Variability in Gene Expression in Healthy Volunteers

Andrew Strahs, Ph.D.
Global Biostatistics & Programming

BASS
November 7, 2006

Wyeth Research

Biomarkers

- **A Biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working Group)**
- **Drugs often fail in clinical development due to safety concerns or poor efficacy despite costly R&D efforts**
- **Biomarkers have the potential to change the risk:benefit ratio in drug development and in the clinic**
- **There is a need for surrogate and predictive biomarkers of safety and efficacy**

Expression Profiling Pharmacogenomics

- **Expression profiling at the RNA or protein level may elucidate mechanism or identify biomarkers of disease or drug response**
- **Dynamic biomarkers that change with disease state or severity and in response to treatment may be particularly useful**
- **Profiling may be performed in disease tissue or surrogate tissue such as peripheral blood mononuclear cells or plasma**
- **Techniques may be applied in preclinical studies, clinical trials, and in the clinic**

Clinical Pharmacogenomics

Target Disease Tissue Biopsies and Blood at Baseline and During Treatment

Expression Profiles (RNA, protein) of Disease-associated and Drug-responsive Genes

Identify Biomarkers of Disease, Drug Activity and Response

Diagnostic Assay

- **Disease-Associated Biomarkers and Mechanism**
  - Indication-specific biomarkers as response measures
- **Predictive Markers of Response or Risk**
  - Patient selection/stratification pretreatment
- **Pharmacodynamic markers of exposure/response**
  - Support dose selection and measure biological activity
- **Surrogate Endpoints**
  - Validated as indicator of clinical benefit

Rationale for Healthy Volunteers Observational Study

- **Blood is a source of potential mRNA biomarkers**
  - Disease-associated biomarkers (Twine et al., 2003)
  - Outcome-associated biomarkers (Burczynski et al., 2005)
- **Lack of reference levels for PBMC transcriptome is an obstacle to assessing potential of blood biomarkers**
- **Whitney et al. (2003) profiled 75 healthy subjects at 1 time point and 16 subjects at >1 time point on cDNA microarrays**
- **Radich et al. (2004) profiled 15 subjects at multiple time points on inkjet DNA microarrays**
- **McLoughlin et al. (2006) studied 48 genes in 131 healthy volunteers**
- **Wyeth's Healthy Volunteers Observational Study in 400 subjects at ~4 time points conducted to characterize the intra- and inter-subject variability in peripheral blood mononuclear cells**

Healthy Volunteer Observational Study Design

- **Exploratory, outpatient study of longitudinal gene and protein expression in healthy volunteers over a period of 12 months**
  - **Detailed information about blood chemistry, medication usage, and general health, also collected**
- **Subjects enrolled in parallel into one of 18 strata defined by combinations of**
  - **Age (20-39, 40-59, 60-79 years)**
  - **Sex (male, female)**
  - **Race (Asian, Black, White)**
- **2 strata of smokers (males and females, aged 20-39) also included**
- **Each stratum consisted of ~20 subjects.**
- **Total enrollment of ~400 subjects X 6 timepoints = 2400 samples**
- **Evaluation of ~400 subjects x ~4 timepoints = ~1600 evaluable profiles**

Sample Collection and Processing

- **Blood collected into Vacutainer BD-CPT tubes and shipped overnight at ambient temperature**
- **Peripheral blood mononuclear cells (PBMCs) purified, then frozen at –80C**
- **RNA prepared from PBMCs and analyzed on human U133A Affymetrix Gene Chip**
  - 22,283 human sequences on chip

Analysis Data set
'PG Evaluable' subjects

- **382 subjects***
  - 117 Asians / 117 Blacks / 148 whites
  - 163 Canadians / 75 Germans / 144 Americans
  - 197 females / 185 males
  - 165 Aged 20-39 / 117 Aged 40-59 / 100 Aged 60-79
  - 40 Smokers / 342 Non-Smokers

