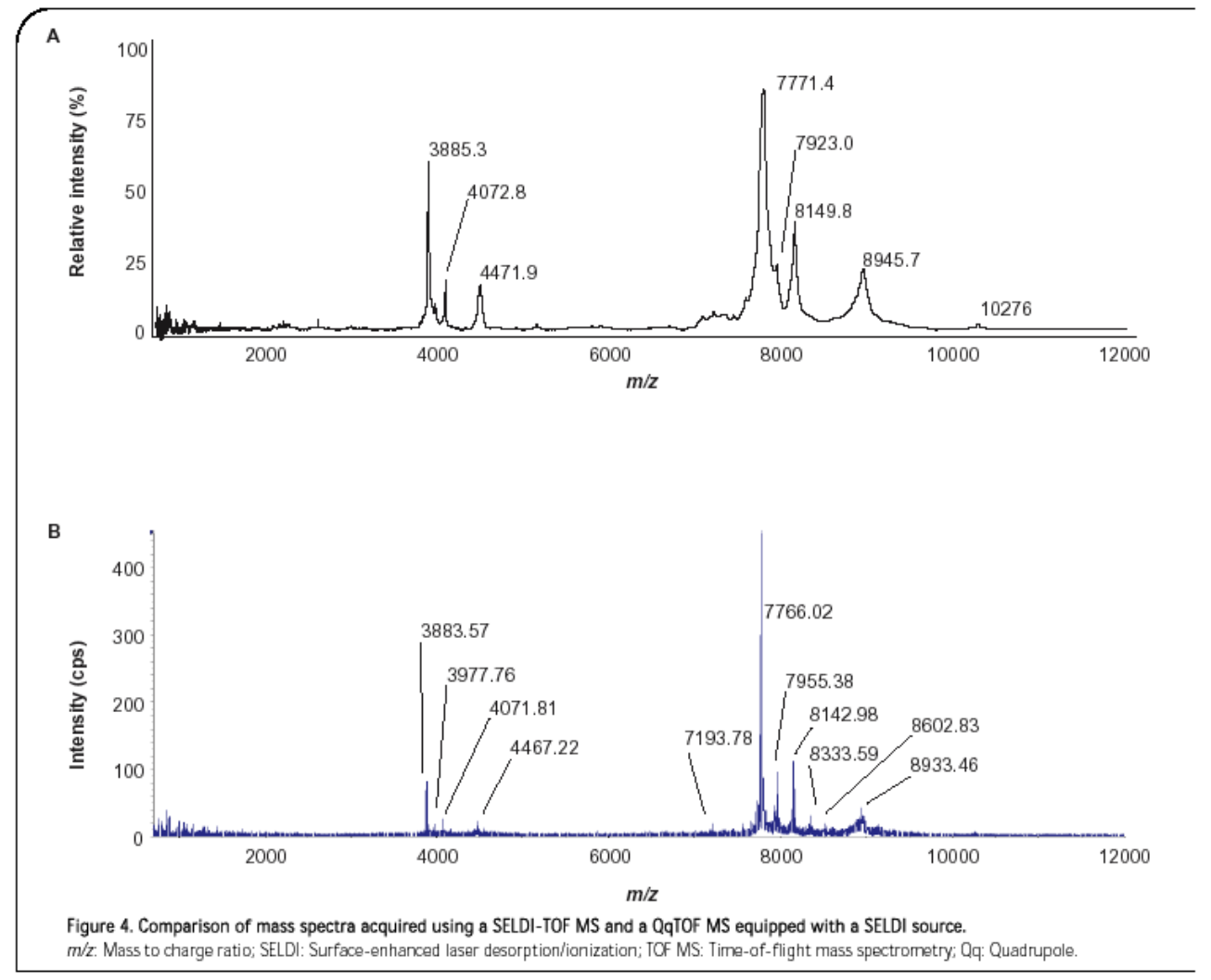
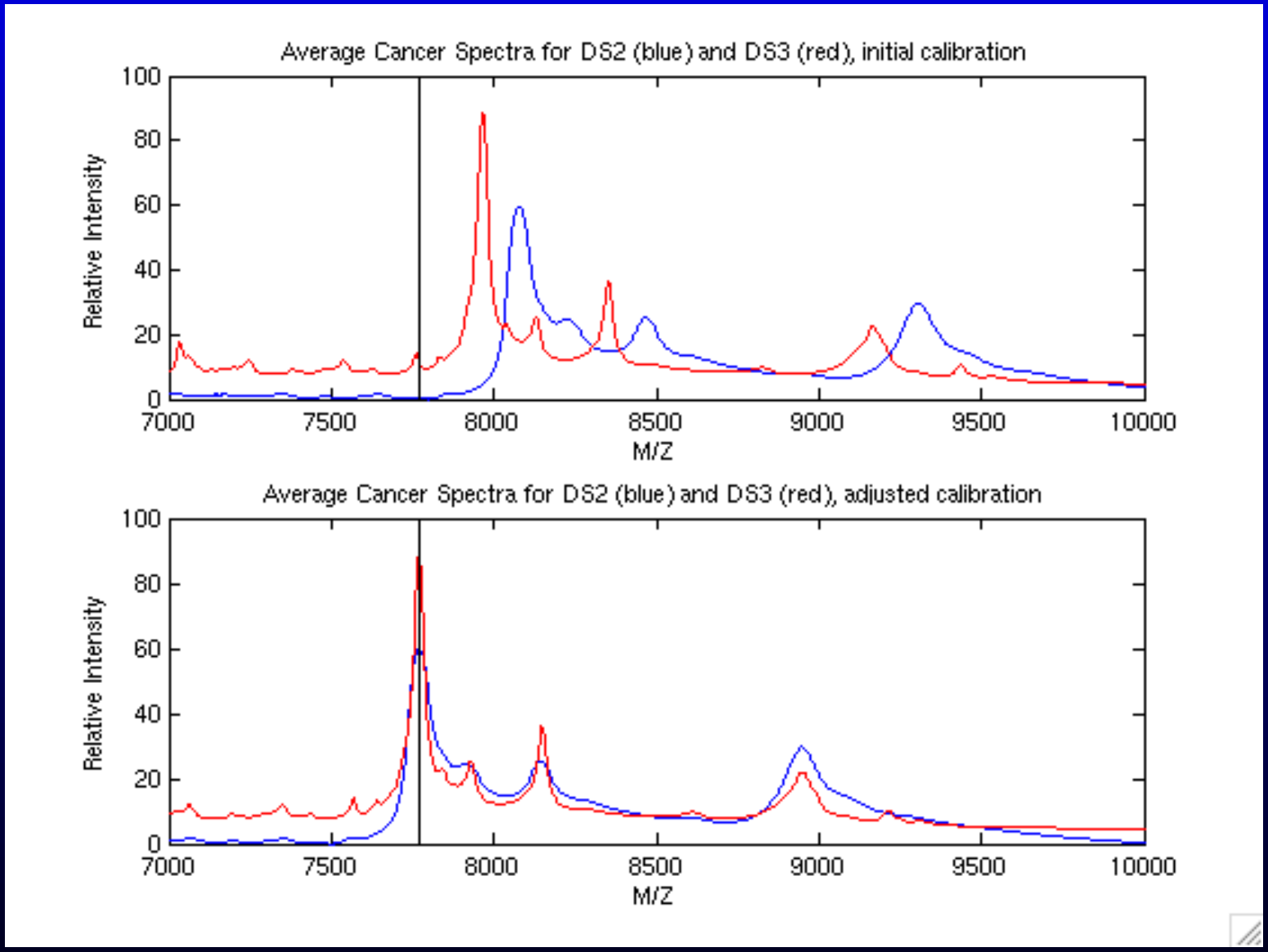


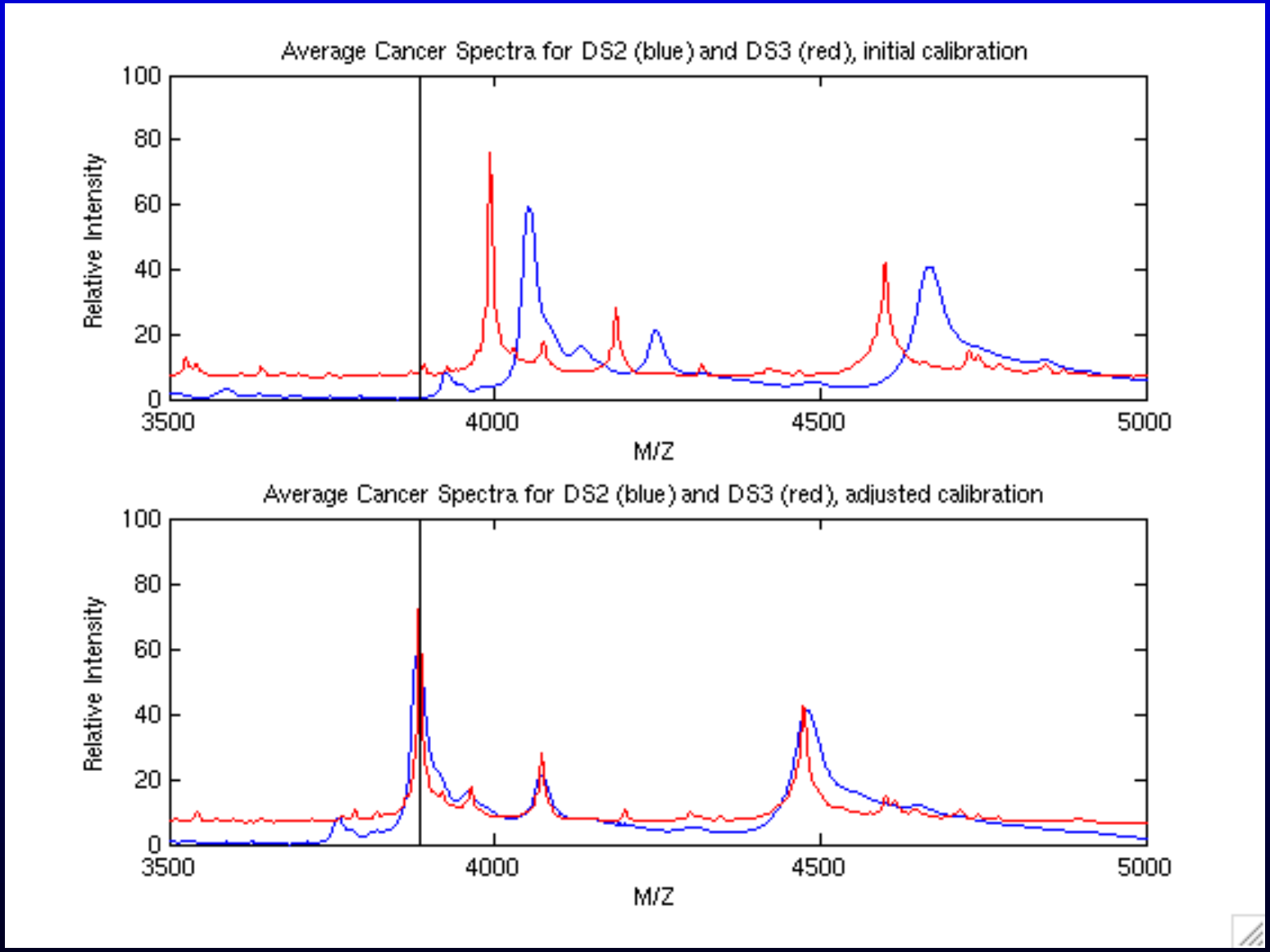
Figure 4. Comparison of mass spectra acquired using a SELDI-TOF MS and a QqTOF MS equipped with a SELDI source.
m/z: Mass to charge ratio; SELDI: Surface-enhanced laser desorption/ionization; TOF MS: Time-of-flight mass spectrometry; Qq: Quadrupole.

