

Characterization of Chronic Fatigue Syndrome Using Affective Disorder and Immune System Pathways

Earl F. Glynn¹

Chris Seidel²

Arcady R. Mushegian^{1,3}

Hua Li¹

Frank Emmert-Streib¹

Jie Chen⁴

¹Stowers Institute for Medical Research, Bioinformatics, Kansas City, MO

²Stowers Institute for Medical Research, Microarray Group

³University of Kansas Medical Center, Kansas City, KS

⁴University of Missouri – Kansas City, Dept. of Mathematics and Statistics

<http://research.stowers-institute.org/efg/2006/CAMDA>

Critical Assessment of Microarray Data Analysis Conference

Duke University

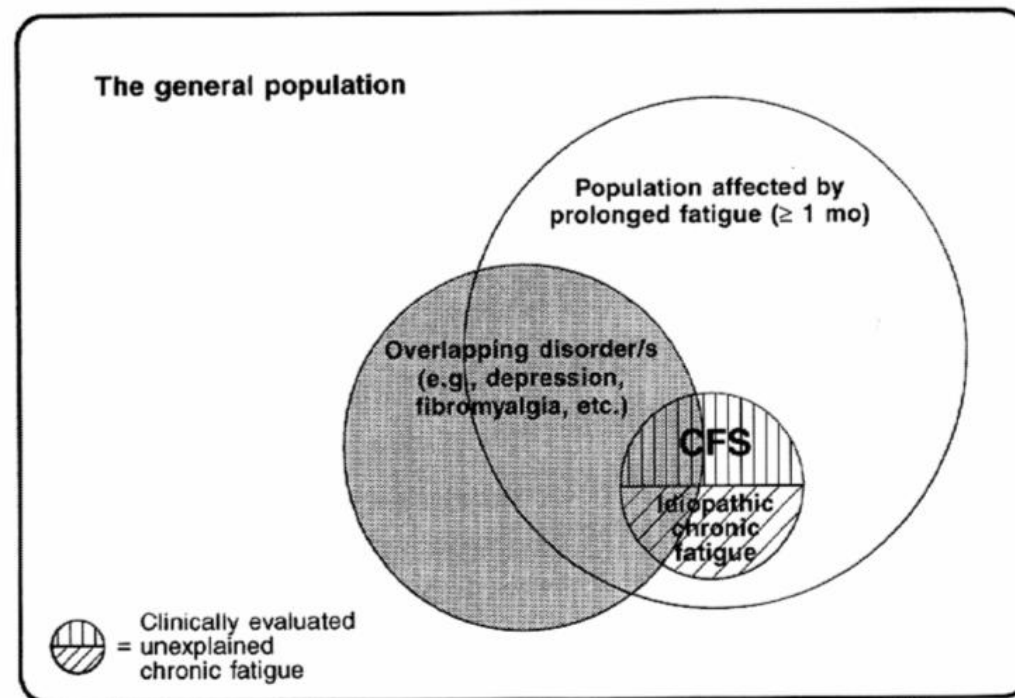
8-9 June 2006

Characterization of Chronic Fatigue Syndrome Using Affective Disorder and Immune System Pathways

- Chronic Fatigue Syndrome Overview
- Data Sources
- SNP Analysis and Results
- Microarray Analysis and Results
- Conclusions

Chronic Fatigue Syndrome Overview

A conceptual framework of abnormally fatigued populations, including those with the chronic fatigue syndrome (CFS) and overlapping disorders



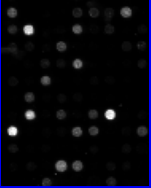

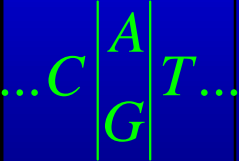


Fukuda, K. et. al. Ann Intern Med 1994;121:953-959

Annals of Internal Medicine

Wichita Chronic Fatigue Syndrome Study

Data Sources

	Clinical Survey Data
	Blood Data
	Gene Expression Data
	Proteomics Data (ignore)
	SNP Data (Single Nucleotide Polymorphism)

How to integrate clinical, blood, microarray and SNP data in analysis?



Clinical Survey Data

BMC Medicine

Published: 15 December 2005



Research article

Chronic Fatigue Syndrome – A clinically empirical approach to its definition and study

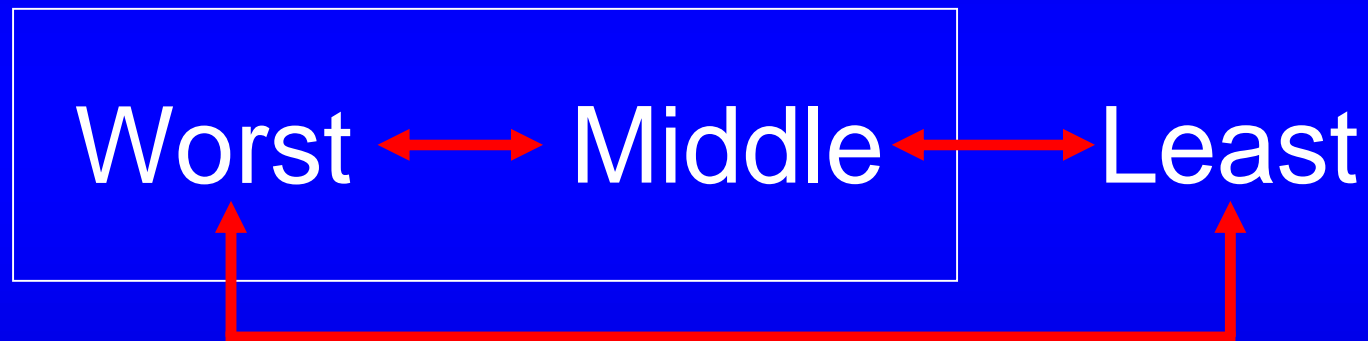
William C Reeves*¹, Dieter Wagner¹, Rosane Nisenbaum^{1,2}, James F Jones¹, Brian Gurbaxani¹, Laura Solomon^{1,3}, Dimitris A Papanicolaou^{4,5}, Elizabeth R Unger¹, Suzanne D Vernon¹ and Christine Heim⁶