Country vs. Race	**Race**		
Country	Asian	Black	White
Canada	41	68	54
Germany	1	0	74
USA	75	49	20

***Evaluable subjects contributing qc'd chips**

Analysis Data set
'PG Evaluable' Gene Chips

- **1634 gene chips**

Numbers of Chips in analysis data set, per subject							
#chips	1	2	3	4	5	6	7
#subjects	2	11	41	211	68	40	9

- Intervals between samples for chips in analysis data set: (days)

2, 8, 11, 13, 13, … , 251, 252, 254, 266, 357

25th percentile: 54 days

50th percentile: 82 days (=mean)

75th percentile: 88 days

Regression Analysis

- **'Primary' analysis is mixed model ANCOVA (one model per probe set) with random subject effect (compound symmetry)**
  - Response variable is $\log_2$(signal); signal from MAS 5.0, signal algorithm
  - Explanatory variables
    - Factors: age (20-39, 40-59, 60-79), sex (M, F) , race (W, B, A) , country (USA, CAN, DEU),
    - Continuous covariates: ratio of monocytes to lymphocytes (sample cell composition), average of beta-actin and GAPDH ratios (RNA quality)
- **Results include the following:**
  - Intra- and inter- subject variance estimates
  - F-test P-values for significance of each of the explanatory variables
  - Estimates of marginal means of expression for each level of every stratifying variable or factor used in computation of fold differences

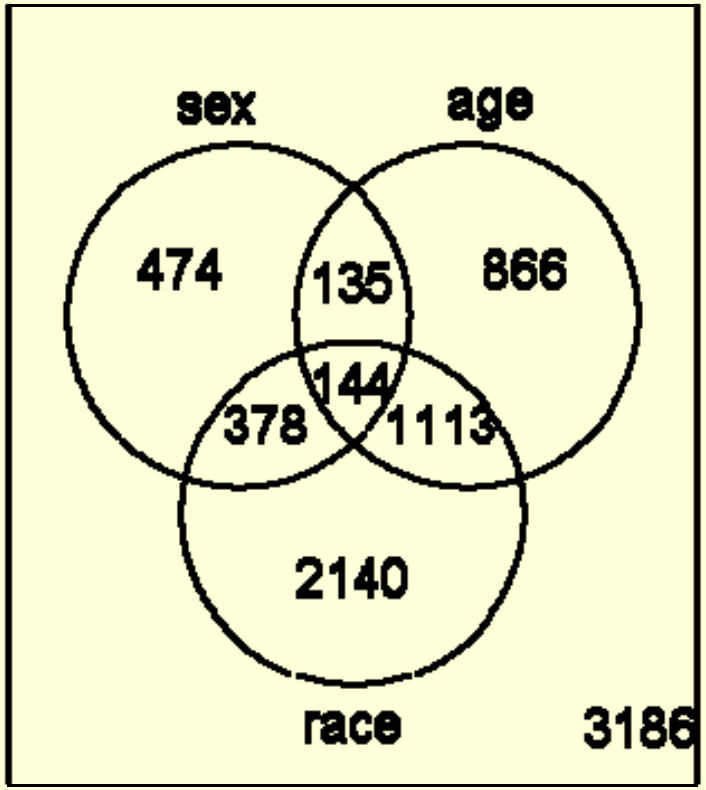
Effects of Covariates on 8436 Probe Sets (out of 22283) with Geometric Mean of Signal > 50

Variable	P<0.05	FDR*<0.0001	2-fold◑	
Sex	4158	1131	5	
Race	6650	3775	12/6/8	(W:B / W:A / B:A)
Age (bin)	5640	2258	0/1/0	(20-39:40-59 / 20-39:60-79 / 40-59:60-79)
Country	5968	2866	1/38/54	(USA:CAN / USA:GER / CAN:GER)
Mono:Lymph	6294	4305	NA	
BA GAPDH	6951	5528	NA	

