EXPRESSION ANALYSIS™

Comparative Analysis using the Illumina DASL assay with FFPE tissue

Wendell Jones, PhD

Vice President, Statistics and Bioinformatics

Background

- EA has examined several protocol assay possibilities for reliably analyzing FFPE tissue since 2003
  - Traditional and specialized protocols for 3'-oriented Affymetrix GeneChips
  - Traditional protocols for 3'-oriented Illumina BeadChips
- In short, we have found traditional and specialized methods to either be lacking in detection capability, having high variability, or both
  - Detection rates often as low as background levels
  - High Variability as characterized by > 30% (sometimes 50%) typical Coefficient of Variation (CV) with replicate processing
    - Typical CV for intact RNA is around 5-10%

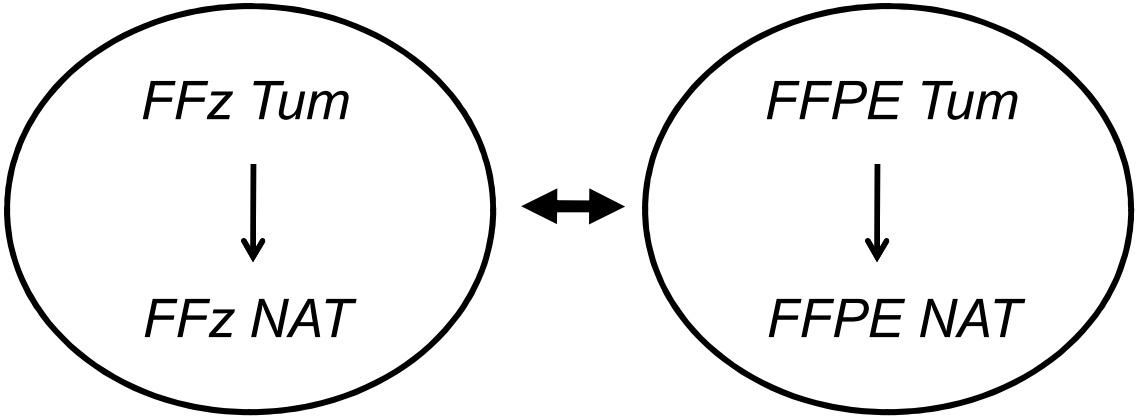
Important Terminology

- Specimen – a biological sample of mostly homogeneous tissue taken from a host
- Sample – a statistical term used in this presentation to represent a collection of specimens that are representative of a population of specimens

Experimental Design

- Fresh Frozen(FFz) vs FFPE RNA (with matched tissue)
  - Illumina WG-DASL protocol and array
  - 4 processing replicates from Lung Tumor and NAT
  - Matched tissue specimen stored FFz and in FFPE

- Compare differential expression between matched tissues when examining preservation method



Measurement Primer

- Microarrays are NOT absolute measurement devices. They do not tell you how many of a given molecule species exists in your RNA specimen. They cannot even provide an ordered list of counts of molecules for different transcripts within an individual specimen.
  - Nonlinear biases in amplification
  - Nonlinear biases in labeling
  - Nonlinear biases in hybridization
  - …
- However, these biases seem to be persistent at the individual transcript level within batches so that one can determine the relative magnitude of a given molecule between experimental groups.

Statistical and Measurement Primer

- For differential detection in populations, sensitivity to differential detection is affected by 4 primary attributes:
  - Biological variation between populations
  - Variation in measurement not due to biology (technical error and bias)
  - Magnitude differences between populations
  - Statistical sample size
  - Number of analytes being tested

- The assay only impacts the 2nd, 3rd and 5th bullet directly and the 4th indirectly related to cost. It cannot impact the 1st bullet whatsoever (inherent variance). The 5th bullet is the multiple testing problem and highlights the statistical dilemma that the more you test, the less sensitive you will be to a particular transcript if you keep the specificity rate fixed.

Statistical and Measurement Primer

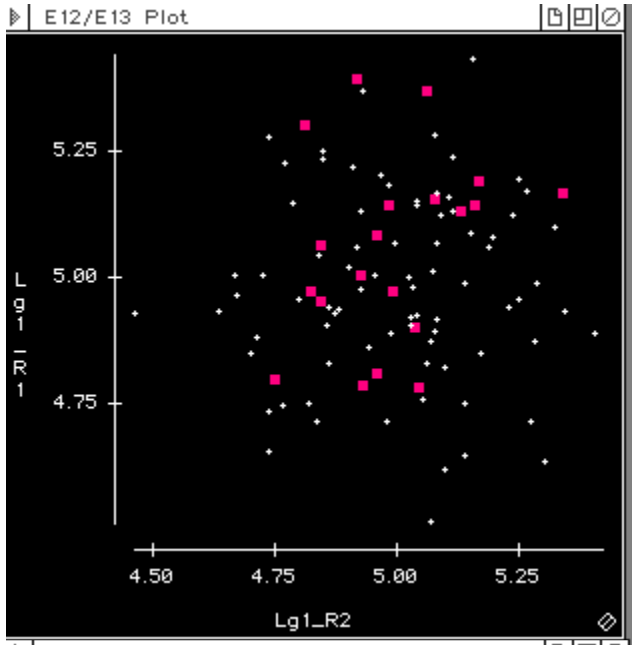
- Variation in measurement not due to biology
  - The preferred way of measuring this attribute is to examine the CV (coefficient of variation) distribution, especially if you feel the measurement error has a strong multiplicative component
- Magnitude differences between populations
  - Surprisingly or not, some platforms “artificially” lower their CVs by compressing signal (eg, PLIER+16) sometimes by biasing the signal. Besides lowering the CV, another ramification is that true differences appear smaller than they really are. Protocols can have this type of impact as well.
- Impact of Detection on both
  - Detection (or LOD) is a key measurement related to both. If you cannot detect the transcript, its variation is effectively very large and one cannot easily distinguish magnitude differences
  - Detection is, in essence, the “table stakes” of differential analysis

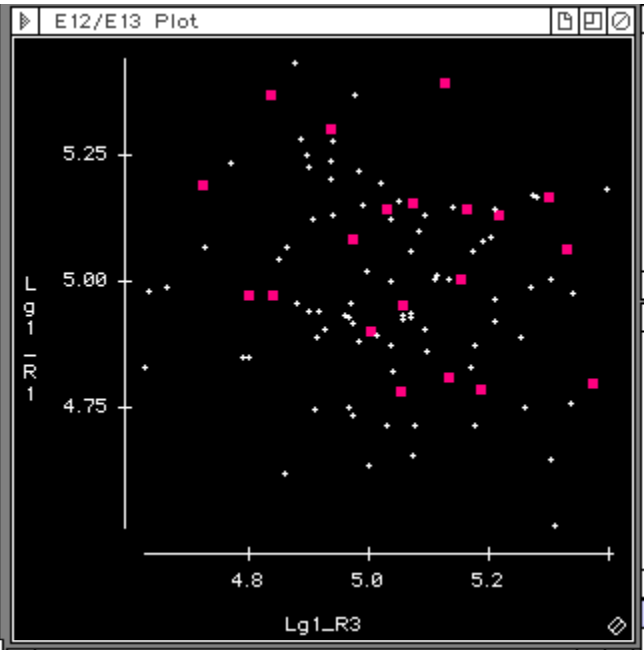
Statistical and Measurement Primer

- Reasons why the CV (distribution) is a more useful measure than pairwise correlation or r^2 between replicates across the transcripts
  - CV is directly related to the power of an experiment
  - CV measures individual transcripts (which is what the statistical tests are addressing)
  - Correlation of signal depends as much on the variation between transcripts as it does the variation within a transcript. The variation between transcripts is largely irrelevant in assessing the differences between populations
  - One can have large or small correlations between replicates while the CV is unchanged

A Simple Simulation Illustrating CV and r

- Take a smaller-than-Whole Genome set of transcripts (100)
- Assume these 100 transcripts all have an average signal value around 32 which is above the noise of the microarray
- Assume that all 100 transcripts have a CV of ~10% (based on MAQC and other replicate results for Illumina)
- What is the expected replicate correlation?