Conrads et al (2003), Exp Rev Mol Diag, 3:411-420.

Fixing the Biggest Peak



Fixing the Second Biggest Peak



How Far Off Were We?

Checking (DS3-real)/real, we're off

Between 2.5-2.7%.

Checking (DS2-real)/real, we're off

Between 3.9-4.3%.

“Producers and Consumers”

Recently, Petricoin et al. have replied to our analysis and that of Sorace and Zhan (BMC Bioinformatics, 2003) in

“Proteomic Pattern Diagnostics: Producers and Consumers in the Era of Correlative Science”.

“Producers and Consumers”

Recently, Petricoin et al. have replied to our analysis and that of Sorace and Zhan (BMC Bioinformatics, 2003) in

“Proteomic Pattern Diagnostics: Producers and Consumers in the Era of Correlative Science”.

Main Gist: There is a disconnect (lack of communication) between the “producers” of data (them) and the “consumers” of data (us).

“Producers and Consumers”

Recently, Petricoin et al. have replied to our analysis and that of Sorace and Zhan (BMC Bioinformatics, 2003) in

“Proteomic Pattern Diagnostics: Producers and Consumers in the Era of Correlative Science”.

Main Gist: There is a disconnect (lack of communication) between the “producers” of data (them) and the “consumers” of data (us).

Specific Contention: Because we didn’t talk with them, we have made some basic errors in our analyses.

"Producers and Consumers"

Recently, Petricoin et al. have replied to our analysis and that of Sorace and Zhan (BMC Bioinformatics, 2003) in

"Proteomic Pattern Diagnostics: Producers and Consumers in the Era of Correlative Science".

Main Gist: There is a disconnect (lack of communication) between the "producers" of data (them) and the "consumers" of data (us).

Specific Contention: Because we didn't talk with them, we have made some basic errors in our analyses.

Suggested Remedy: Consumers should work more closely with producers to ensure that the consumers are fully informed.

Specific Errors Mentioned

Start with (2.79,245.2) and structure in DS3:

- Their analyses didn't go down this far; values this low are not relevant. Shouldn't trust these numbers beccause they're lower than their lowest calibrant m/z value: 1084. M/Z values below this "should not be used for pattern classification".

Specific Errors Mentioned

Start with (2.79,245.2) and structure in DS3:

- Their analyses didn't go down this far; values this low are not relevant. Shouldn't trust these numbers beccause they're lower than their lowest calibrant m/z value: 1084. M/Z values below this "should not be used for pattern classification".
- 2.79 cannot be bias, because they were blinded and the samples were randomized with "cases and controls run intermingled on the same chips".

Specific Errors Mentioned

Start with (2.79,245.2) and structure in DS3:

- Their analyses didn't go down this far; values this low are not relevant. Shouldn't trust these numbers beccause they're lower than their lowest calibrant m/z value: 1084. M/Z values below this "should not be used for pattern classification".
- 2.79 cannot be bias, because they were blinded and the samples were randomized with "cases and controls run intermingled on the same chips".
- This must be real biology.

Specific Errors Mentioned

Start with (2.79,245.2) and structure in DS3:

- Their analyses didn't go down this far; values this low are not relevant. Shouldn't trust these numbers beccause they're lower than their lowest calibrant m/z value: 1084. M/Z values below this "should not be used for pattern classification".
- 2.79 cannot be bias, because they were blinded and the samples were randomized with "cases and controls run intermingled on the same chips".
- This must be real biology.
- Assertions that 2.79 shows bias are " judgementally biased and scientifically unfounded".

Specific Errors Mentioned

Start with (2.79,245.2) and structure in DS3:

- Their analyses didn't go down this far; values this low are not relevant. Shouldn't trust these numbers beccause they're lower than their lowest calibrant m/z value: 1084. M/Z values below this "should not be used for pattern classification".
- 2.79 cannot be bias, because they were blinded and the samples were randomized with "cases and controls run intermingled on the same chips".
- This must be real biology.
- Assertions that 2.79 shows bias are " judgementally biased and scientifically unfounded".

Specific Errors Mentioned, Continued

We confuse “noise” with “bias”.

Specific Errors Mentioned, Continued

We confuse “noise” with “bias”.

Extrapolation of DS3 problems to DS2 and DS1 unwarranted (mostly at Sorace?). Extrapolation to prostate data?

Specific Errors Mentioned, Continued

We confuse “noise” with “bias”.

Extrapolation of DS3 problems to DS2 and DS1 unwarranted (mostly at Sorace?). Extrapolation to prostate data?

We only look at the SELDI data, and not at the Qstar data which is the current state of the art. Their earlier stuff should be treated more as a proof of principle.

Specific Errors Mentioned, Continued

We confuse “noise” with “bias”.

Extrapolation of DS3 problems to DS2 and DS1 unwarranted (mostly at Sorace?). Extrapolation to prostate data?

We only look at the SELDI data, and not at the Qstar data which is the current state of the art. Their earlier stuff should be treated more as a proof of principle.

Our contention that you can’t find consistent structure in DS2 and DS3 is contradicted by a recent PNAS article by Zhu et al.

Specific Errors Mentioned, Continued

We confuse “noise” with “bias”.

Extrapolation of DS3 problems to DS2 and DS1 unwarranted (mostly at Sorace?). Extrapolation to prostate data?

We only look at the SELDI data, and not at the Qstar data which is the current state of the art. Their earlier stuff should be treated more as a proof of principle.

Our contention that you can’t find consistent structure in DS2 and DS3 is contradicted by a recent PNAS article by Zhu et al.

While we “wondered about their calibration method” in our article, they adhere to a strict SOP for data collection, and all spectra are calibrated at the time of collection.

Do We Agree?

No.

Do We Agree?

No.

We will now attempt to address the points that they raise, each in turn, but our overall assessment is that we think most of their criticisms are flawed.

We Shouldn’t Use Values Below the Lowest Calibrant

What have they used?

We Shouldn't Use Values Below the Lowest Calibrant

What have they used?

Their reported key m/z values:

DS1: 534.82, 989.15, 2111.71, 2251.18, 2465.02

DS2: 467.18, 500.85, 502.10, 664.92, 12354.27

DS3: 435.46, 465.57, 2760.67, 3497.55, 14051.98, 19643.41

We Shouldn't Use Values Below the Lowest Calibrant

What have they used?

Their reported key m/z values:

DS1: 534.82, 989.15, 2111.71, 2251.18, 2465.02

DS2: 467.18, 500.85, 502.10, 664.92, 12354.27

DS3: 435.46, 465.57, 2760.67, 3497.55, 14051.98, 19643.41

Of the peaks they report, 8/17 are below their lowest calibrant: 2/5 in DS1, 4/5 in DS2, and 2/7 in DS3. As we've seen, the two lowest values in DS3 are the most important for separation.

We Shouldn't Use Values Below the Lowest Calibrant

Stuff in more recent papers:

Alexe et al (Proteomics, 2004, Petricoin and Liotta are coauthors) exploit the low m/z region. They don't go below 200 because they exclude it by fiat.

Zhu et al (PNAS, Dec 9, 2003), which they cite to refute us, find most of their structure in the low m/z range (including one below 245).

(both of these papers use the posted m/z values, which are wrong due to miscalibration).

We Shouldn't Use Values Below the Lowest Calibrant

Stuff in more recent papers:

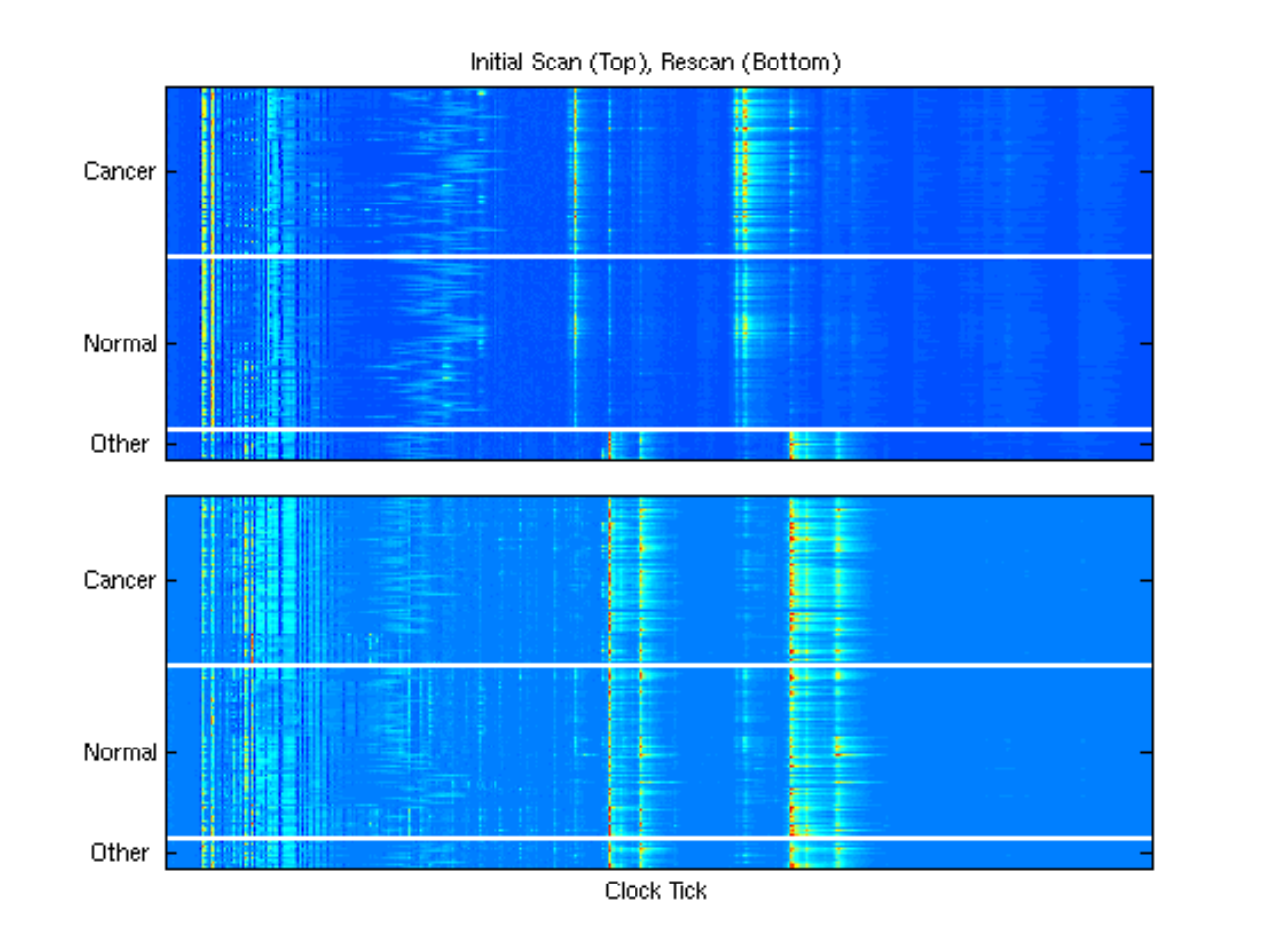
Alexe et al (Proteomics, 2004, Petricoin and Liotta are coauthors) exploit the low m/z region. They don't go below 200 because they exclude it by fiat.

Zhu et al (PNAS, Dec 9, 2003), which they cite to refute us, find most of their structure in the low m/z range (including one below 245).

(both of these papers use the posted m/z values, which are wrong due to miscalibration).

They can't have it both ways. If they can report it, others can look at it.

Their Randomization Should Preclude Bias



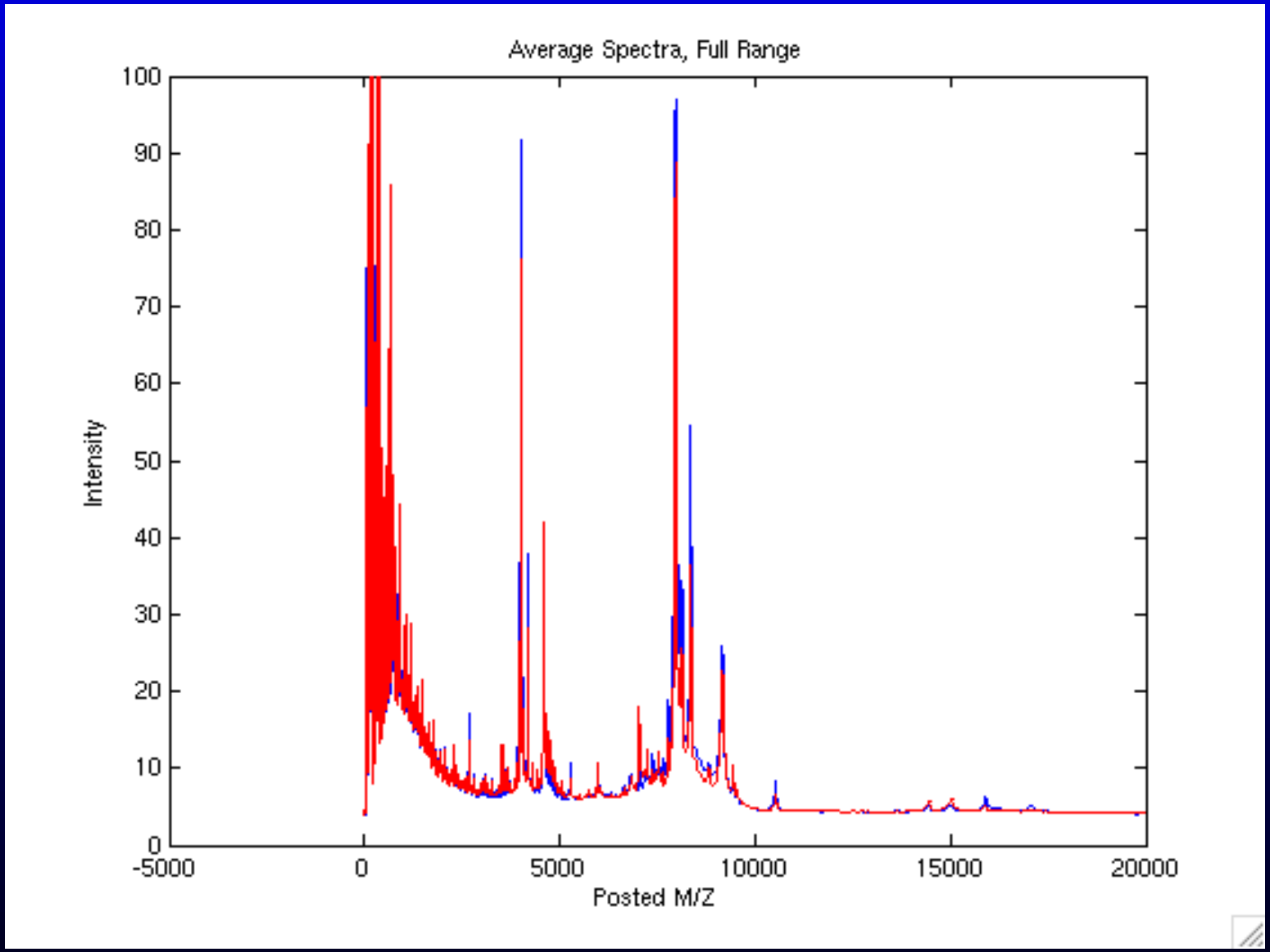
Same statement made in Lancet paper. The benign disease samples were clearly not randomized.

Their Randomization Should Preclude Bias

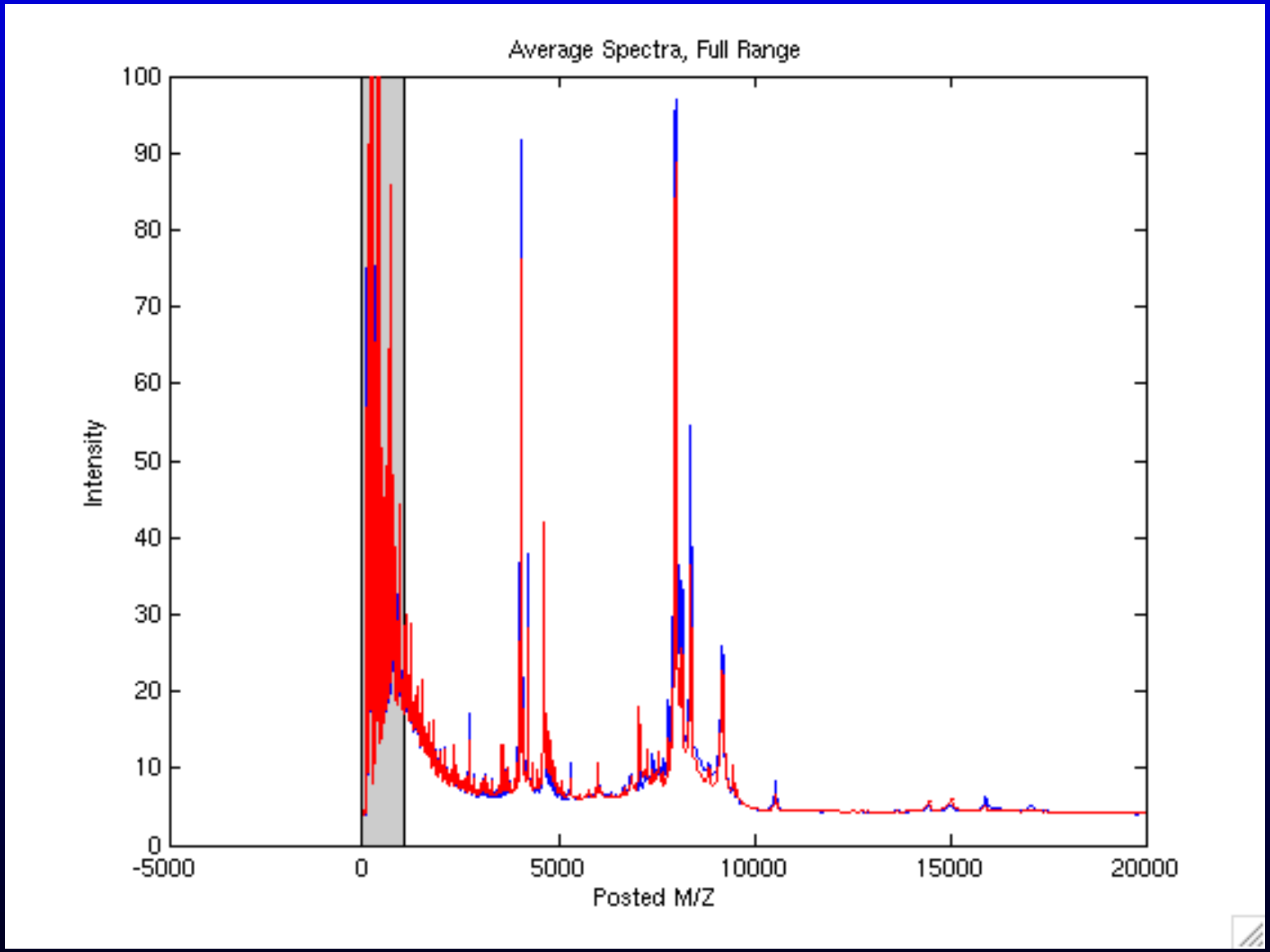
There is other evidence causing us to suspect the randomization of DS3. The files were posted twice, and the earlier posting suggests that they were not randomized.