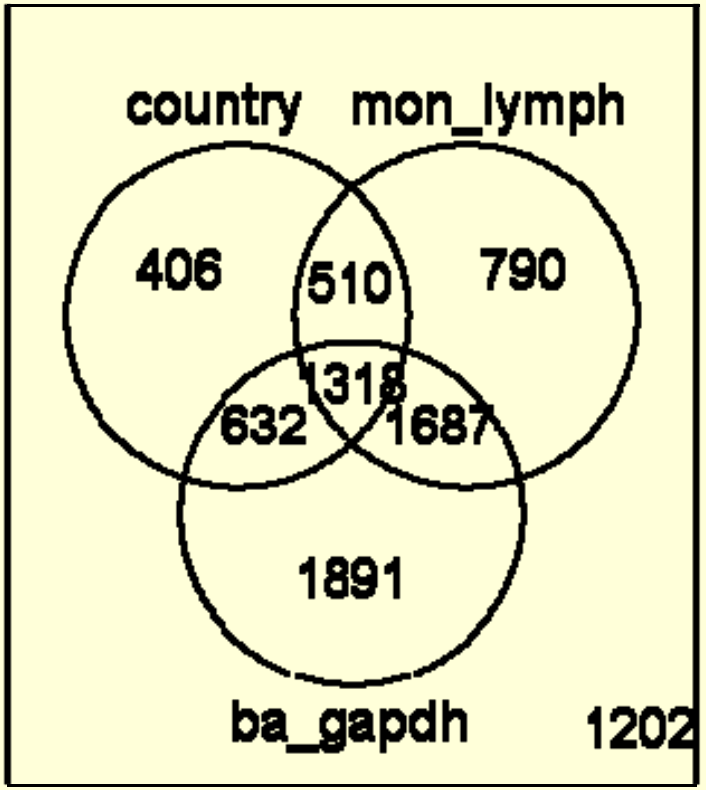
*FDR adjustments are made for each covariate separately

Effects of Covariates, cont.

Probesets assoc. with demography



Probesets assoc. with sample characteristics



Venn diagrams describe overlap in associated (FDR<0.0001) probesets

Wyeth Research

Expression Level vs. Variability

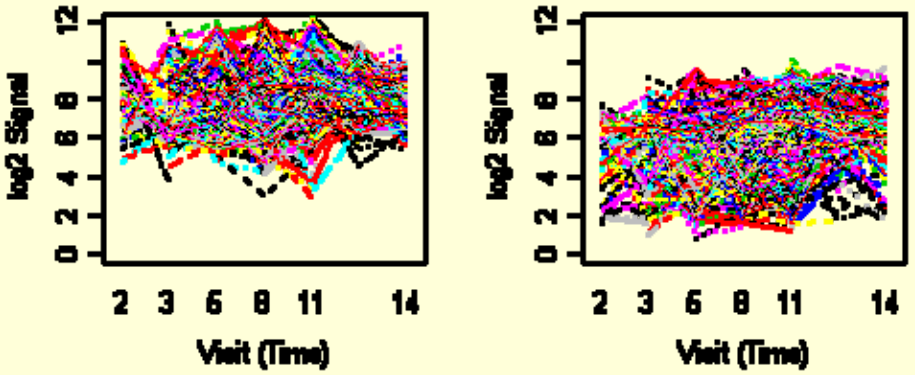
22,283 probesets, partitioned by expression level and within-subject coefficient of variation

Covariates included*	Within-subject CV		
Expression level (Geometric mean of signal)	<25%	25-50%	>50%
<50	91	4535	9221
50-500	3023	3852	268
>500	975	307	11

