



ViaLogy VMaxS Calibration Study: A Case Study

Sandeep Gulati
Chief Science Officer
ViaLogy Corp.

Prof. Charles Wang
Director, Microarray Core
Cedars-Sinai Medical Center

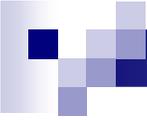
May 2, 2005





Calibration Rationale

- As Analyte Volumes Get Smaller It Becomes Increasingly Difficult (Impossible) To Accurately Measure Linear Dose Response Due To Multiple, Random Non-Linear Interactions of Analytes and Background
- There Is Valuable Expression Information In The ‘Noise’
- Platform Calibration Increases Detection Sensitivity, Linear Dynamic Range And Improves Platform Resolution
- The Result – Greater Dynamic Range With Better Reproducibility



Study Goals

- Independent Analysis of the Same Dataset Yields Same Results
- Experiments Using Identical Samples, Protocols and Systems Produce Statistically Similar Results
- Calibrate The Analysis Method to Deliver Standardized Performance on the Leading Microarray Platforms
 - ABI, Affymetrix, Agilent, GE & Illumina
 - Initial Platform: Affymetrix GeneChip®
- Preserve Biological Variance and Dynamic Range
 - Preserve Gene Expression Information ‘In The Noise’
- K. I. S.