We have more details on this.

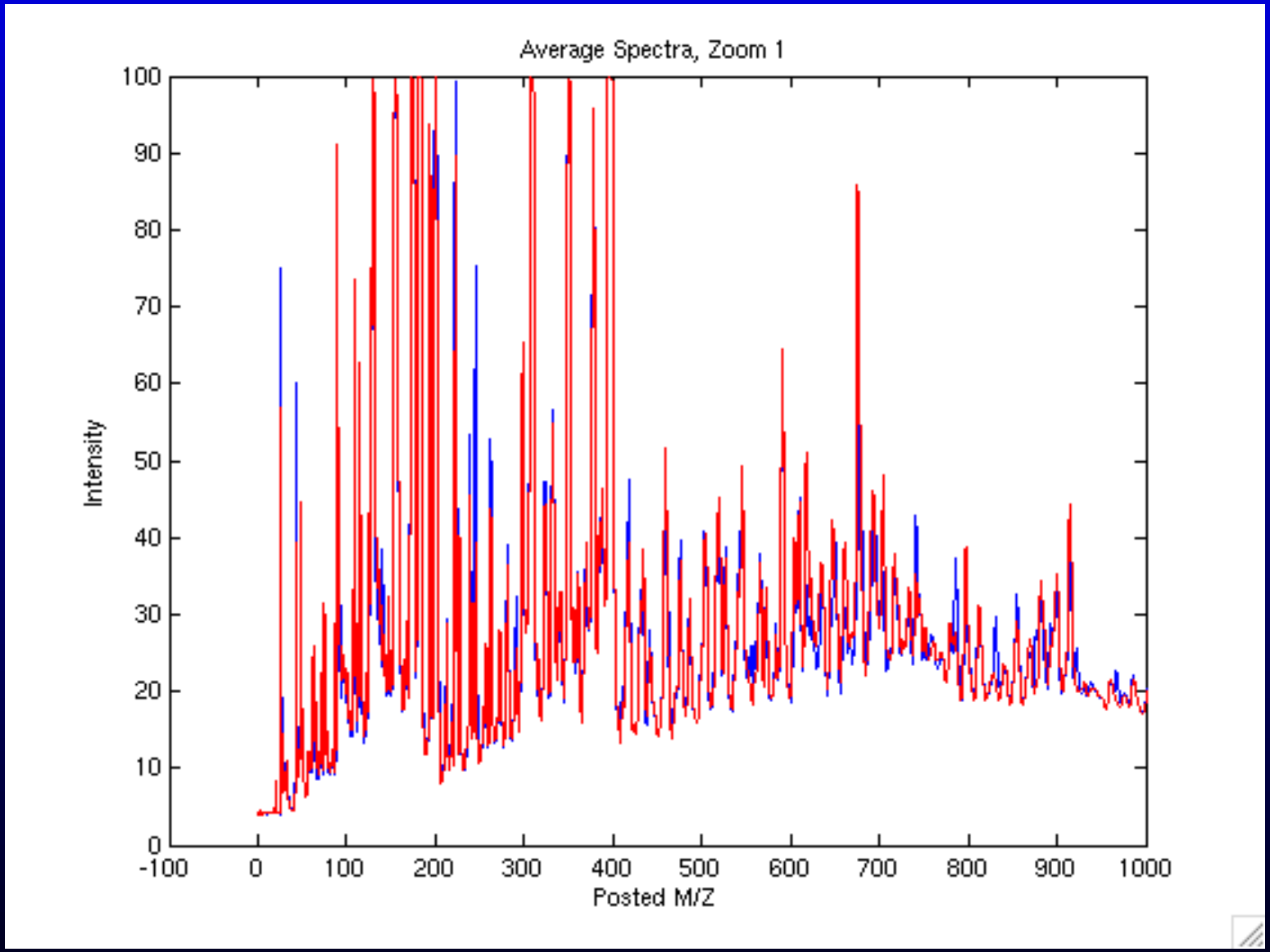
2.79 Must Be Biology?



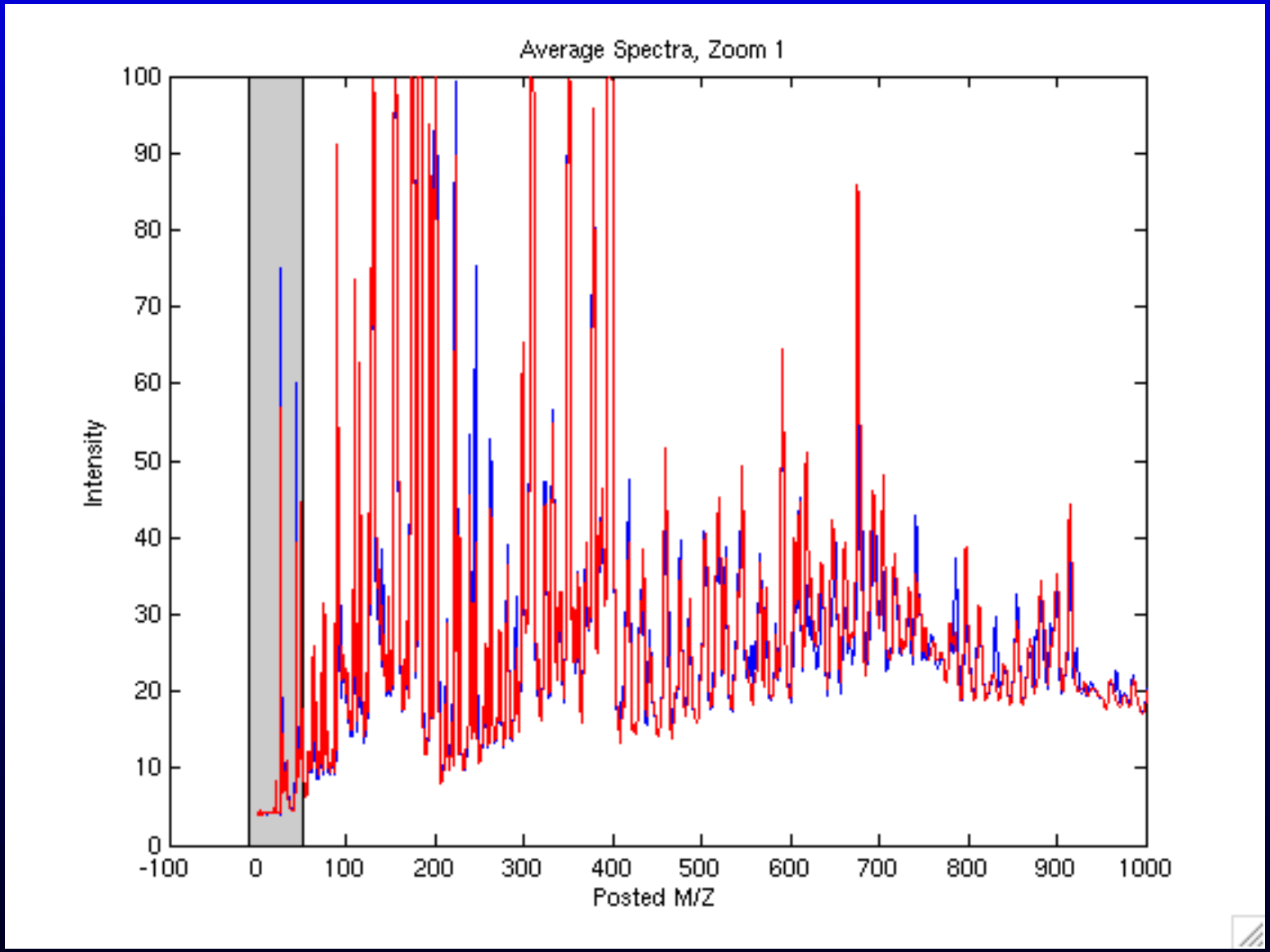
2.79 Must Be Biology? Zooming In...



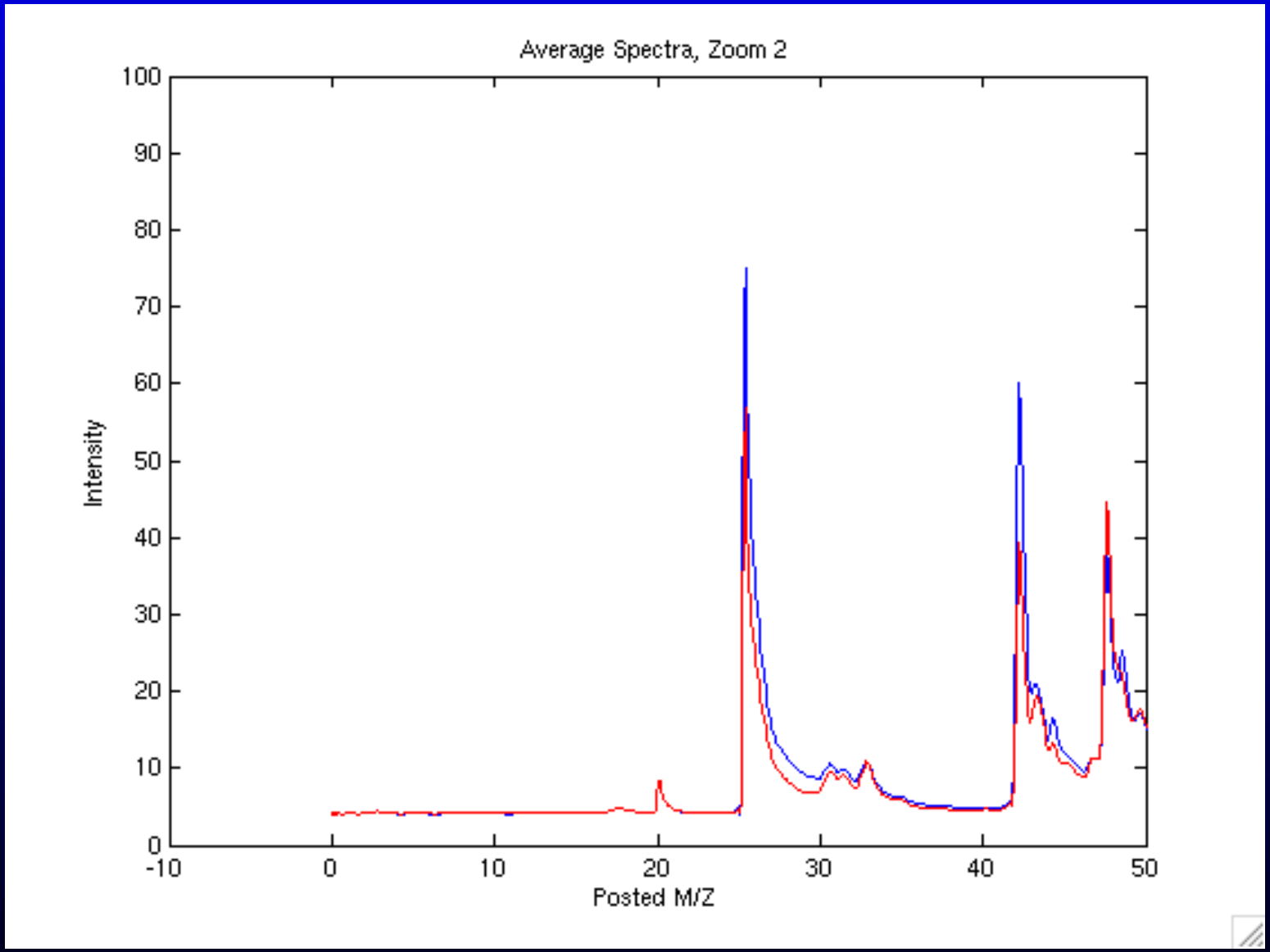
2.79 Must Be Biology? Zooming In...