Group A

Pearson Product-Moment Correlation

cases selected according to **Group A or B**

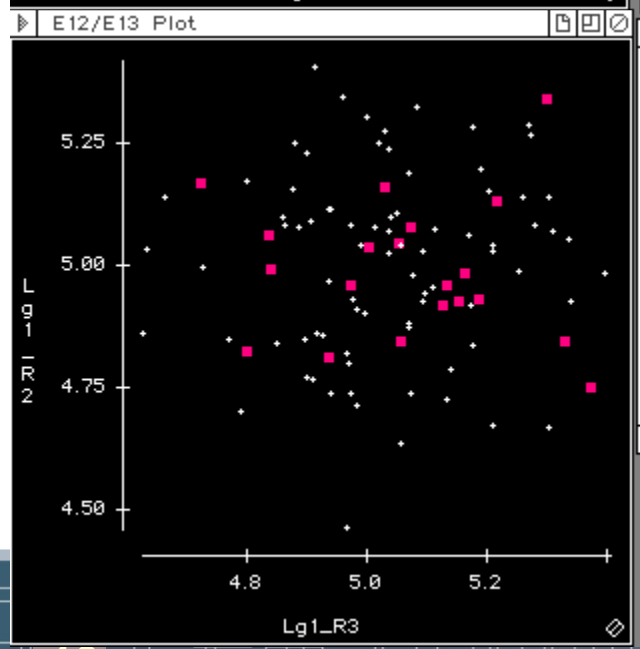
	Lg1_R1	Lg1_R2	Lg1_R3
Lg1_R1	1.000		
Lg1_R2	0.053	1.000	
Lg1_R3	-0.150	0.042	1.000

Correlation

Pearson Product-Moment Correlation

cases selected according to **Group A or B**

	Sig1_R1	Sig1_R2	Sig1_R3
Sig1_R1	1.000		
Sig1_R2	0.046	1.000	
Sig1_R3	-0.154	0.042	1.000



CV G1

Summary of **CV G1**
cases selected according to **Group**
1000 total cases of which 900 are missing

Percentile 10

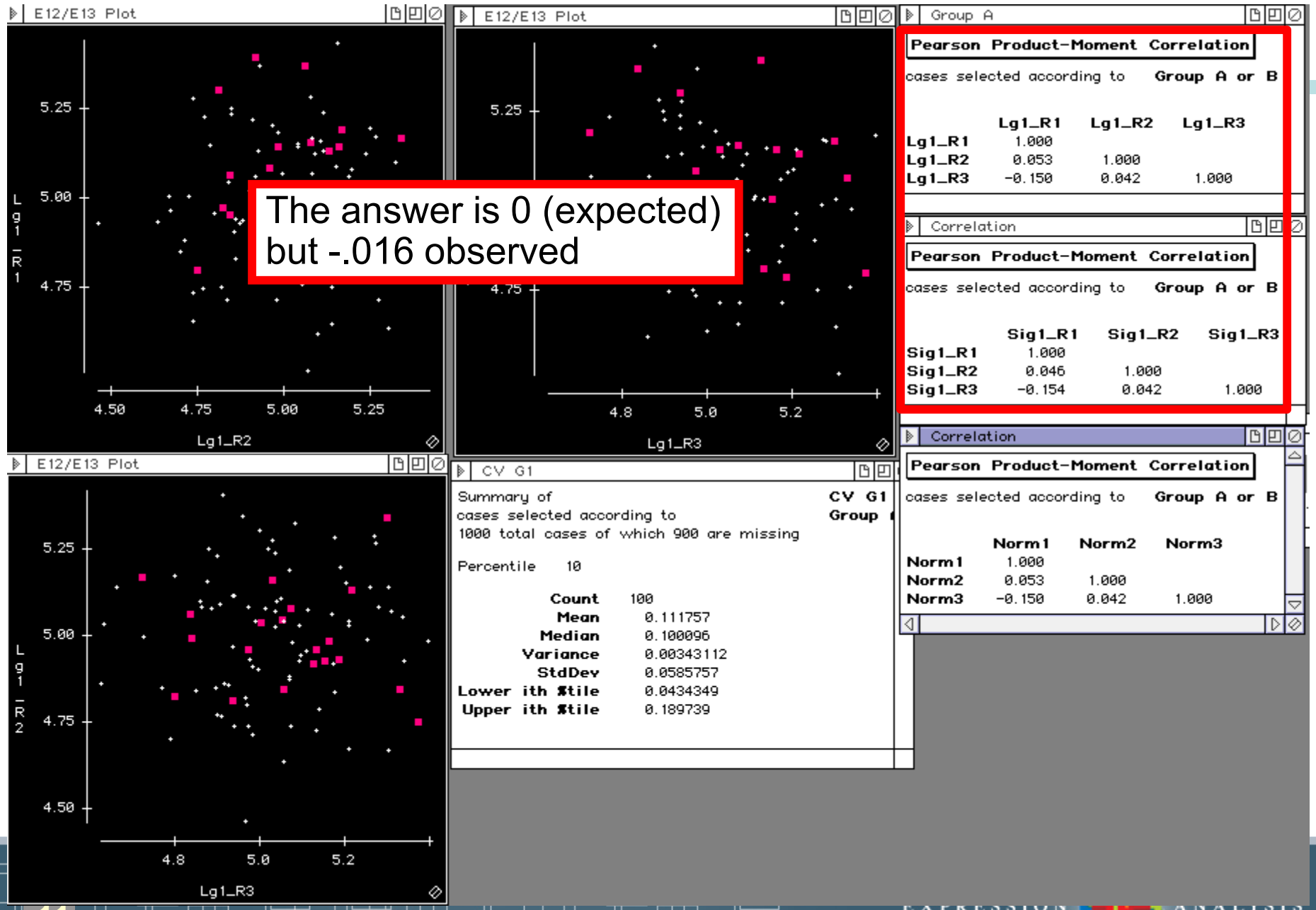
Count	100
Mean	0.111757
Median	0.100096
Variance	0.00343112
StdDev	0.0585757
Lower ith %tile	0.0434349
Upper ith %tile	0.189739