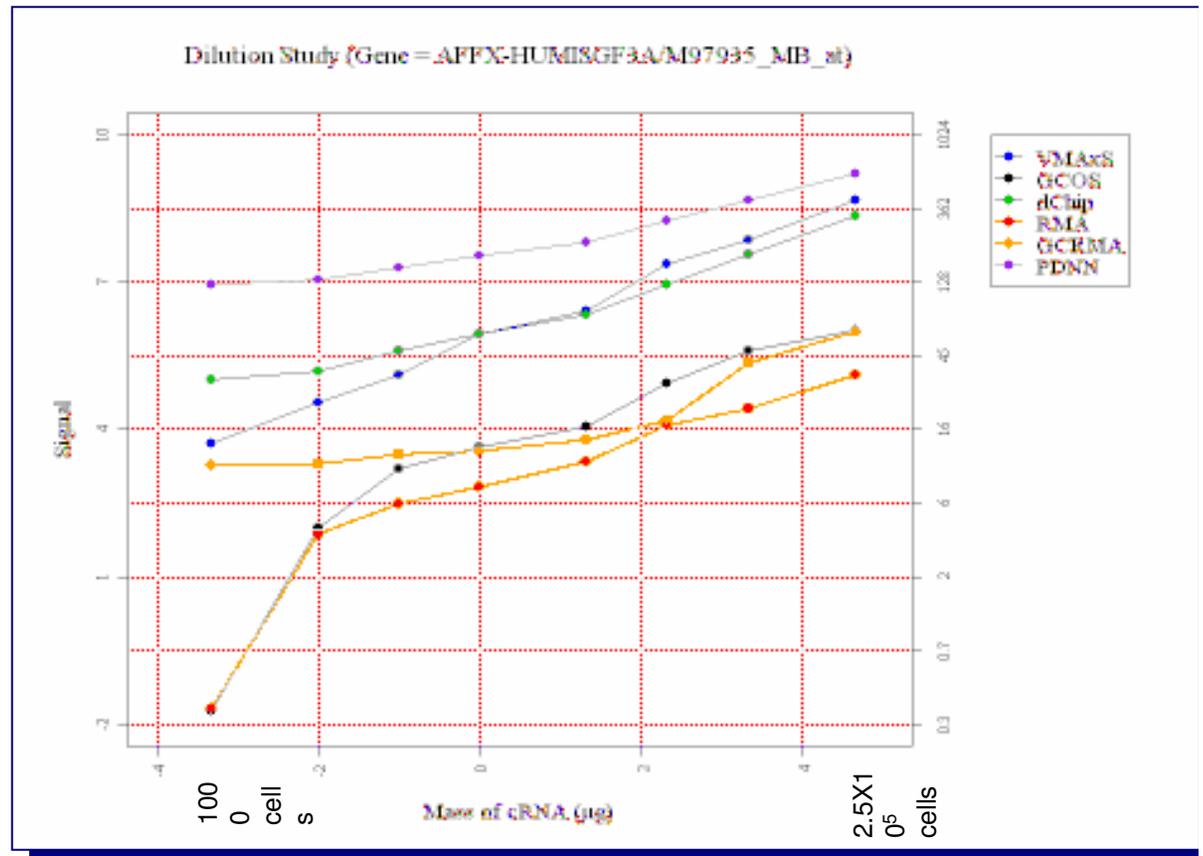


Calibration Objectives

- Establish An Analytical Performance Specification for VMAS
 - Phase 1: 15ug to 100ng starting cRNA
Complex biological backgrounds
<5% between run CVs
 - Phase 2: < 100ng starting cRNA
- Select Optimal Feature-level Probes for Quantitation
 - Assess response linearity, sensitivity and specificity
- Develop Linearity Model for High, Moderate, Low & Very Low Expressors
 - Complex samples with genes exhibiting 6 log variation in expression levels

Methodology

- Step 1: Calibrate GeneChip® HG U133 Plus 2.0 system based on carefully controlled hybridizations
- Compare results by 6 analytical methods
- Step 2: Validate with a set of over 1000 experimental arrays



Dilution Study Design (24 arrays)

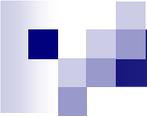
- Stratagene Universal Reference RNA
- Triplicate technical analysis (hybridizations)
- Biotin labeling and fragmentation on bulk sample RNA before aliquots were frozen

Dilution Study	
Sample	Amount of cRNA from Reference RNA
1	25 µg
2	10 µg
3	5 µg
4	2.5 µg
5	1 µg
6	0.5 µg
7	0.25 µg
8	0.1 µg

Tissue Mixing Study Design (16 Arrays)

- Ambion Diploid Lung Total RNA used to emulate a complex background interference
- Duplicate hybridizations

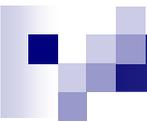
Tissue Mixing Experiments		
Stratagene Universal Total RNA	Ambion Diploid Lung Total RNA	Background Interference Ratio
(Specific RNA)	(Complex Interference to Specific Signal)	
	10μg	1:20
0.5μg +	5μg	1:10
	2.5μg	1:5
	1μg	1:2
	0.5μg	1:1
0.1μg +	10μg	1:100
	1μg	1:10
	0.1μg	1:1



RNA Selection

- Dilution Study
 - Commercially Available Reference Material
 - Maximal GeneChip (55K probe sets) Coverage
 - Abundant Expression Data Available on Stratagene RNA
 - Affy chips, RT-PCR, SAGE
 - Already Used As A Calibration Assay By Most Labs

- Tissue Mixing
 - Commercially Available Material
 - Genes Not Expressed In UHRNA (Lung)
 - Full Characterization of RNA Is Not Germane To This Study
 - But must have some known differences with Reference RNA
 - Purpose: Induce Background Interference to Universal RNA
 - Simulate (Signal + 100 X Background) Interference



Criteria for Selecting Tracking Genes

- Focus Set: 15 Genes Known Present in Stratagene Universal RNA
 - 3 genes in 4 expression groups + 3 house keeping genes
 - Grouping determined by an alternative method

High Exp. : Moderate Exp. : Weak Exp. : Ultra-Weak Exp.
>1000 fold : 100 fold : 10 fold : <1

- Typical Of Most Real-life Experiments
- Different Analytical Methods May Show Up To 2-log Range Variation In Between-run CVs Across The 4 Expression Groups
 - Significant biology is lost if expression quality is not consistent across different expressors
 - Richard Simon (NCI): “.. *very important to have consistent and comparable fold change quality over all types of expressors in same sample so that users can start with small quantities of material and still trust their data..*”