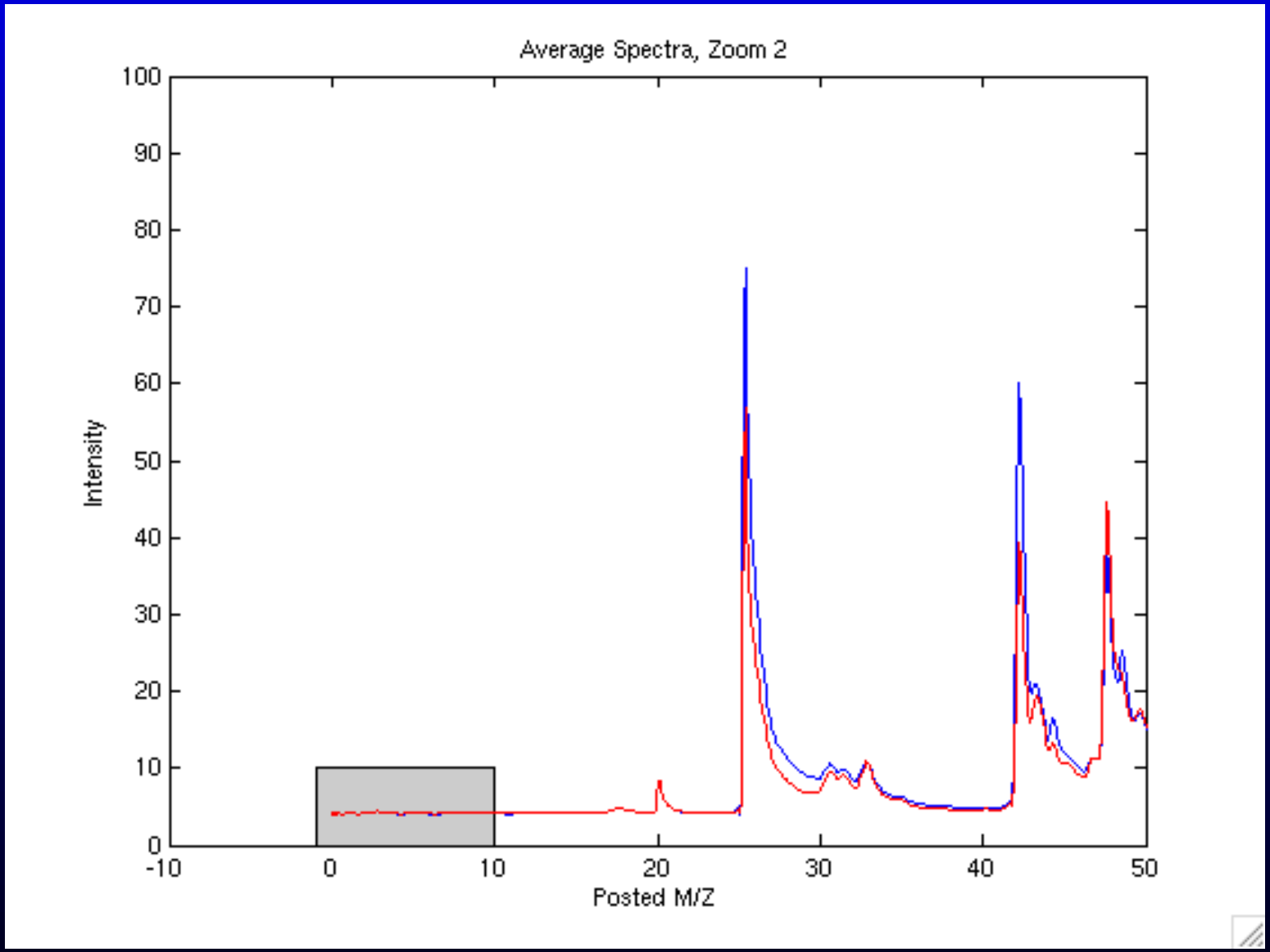
2.79 Must Be Biology? Zooming In...



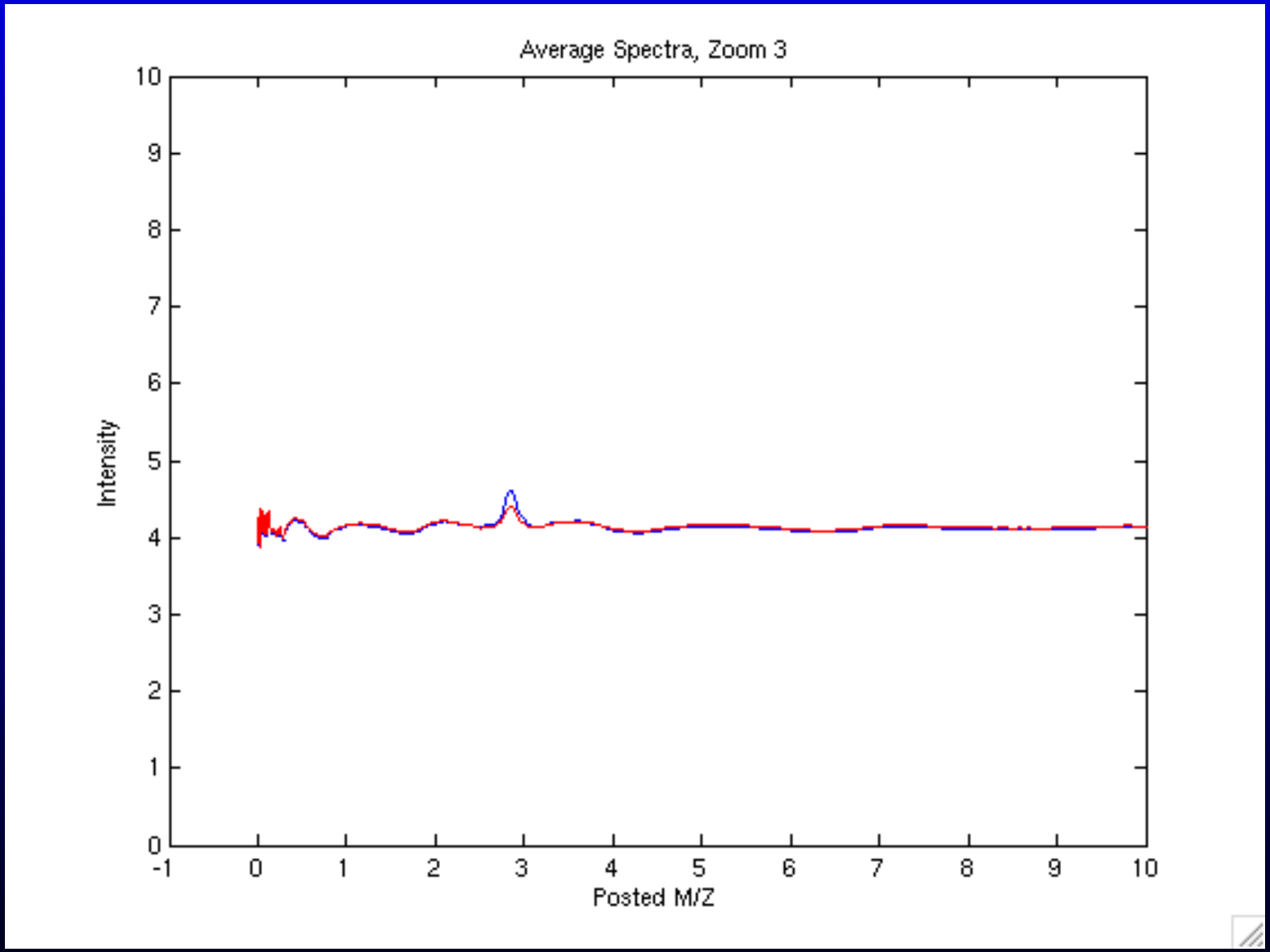
2.79 Must Be Biology? Zooming In...