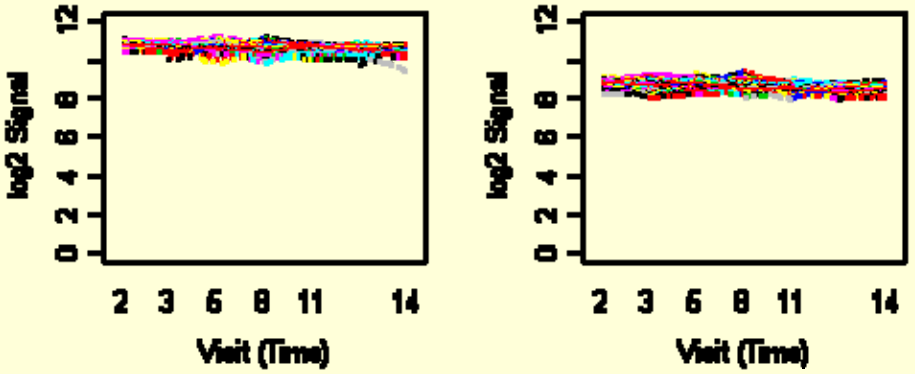
***Counts are similar when only random subject effect is included in model**

Genes with High and Low Intra-subject Coefficient of Variation

High CV:



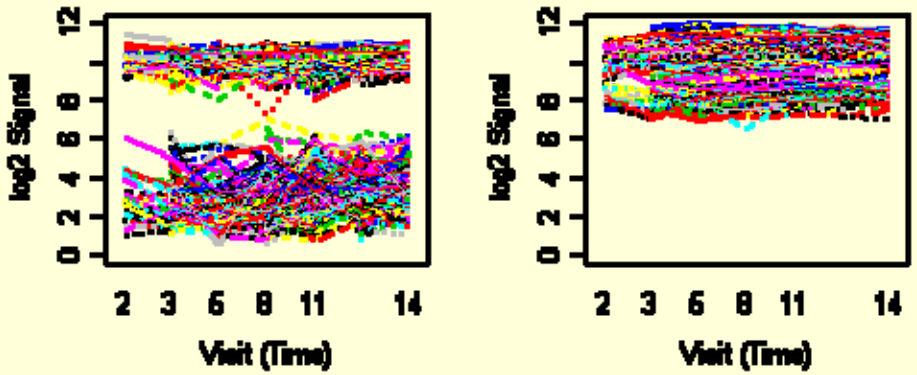
Low CV:

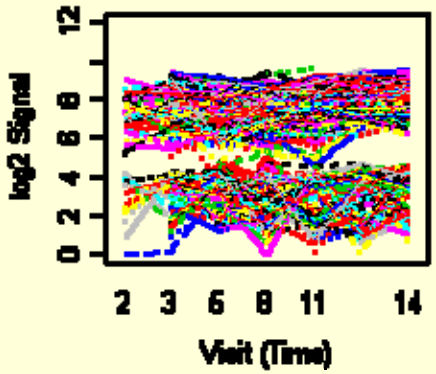


Wyeth Research

Genes with High Intra-Class Correlation

Probe sets with high ICC:





$$\text{intra}-\text{class correlation} = \frac{\text{inter}-\text{subject variance}}{(\text{inter}-\text{subject variance})+(\text{intra}-\text{subject variance})}$$

Wyeth Research

Analysis: Conclusions

- **Wyeth's Healthy Volunteers Observational study has characterized intra- and inter-subject variability in gene expression in large set of healthy volunteers**
- **This study is currently being utilized in Wyeth's mRNA biomarker efforts.**
  - E.g., Burczynski et. al. "Molecular classification of Crohn's disease and ulcerative colitis patients by global expression analysis of peripheral blood mononuclear cells". To appear in Journal of Molecular Diagnostics.

Acknowledgements

This presentation describes joint work with

- **Michael Burczynski, Ph.D.**
- **Andrew Dorner, Ph.D.**
- **Fred Immermann, M.S.**
- **Judy Oestreicher, M.S.**
- **John Ryan, M.D., Ph.D.**

Wyeth Research