Selected Tracking Genes

- Existing StaRT-PCR Data on Reference RNA
- SAGE Tag Performance

StaRT-PCR mRNAs per 10 ⁶ b-actin	SAGE 51973 tags total	Gene symbol	Affy-probeset	Description
5.3E+05	123	LGALS1	201105_at	lectin, galactoside-binding, soluble, 1 (galectin 1)
2.6E+05	9	HINT	1555960_at	histidine triad nucleotide-binding protein
1.4E+05	4	FN1	1558199_at	fibronectin 1
5.7E+04	1	MSH6	202911_at	mutS (E. coli) homolog 6
1.1E+04	1	GSS	201415_at	glutathione synthetase
5.7E+03	1	ANXA4	201302_at	annexin A4
1.1E+03	0	PAK1	1565772_at	p21/Cdc42/Rac1-activated kinase 1 (yeast Ste20-related)
5.5E+02	0	RABIF	204477_at	RAB interacting factor
1.3E+02	0	IGF1	209540_at	insulin-like growth factor 1 (somatomedin C)
5.4E+01	0	MAPK8	210477_at	mitogen-activated protein kinase 8
1.8E+01	0	CYP2E	1431_at	cytochrome P450, subfamily IIE (ethanol-inducible)
6.2E+00	0	KDR	023934_at	kinase insert domain receptor (a type III receptor tyrosine kinase)
0.0E+00	0	TNF	207113_s_at	tumor necrosis factor (TNF superfamily, member 2)



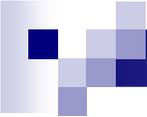
Data Quality Metrics

- **Data Compression: Fold Change As Compared to RT-PCR Taqman Assay (Δ_{CT})**
 - In general it has been observed that microarray expression levels are compressed by 3:1; 5:1; 7:1; ... depending on the different expression groups
 - Degradation of compression ratios occurs in the presence of interfering background
- **Optimal Probes Selection**
 - Determine features that exhibit monotonic expression response to dilution; and,
 - Can be used for calibration and standardization of expression



Data Quality Metrics

- Reproducibility
 - Across Runs/Triplicates
 - Highly Controlled Experiments Performed to Limit Variability
 - Operator, batch, instrument, labeling
 - All Credible Users/Core Labs Today Employ Tight Controls
 - In more complex backgrounds (tissue mixing experiments)
 - Assess %CV Expression Degradation Across the 4 Groups



Dynamic Range

- Analytical Tools Must Address the Range of Biology
- Study Design
 - Capture 5 log Variation in Expression for Different Genes
 - 2.5 log Variation in Starting Material
 - 2 log Variation in Noise
 - Overall 6 – 7 log Variation in Expected Expression Performance
 - (with 5:1 compression over biology)
- Linearize Expression to Match Taqman CT range: 18 ± 1 to 36 ± 2
- Rationale: Do Not Focus Only On A Subset of Genes That Are Concordantly Observed On All Platforms By All Laboratories.
- RE: High Risk of Missing Dynamical Range of Intrinsic Biology
 - “Common Good” Likely Is Not Adequate Standard for Clinical Data



Starting Data

- Processing Raw Data Files (e.g., DAT file for GeneChip®)
Produces Better Results Than Starting With Preprocessed CEL Files
- Recommend Having Raw Data Available To Analyze As New
Signal Processing Tools Become Available
 - It is suggested that GeneChip data be submitted as DAT files



Summary

- Clinical Quality Results Are Achievable
- Standardization and Calibration Assays and Protocols Will Produce Data Required for Clinical Quality Standards and Advance Biological and Clinical Decisions
- Broad Platform Calibration Is A Must
 - Cost-prohibitive To Validate Every Microarray Result With A Reference Method.
- Strongly Suggest Including Genes With Low Expression In The Control Group For MAQC Study

Acknowledgements

- FDA Leming Shi, Weida Tong
- Cedars-Sinai Charles Wang, Jun Xu
- Stratagene Gavin Fisher
- Ambion Bob Setterquist, David Doris
- ViaLogy Bud Bromley, Jim Breaux,
Cecilie Boysen, Sandeep Gulati
- Contact: Bud Bromley bud.bromley@vialogy.com
+1(858) 832-1099