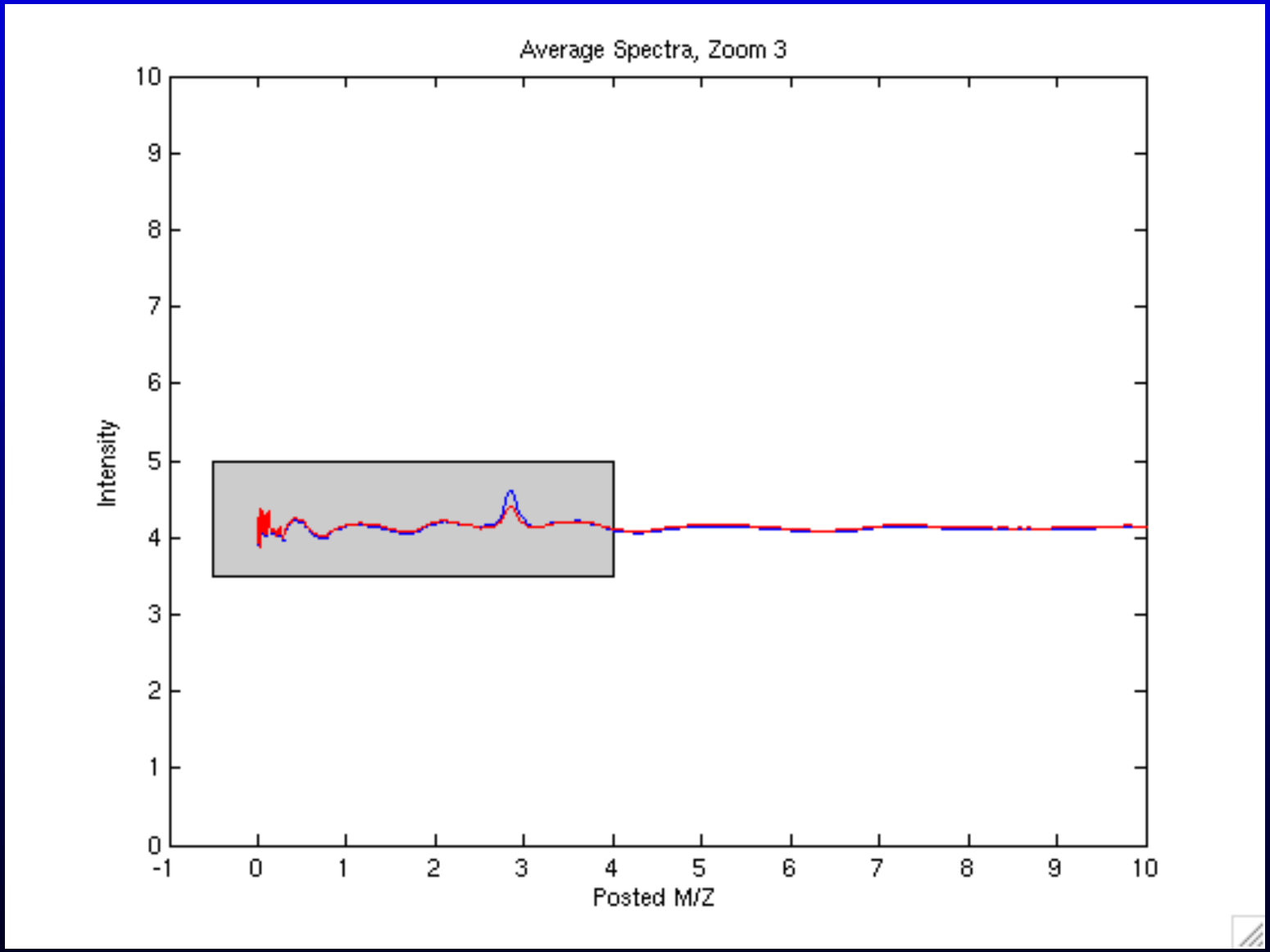
2.79 Must Be Biology? Zooming In...



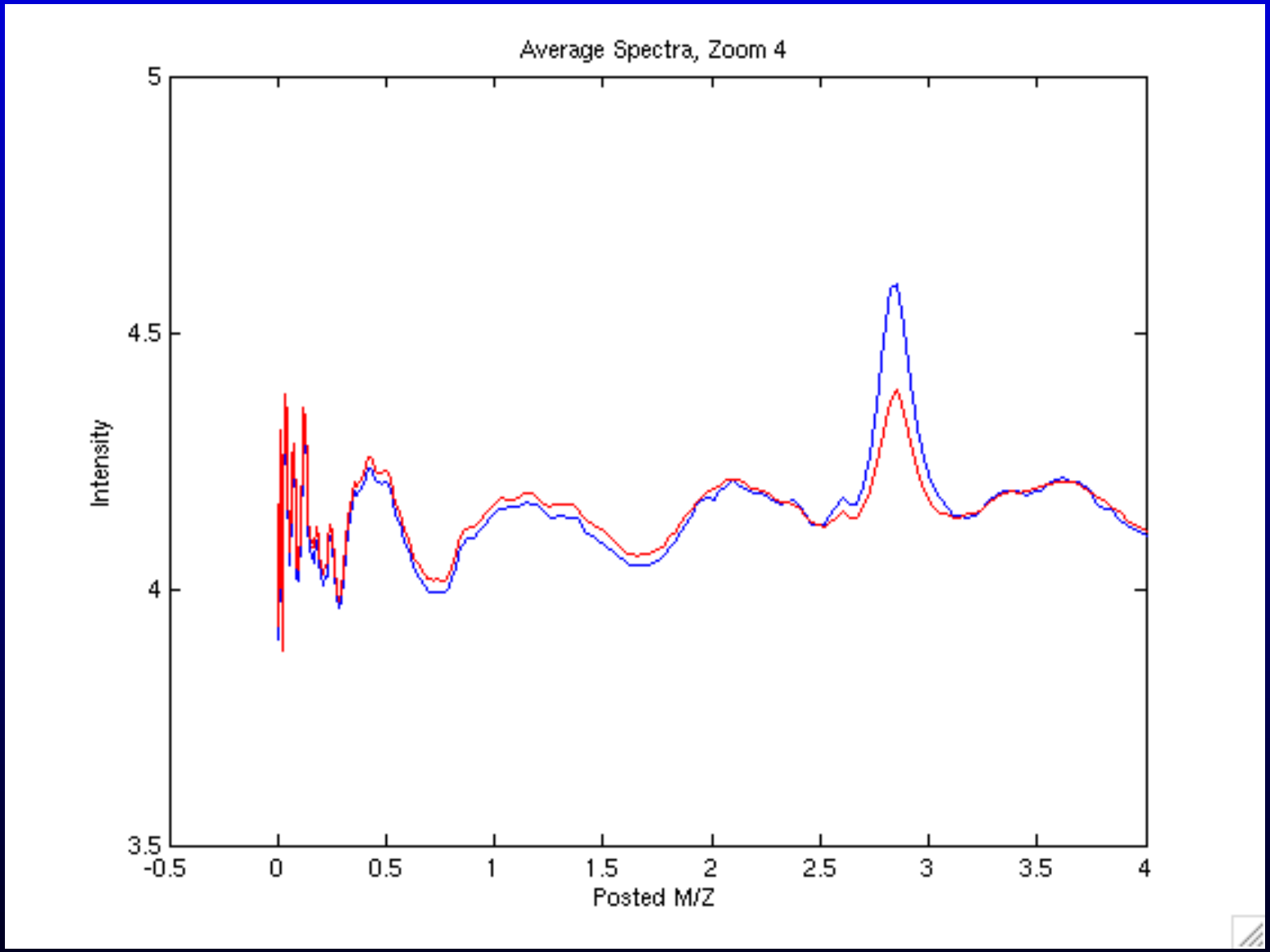
2.79 Must Be Biology? Zooming In...



2.79 Must Be Biology? Zooming In...



2.79 Must Be Biology? Zooming In...



2.79 Must Be Biology?

There’s nothing else close by.

The signal is very weak.

We believe that there is a pattern in the signal that may be associated with run order.

Our Assertions about 2.79

We do not feel that the points that we raise are “judgementally biased” or “scientifically unfounded”.

We would welcome comments as to how to address this conflict in more detail.

We Confuse Noise With Bias

This is a semantic issue. We meant that we were able to find structure that separated cancers from normals in regions of the spectra that we believe are in regions unaffected by biology. These regions of the spectra are affected by random electrical signals associated with the machine operation, and we call this electronic noise.

In this region, we should not be able to find structure; the signals should be statistically distributed as noise.

Given that we do find structure where there should only be noise, wee postulate the existence of bias.

Extrapolation to Other Data Sets is Unwarranted

We suspect that this point was meant to be directed only at Sorace and Zhan, who looked just at DS3.

Most of the cautions they raised have to do with experimental design, and we agree with those. Further, we have looked at all of the posted SELDI data sets, so we are not “extrapolating”.

Along those lines, we will note that the posted SELDI prostate spectra are also baseline subtracted and use the default calibration (which is wrong).

We Only Looked at SELDI Data, Not Qstar

Earlier stuff proof of principle, but not cutting edge.

We Only Looked at SELDI Data, Not Qstar

Earlier stuff proof of principle, but not cutting edge.

What principle?

We Only Looked at SELDI Data, Not Qstar

Earlier stuff proof of principle, but not cutting edge.

What principle?

Most of the issues we raise with respect to the SELDI data have to do with experimental design. These questions are still relevant.

We Only Looked at SELDI Data, Not Qstar

Earlier stuff proof of principle, but not cutting edge.

What principle?

Most of the issues we raise with respect to the SELDI data have to do with experimental design. These questions are still relevant.

The DS3 chips were used in the posted Qstar experiment, so problems there could carry over.

We Only Looked at SELDI Data, Not Qstar

Earlier stuff proof of principle, but not cutting edge.

What principle?

Most of the issues we raise with respect to the SELDI data have to do with experimental design. These questions are still relevant.

The DS3 chips were used in the posted Qstar experiment, so problems there could carry over.

We didn’t comment on the Qstar data due to time of writing. We’ve now looked at the Qstar data too, and similar issues of experimental design come up.

Others Find Structure; It Must Be There!

Zhu et al (PNAS, Dec 9 2003) identify 18 m/z values with large t-statistics that allow them to differentiate normals and cancers in DS2 and DS3.