Correlation

Pearson Product-Moment Correlation

cases selected according to **Group A or B**

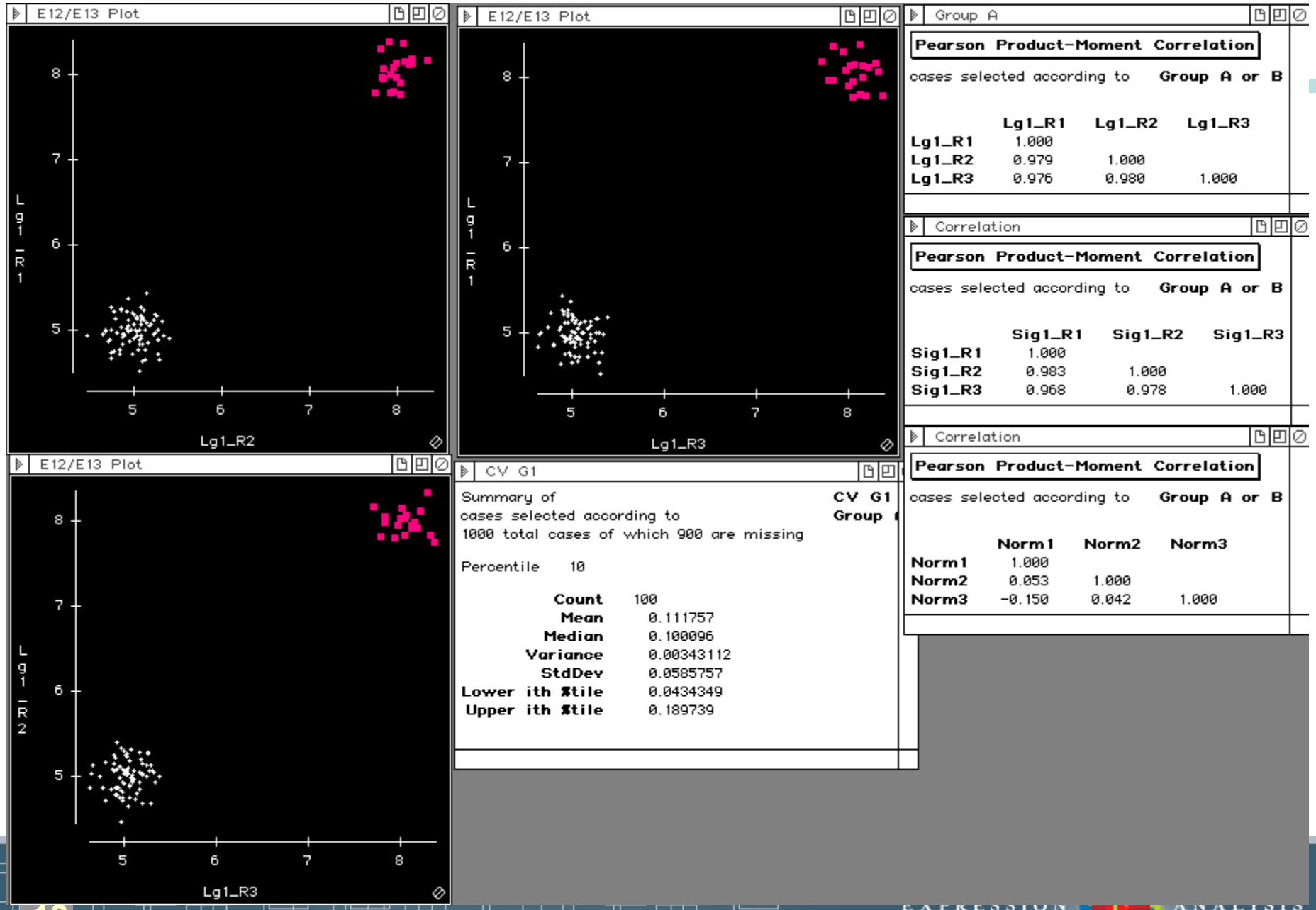
	Norm1	Norm2	Norm3
Norm1	1.000		
Norm2	0.053	1.000	
Norm3	-0.150	0.042	1.000

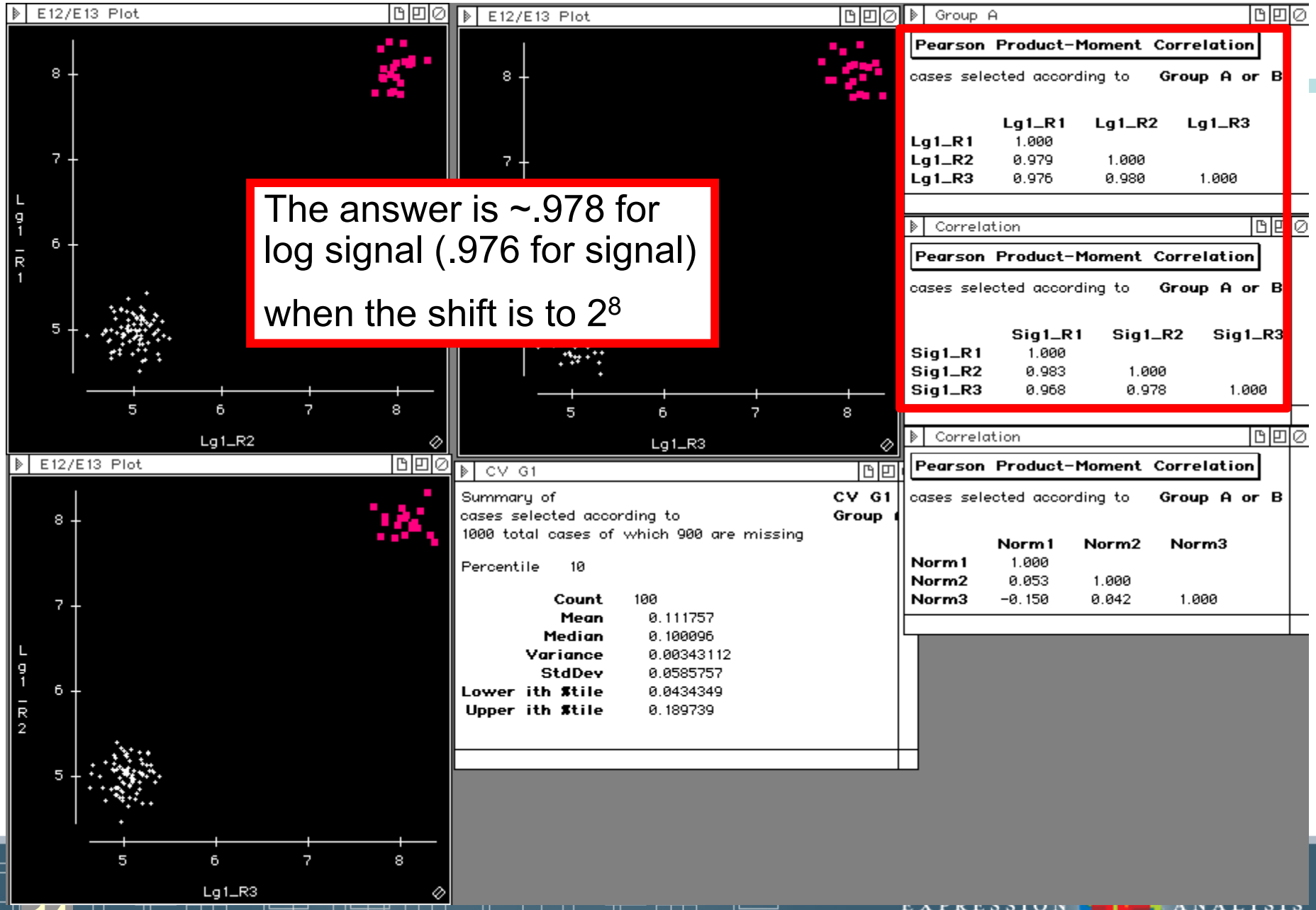


The answer is 0 (expected) but -.016 observed

A Simple Simulation Illustrating CV and r

- Now take these 100 transcripts and divide them into a group of 20 (violet) and a group of 80
- Allow the violet group to change its expected or average signal value (allow it to be higher) up to 2^8 but keep the variation of replicates the same
  - All we are doing is adjusting the mean at the (log) signal level, not the variation
- What is the expected replicate correlation?