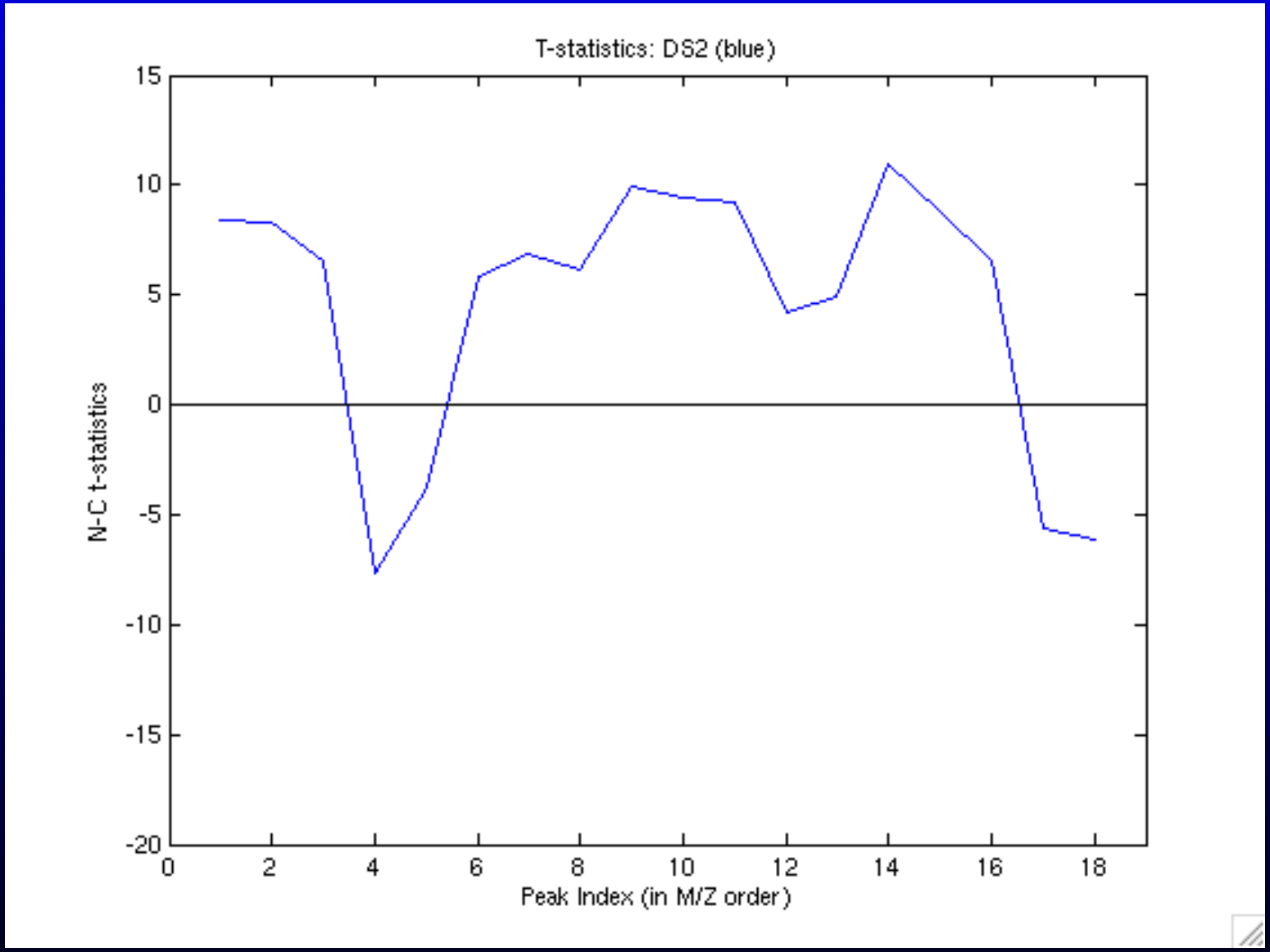
Their 18 m/z values: 167.80, 321.42, 322.42, 359.63, 385.57, 413.17, 433.91, 434.69, 444.47, 445.26, 1222.18, 1528.34, 3345.80, 3449.15, 3473.31, 3528.53, 6101.63, 6123.52.

9 values are below 500.

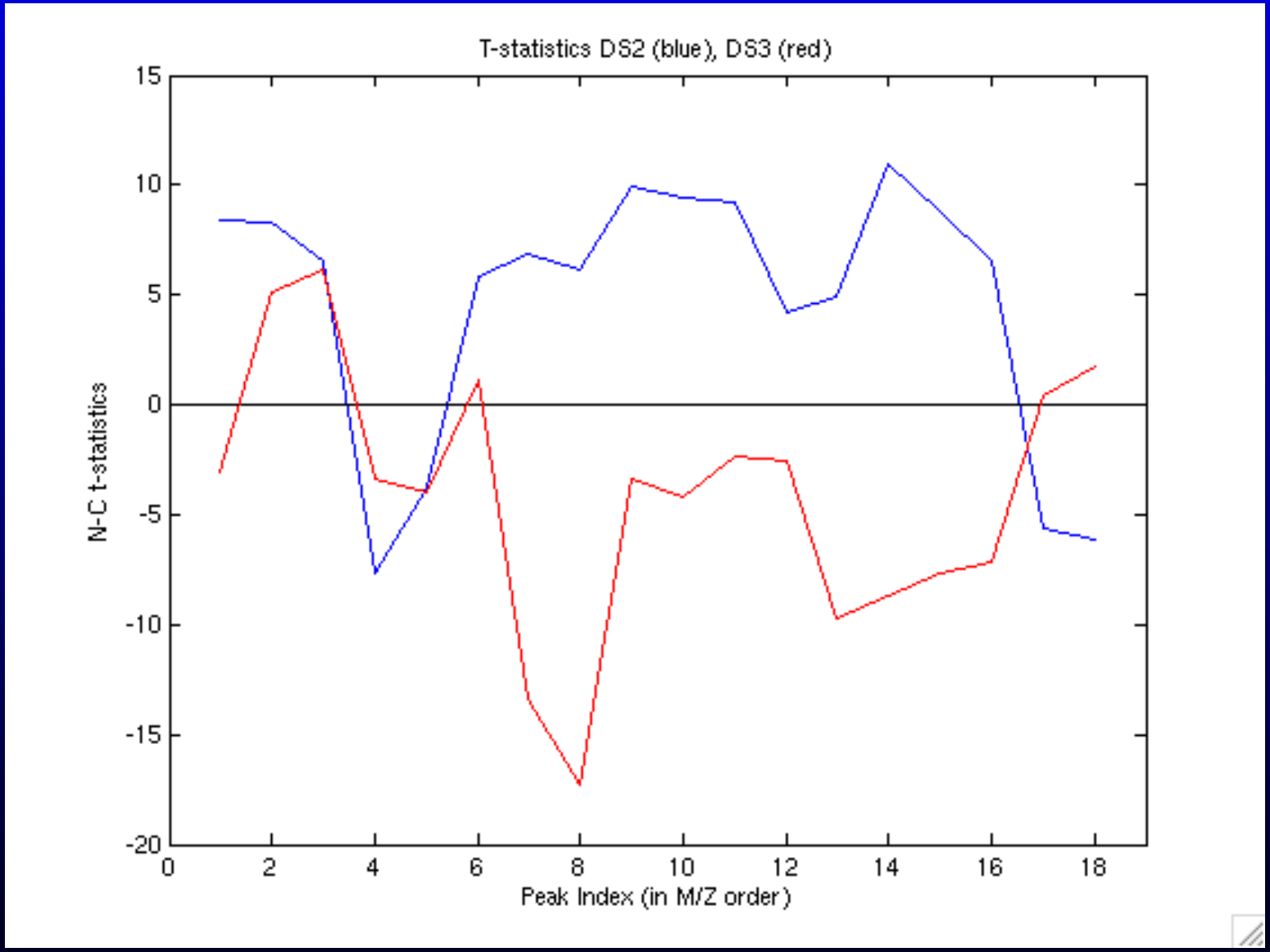
No comment about baseline correction or about calibration shifts.

Let’s look at the t-statistics...

What Are the T-Statistics?



13/18 Flip Sign?!?



Their SOP Means Their Calibration is OK

Actually, we believe that they do have a SOP for calibration. We still contend, however, that the data posted to the web have not been calibrated.

The paper by Conrads et al (Exp. Rev. Mol. Diag., 2003, Petricoin and Liotta coauthors) shows a SELDI spectrum derived from DS3 with a maximum located at m/z 7771. The corresponding maxima locations in the posted data are at 7966.6 for DS3 and 8076.9 for DS2. The mass errors for these two data sets are about 2.5% and 3.9% respectively.

The Conrads et al Figure

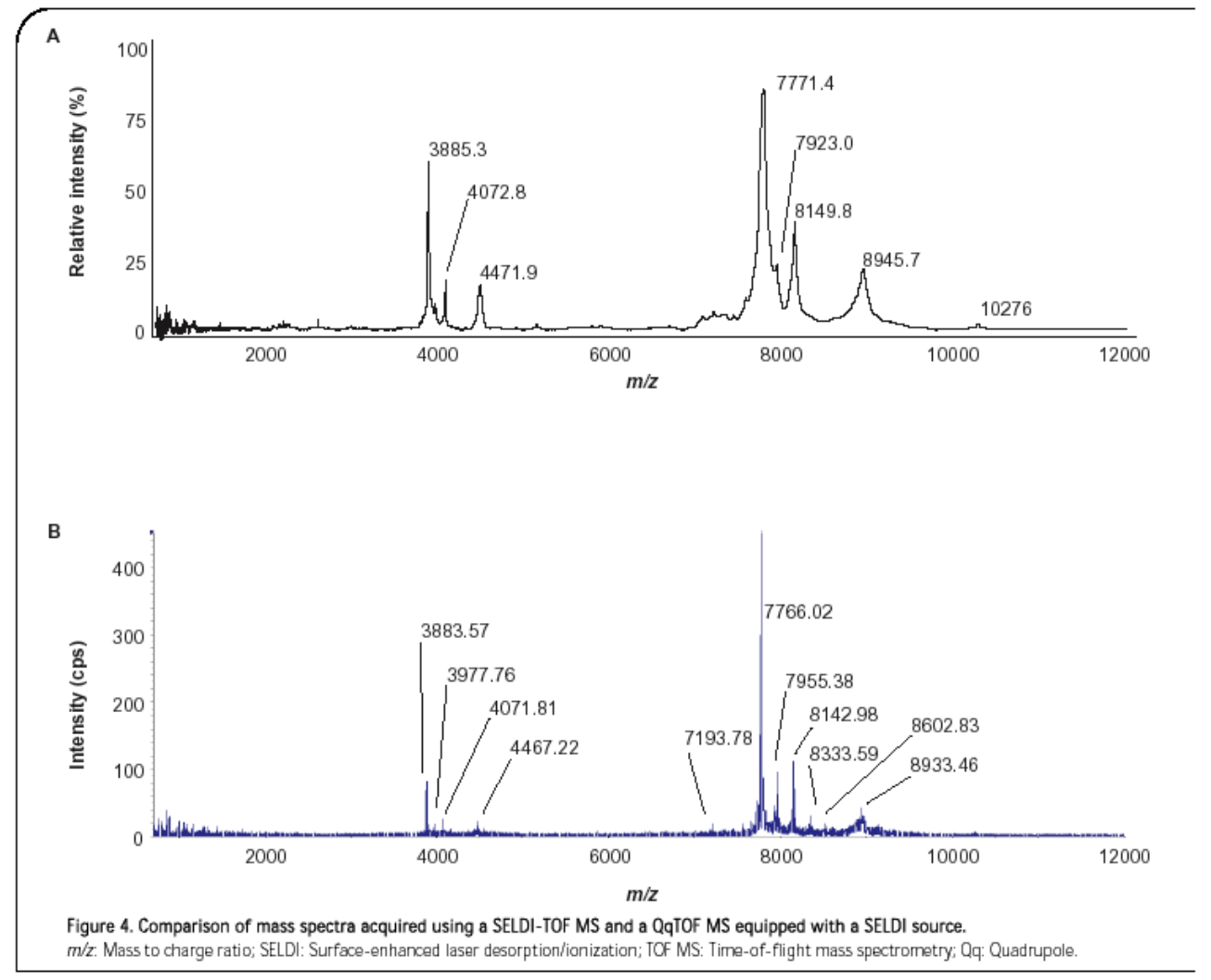
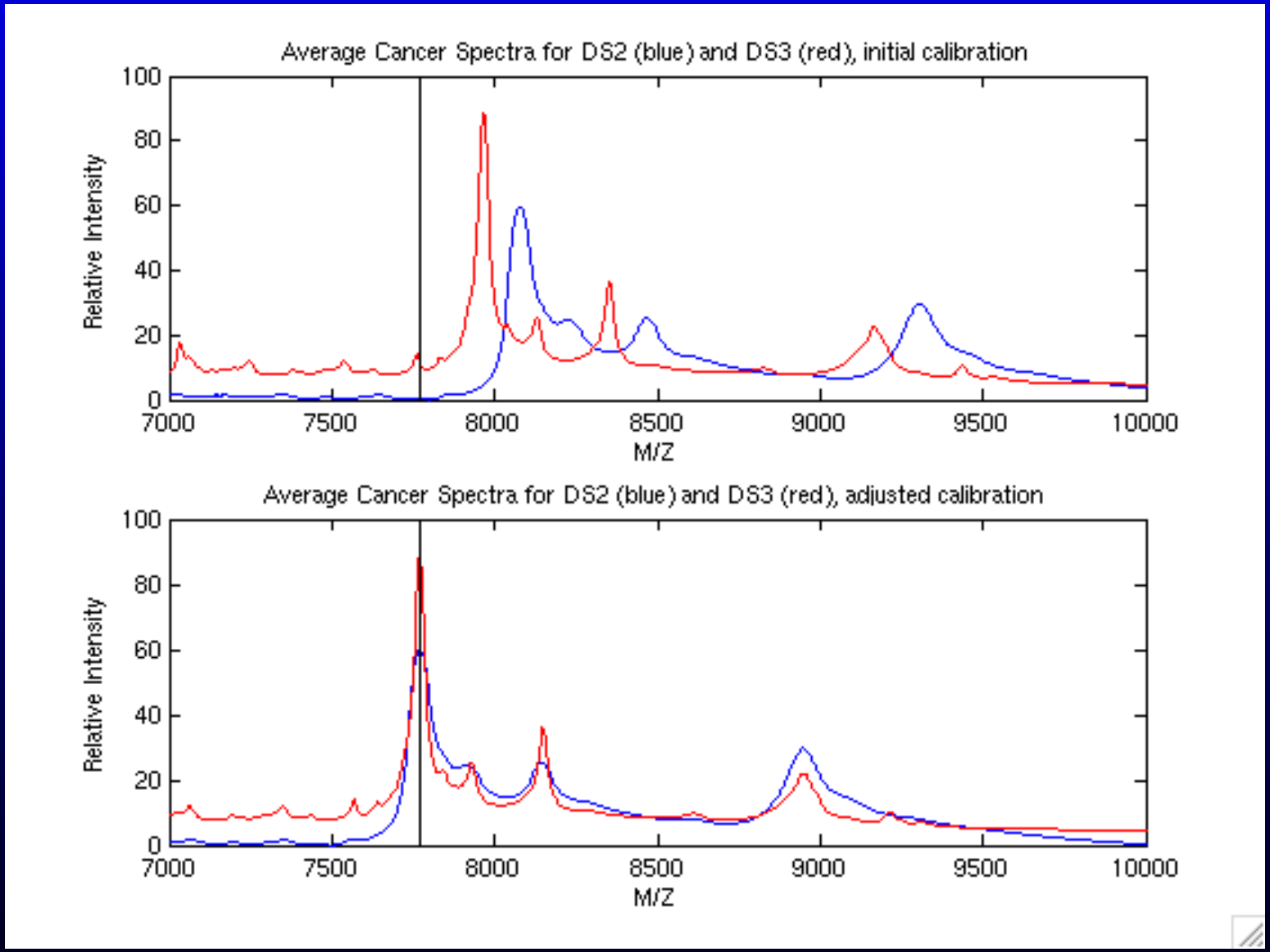


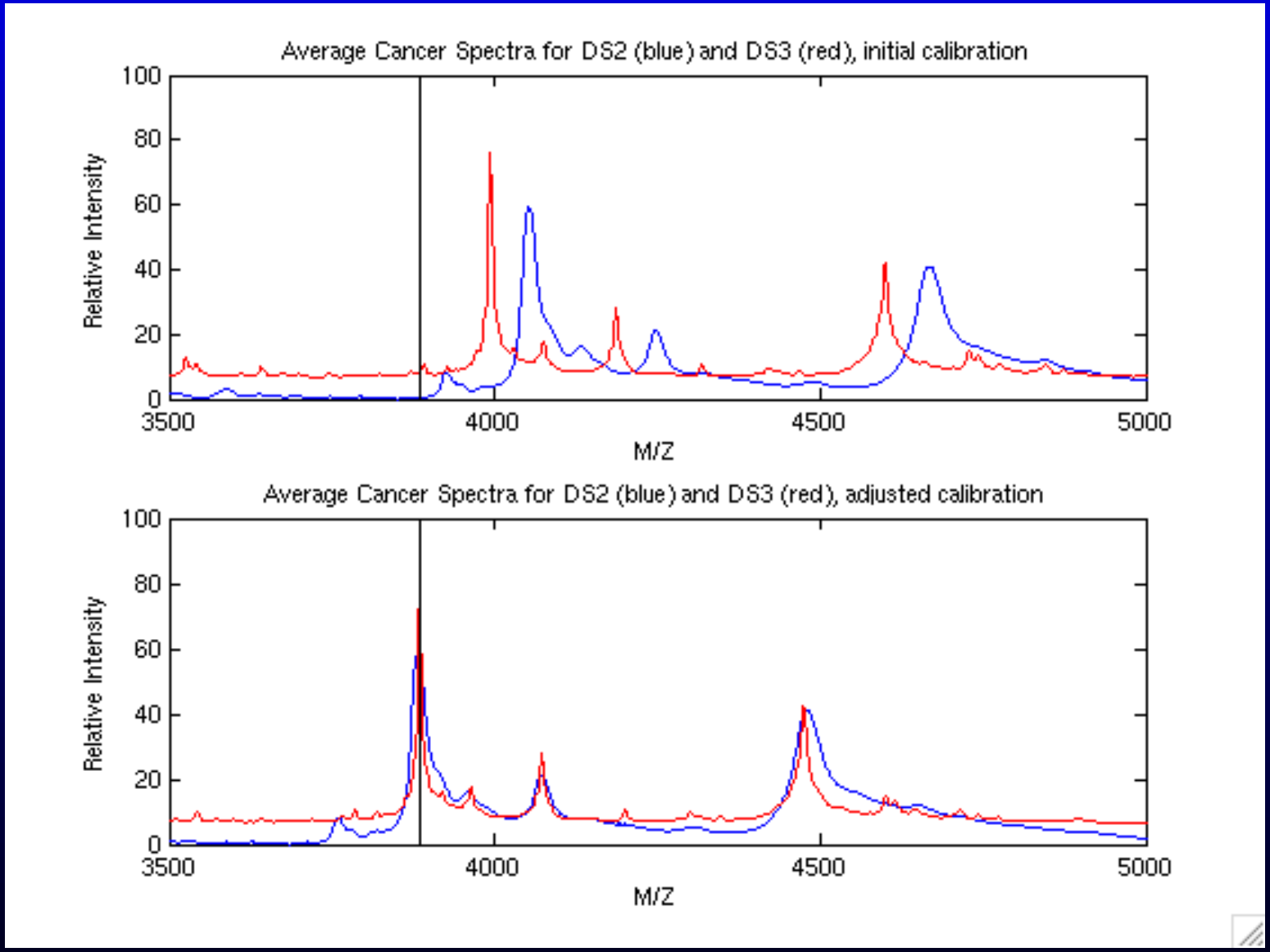
Figure 4. Comparison of mass spectra acquired using a SELDI-TOF MS and a QqTOF MS equipped with a SELDI source.
m/z: Mass to charge ratio; SELDI: Surface-enhanced laser desorption/ionization; TOF MS: Time-of-flight mass spectrometry; Qq: Quadrupole.

Conrads et al (2003), Exp Rev Mol Diag, 3:411-420.

Fixing the Biggest Peak



Fixing the Second Biggest Peak



Their SOP Means Their Calibration is OK

To us, this mistake suggests an error in file export more than a failure to attempt calibration.

Their SOP Means Their Calibration is OK

To us, this mistake suggests an error in file export more than a failure to attempt calibration. But there was a mistake.

Consumers Need to Talk With Producers

They do have a point, it's useful to talk. We sent them some of our concerns months before submitting our stuff for publication.

Consumers Need to Talk With Producers

They do have a point, it's useful to talk. We sent them some of our concerns months before submitting our stuff for publication.

In general, though, we disagree that this needs to be an absolute prerequisite.

Consumers Need to Talk With Producers

They do have a point, it's useful to talk. We sent them some of our concerns months before submitting our stuff for publication.

In general, though, we disagree that this needs to be an absolute prerequisite.

If we don't have access to all of the relevant data, then we need to define standards so that when proteomic papers are published, all data relevant to the reproduction of the results should be made available. The Ciphergen XML files are a good start.

Index