E12/E13 Plot
Lg1_R1
Lg1_R2
E12/E13 Plot
Lg1_R1
Lg1_R3
E12/E13 Plot
Lg1_R2
Lg1_R3
CV G1
Summary of
cases selected according to
1000 total cases of which 900 are missing
CV G1
Group
Percentile 10
Count 100
Mean 0.111757
Median 0.100096
Variance 0.00343112
StdDev 0.0585757
Lower ith %tile 0.0434349
Upper ith %tile 0.189739
Group A
Pearson Product-Moment Correlation
cases selected according to Group A or B
Lg1_R1 Lg1_R2 Lg1_R3
Lg1_R1 1.000
Lg1_R2 0.979 1.000
Lg1_R3 0.976 0.980 1.000
Correlation
Pearson Product-Moment Correlation
cases selected according to Group A or B
Sig1_R1 Sig1_R2 Sig1_R3
Sig1_R1 1.000
Sig1_R2 0.983 1.000
Sig1_R3 0.968 0.978 1.000
Correlation
Pearson Product-Moment Correlation
cases selected according to Group A or B
Norm1 Norm2 Norm3
Norm1 1.000
Norm2 0.053 1.000
Norm3 -0.150 0.042 1.000

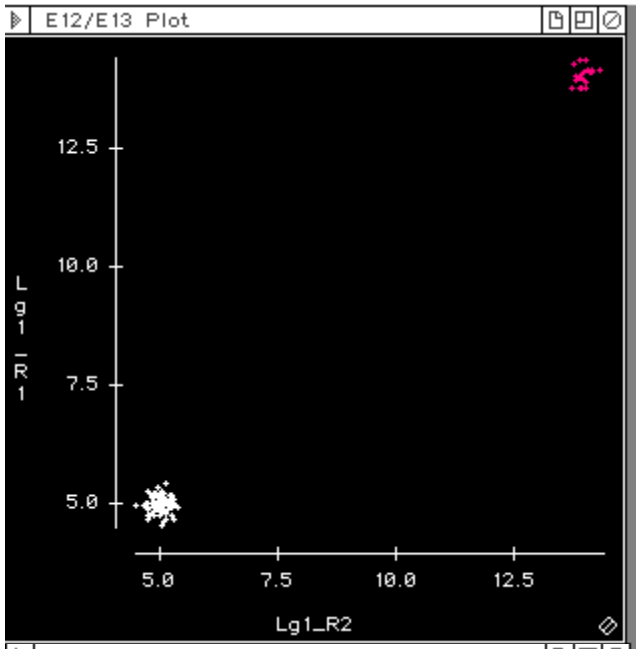


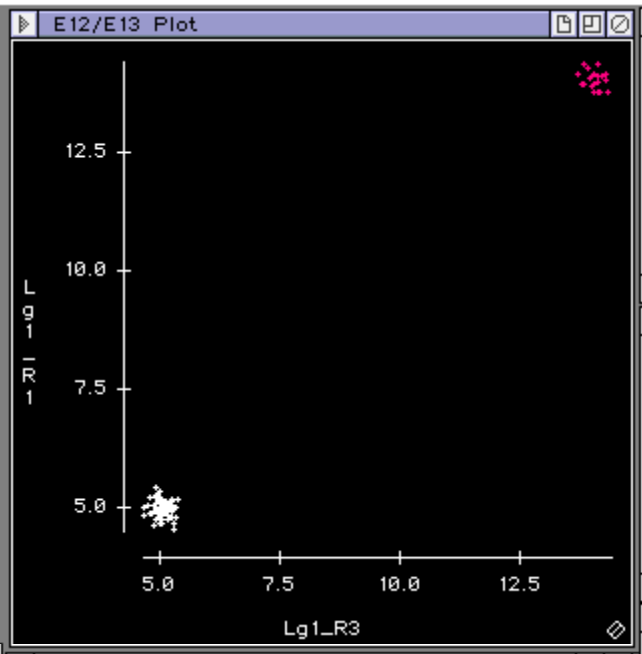
The answer is ~.978 for log signal (.976 for signal)

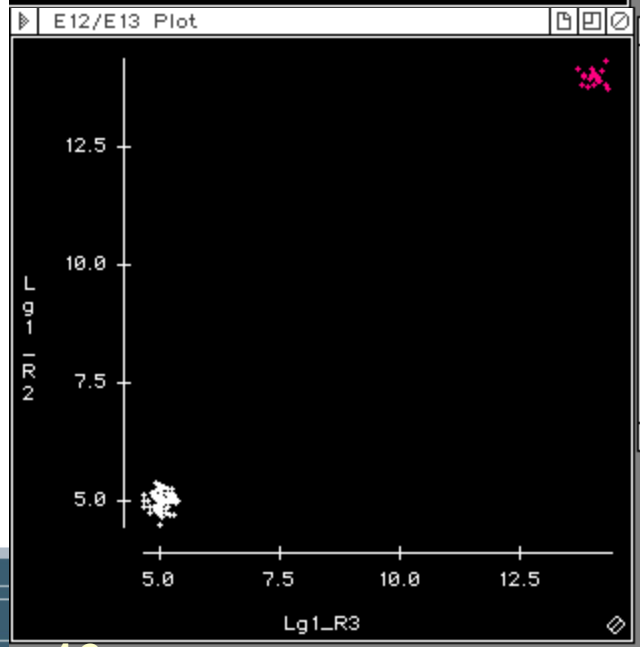
when the shift is to 2^8

A Simple Simulation Illustrating CV and r

- By shifting a smaller subset of the points an order of magnitude (base 10) higher (32 to 258) in expected value, I have changed the average replicate correlation from 0.0 to 0.978
- However, microarray data typically vary by 3 orders of magnitude (base 10)
- What is the expected correlation when I change these 20 to have an expected value of 2^{14}?







CV G1

Summary of — CV G1
cases selected according to — Group
1000 total cases of which 900 are missing

Percentile 10

Count	100
Mean	0.111757
Median	0.100096
Variance	0.00343112
StdDev	0.0585757
Lower ith %tile	0.0434349
Upper ith %tile	0.189739

Group A

Pearson Product-Moment Correlation

cases selected according to **Group A or B**

	Lg1_R1	Lg1_R2	Lg1_R3
Lg1_R1	1.000		
Lg1_R2	0.998	1.000	
Lg1_R3	0.997	0.998	1.000

Correlation

Pearson Product-Moment Correlation

cases selected according to **Group A or B**

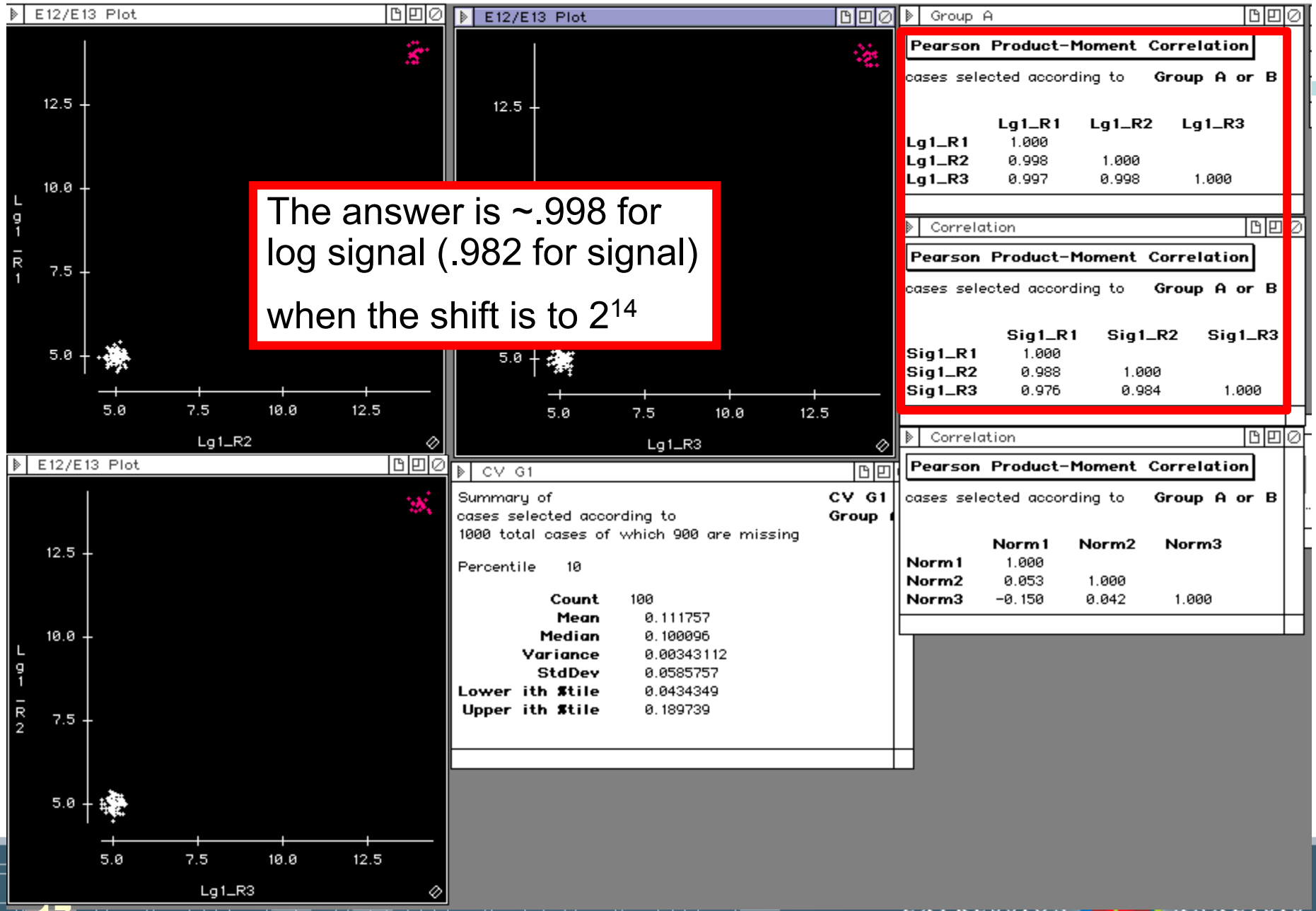
	Sig1_R1	Sig1_R2	Sig1_R3
Sig1_R1	1.000		
Sig1_R2	0.988	1.000	
Sig1_R3	0.976	0.984	1.000

Correlation

Pearson Product-Moment Correlation

cases selected according to **Group A or B**

	Norm1	Norm2	Norm3
Norm1	1.000		
Norm2	0.053	1.000	
Norm3	-0.150	0.042	1.000


E12/E13 Plot
12.5
10.0
7.5
5.0
Lg1_R1
5.0
7.5
10.0
12.5
Lg1_R2
E12/E13 Plot
12.5
5.0
5.0
7.5
10.0
12.5
Lg1_R3
Group A
Pearson Product-Moment Correlation
cases selected according to Group A or B
Lg1_R1 Lg1_R2 Lg1_R3
Lg1_R1 1.000
Lg1_R2 0.998 1.000
Lg1_R3 0.997 0.998 1.000
Correlation
Pearson Product-Moment Correlation
cases selected according to Group A or B
Sig1_R1 Sig1_R2 Sig1_R3
Sig1_R1 1.000
Sig1_R2 0.988 1.000
Sig1_R3 0.976 0.984 1.000
Correlation
Pearson Product-Moment Correlation
cases selected according to Group A or B
Norm1 Norm2 Norm3
Norm1 1.000
Norm2 0.053 1.000
Norm3 -0.150 0.042 1.000
E12/E13 Plot
12.5
10.0
7.5
5.0
Lg1_R2
5.0
7.5
10.0
12.5
Lg1_R3
CV G1
Summary of CV G1
cases selected according to Group
1000 total cases of which 900 are missing
Percentile 10
Count 100
Mean 0.111757
Median 0.100096
Variance 0.00343112
StdDev 0.0585757
Lower ith %tile 0.0434349
Upper ith %tile 0.189739
The answer is ~.998 for log signal (.982 for signal)
when the shift is to 2^{14}

Moral of the Story

- Replicate correlation is a poor measure of replicate variability (but often used)
  - It reflects not only the variation of the signal of the within-transcript replicates but also the between-transcript variation in signal
  - If I have more between-transcript variation, I will automatically increase my correlation and r^2 of replicates no matter the actual variability of the individual transcript measurement
- CV is a more straightforward and appropriate way of measuring replicate variability (repeatability/reproducibility)
  - We have to accommodate the fact that CV may be different for each transcript
  - Summarize in some fashion (using median or other appropriate distribution statistic)

EA Illumina Experiments

- Primary performance measures

  Repeatability

  - Repeatability measures (Coefficient of Variation—Median over the entire array)

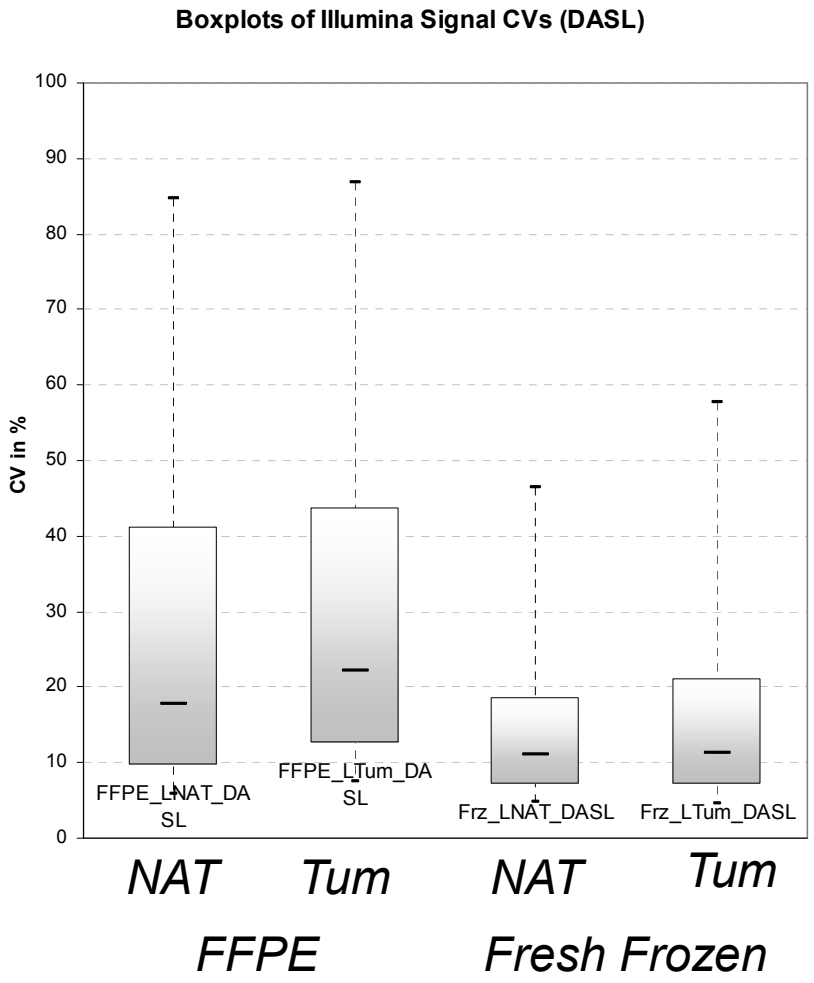
  Transcript Detection

  - Detection rates per hyb and averaged per protocol/source

  Differential Detection

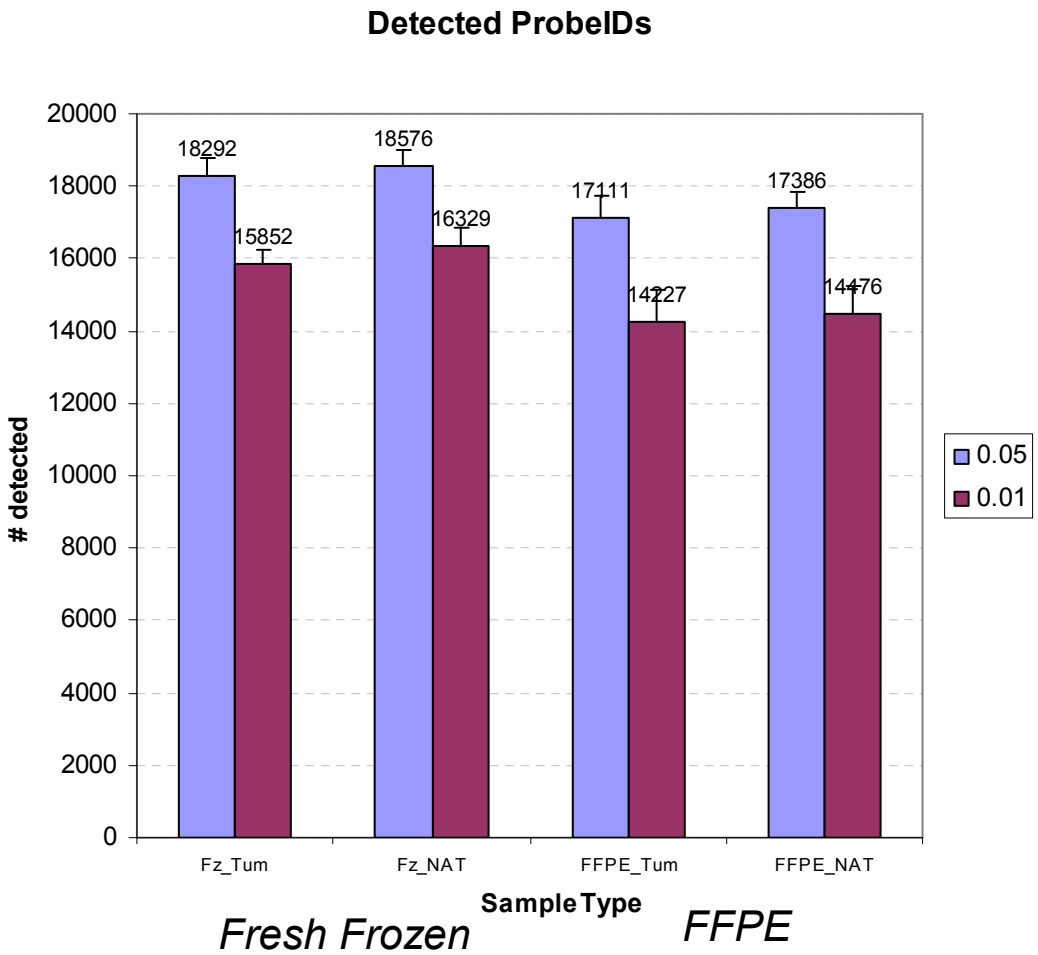
  - Gene List sizes (FDR-based and joint p-value, FC-based rules)
  - Comparability of Log FC or Log Ratio values (correlation)
  - Bias in Log Ratio or FC (magnitude scaling)

Repeatability



- Repeatability
  - Typical CV for NAT FFPE was ~18% and ~22% for Tumor
  - For Fresh Frozen, the typical CVs were just over 11%
- Historic Norms
  - While the CVs are higher for FFPE than with Fresh Frozen, we had not previously seen FFPE protocols having a typical CV below 35% for these tissues

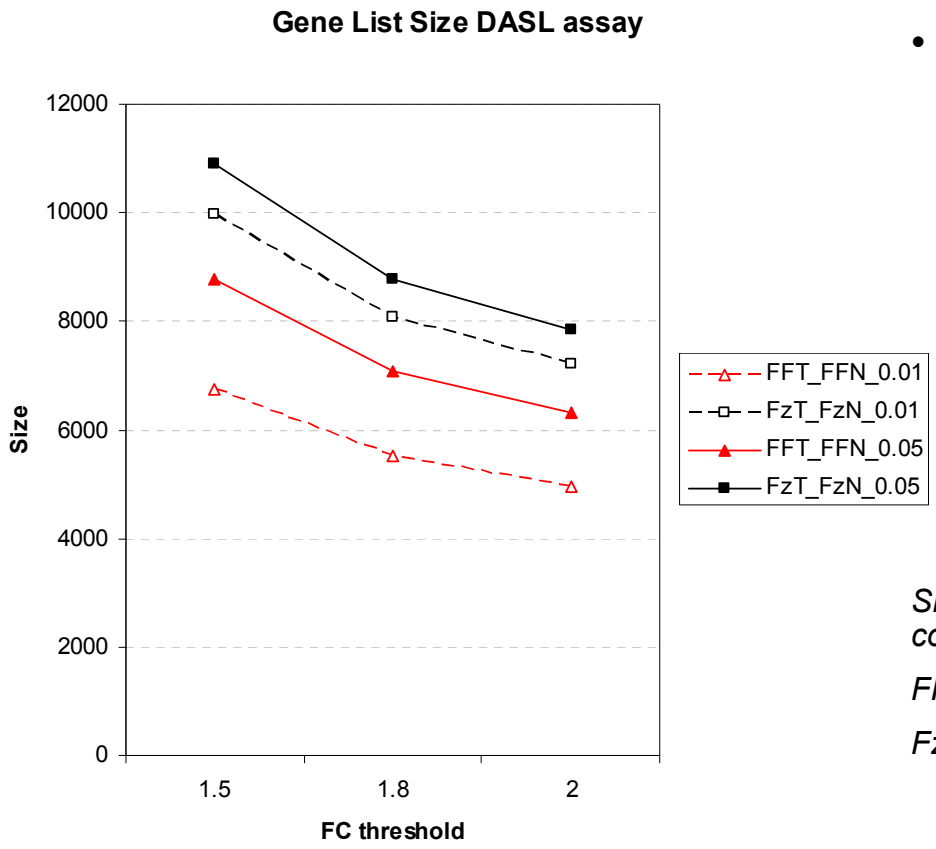
Detection



- The Illumina DASL assay had ~24.5k transcripts available at the time these assays were run.
- FFPE to FFz detection ratios run from 89% to 94%

Average number of transcripts detected in Lung Tumor vs. Lung NAT at the .01 and .05 threshold detection p-value using the DASL assay (out of 24526 total). Bar indicates one standard deviation in replicate detection.

Differential Gene List Sizes



- *For comparison purposes, with the Affymetrix platform for this same tissue type, a particular protocol optimized for FFPE tissue only found 3100 transcripts differentially expressed at p < .05 and FC > 2.*

Size of gene list using the indicated FC threshold combined with a nominal p-value indicated in the legend.

FFT_FFN – FFPE Tumor vs NAT

FzT_FzN – Fresh Frozen Tumor vs NAT

Log Ratio Correlation and Bias - DASL

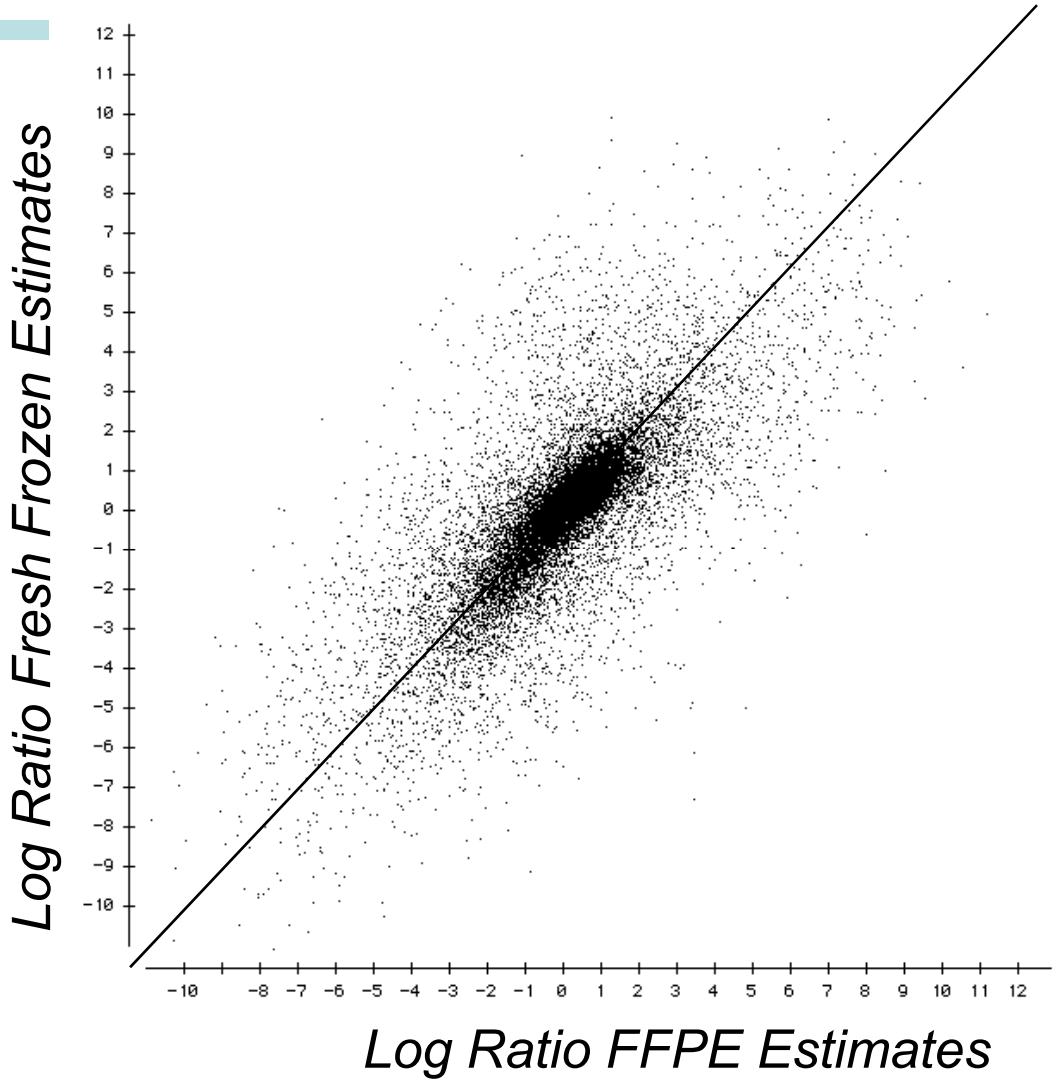
Correlation = .83 using all transcripts, even the ones not detected.

No detectable bias in the log ratio estimates of FFPE (NAT vs Tum)

compared with

Fresh Frozen (NAT vs Tum)

This is different than previous FFPE assays



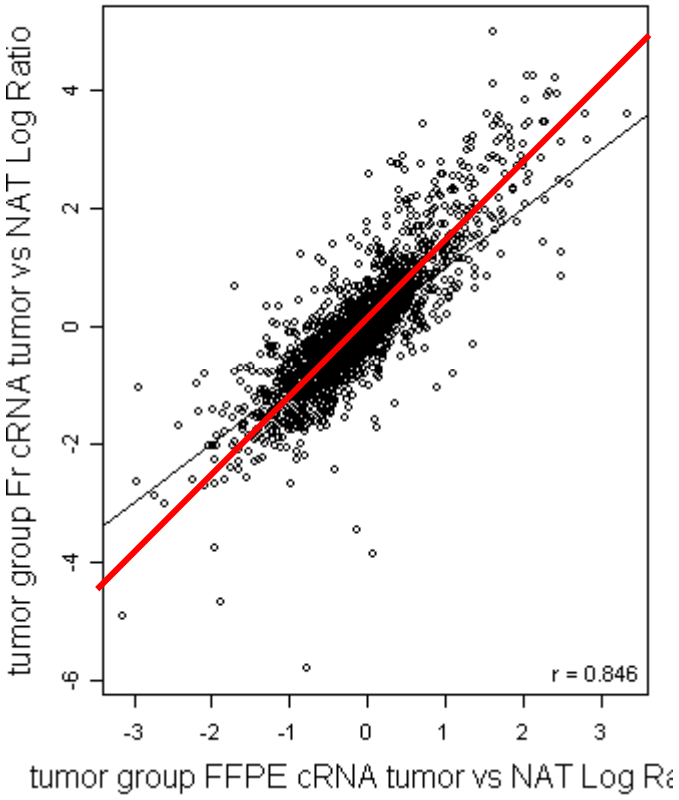
Log Ratio Correlation and Bias – Previous Assays

For this FFPE assay, there is a pronounced bias in the Log Ratio estimates of FFPE (NAT vs Tum)

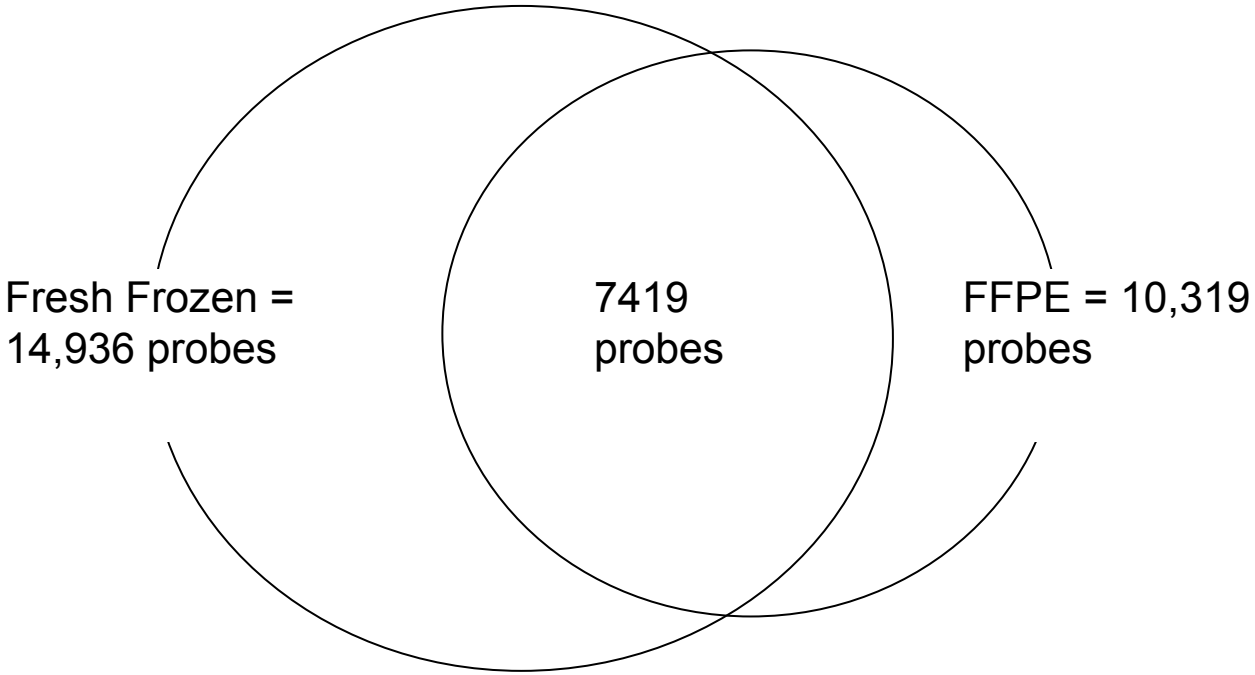
compared with

Fresh Frozen (NAT vs Tum)

The impact is having a smaller gene list when using a threshold-rule based on magnitude of the difference as this assay shrinks the FC estimate.



Differentially Expressed Gene List Comparison



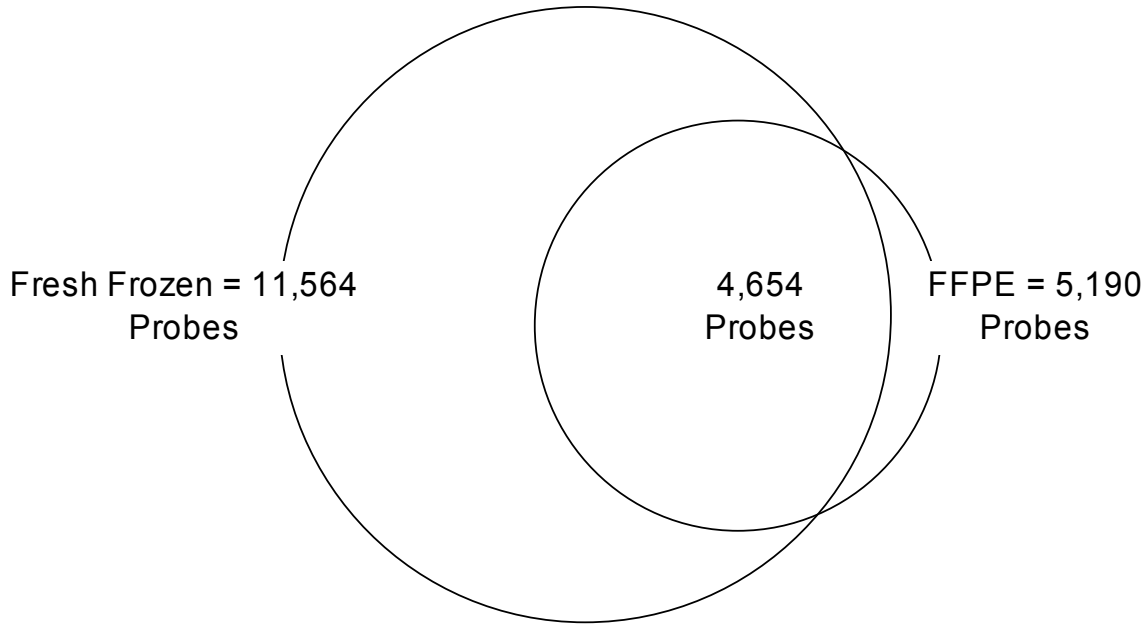
Illumina DASL Assay – Tumor relative to normal adjacent for fresh-frozen and FFPE tissues using permutation analysis with a modified t-statistic and a false discovery rate less than 0.05 for both FFPE and fresh-frozen.

Summary

- Illumina DASL assay
  - Detection rates are roughly 90% of what would be found when using fresh frozen
  - Median CVs are elevated for FFPE tissue compared to fresh-frozen but moderately so and much improved over previous FFPE protocols.
  - The bias in log ratio estimates that compressed apparent differences with other protocols does not seem to exist for the DASL protocol
  - Correlation in log ratio values appears comparable to other methods even when using the entire array (detected and non-detected) for DASL for detected transcripts for other protocols
  - Gene lists sizes for FFPE are large and overlap to a great degree with the fresh-frozen list (but not equivalent due to higher CV).

Differentially Expressed Gene List Comparison

Competing protocol



Tumor relative to normal adjacent for fresh-frozen and FFPE tissues, false discovery rates less than 0.05.