Cluster	Frequency	Description
Worst	30	Most severely ill: “lowest SF-36; highest MFI ...”
Middle	67	Intermediate CFS
Least	67	Least severely ill: “scores essentially reflected population norms.”



Disease Cluster Comparisons

Sick



Apply same survey cluster comparisons to Blood, Gene Expression and SNP data.

...C|A|T...
|G|

SNP Data

POMC / TH / MAOA / MAOB / TPH2 / COMT / NR3C1 / SLC6A4 / CRHR1 / CRHR2



...C|A|T...
|G|

SNP/Gene Expression

POMC / TH / MAOA / MAOB / TPH2 / COMT / NR3C1 / SLC6A4 / CRHR1 / CRHR2

REVIEW ARTICLE

Molecular Psychiatry (2005), 1-22

online publication, 31 May 2005;

Genetic tests of biologic systems in affective disorders

E Hattori^{1,2}, C Liu¹, H Zhu¹ and ES Gershon¹

¹Department of Psychiatry, The University of Chicago, Chicago, IL, USA; ²Laboratory for Molecular Psychiatry, RIKEN Brain Science Institute (BSI), Wako, Saitama, Japan

Group	Description	CAMDA SNP Genes	Hattori Genes	CAMDA Microarray Genes
1	Neurotransmission systems	6	129	119
2	Neuroendocrine system	4	20	20
3	Neurotrophic/growth factors, and intracellular signaling in 1 & 2	-	45	42
4	Circadian rhythm	-	30	26
5	Major affective disorders	-	33	30

Microarray Probes Matched to Genes Using biomaRt Bioconductor package

...C|A|T...
|G|

Related “Psych” Genes

Molecular Brain Research 129 (2004) 193–197

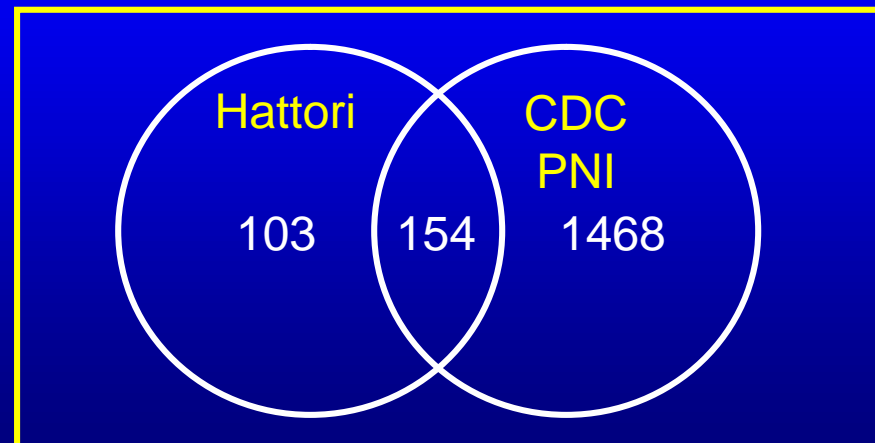
Exploration of neuroendocrine and immune gene expression in peripheral blood mononuclear cells

Ainsley C. Nicholson, Elizabeth R. Unger, Rajeevan Mangalathu, Hanna Ojaniemi, Suzanne D. Vernon*

- CDC Psycho-Neuroendocrine-Immune (PNI) Database
- 1058 genes detected in peripheral blood

Endocrine	323
Immune	618
Neuronal	263
Other	418
Total	1622

1725 genes



Research Question

- **Assume:** Cluster classifications by Reeves, et al, based on clinical data are “correct” disease state assignments.

Hattori's Affective Disorder “psych” genes and/or genes in CDC's Psycho-Neuroendocrine-Immune Systems may be involved in chronic fatigue.

- **Question:** Can affective disorder/immune system genes in objective microarray gene expression and SNP data characterize chronic fatigue patients as well as or better than subjective clinical assessment surveys?

Can microarray and SNP data indicate CFS?

...C|A|T...
G

SNP Analysis

- Hardy-Weinberg Equilibrium
- Bagged Logic Regression
 - Logic Regression
 - Bootstrap Aggregating Logic Regression

...C|A|T...
|G|

SNP Analysis

Hardy-Weinberg Equilibrium

- Let p = frequency of one of two alleles
 q = frequency of other allele

$$p + q = 1$$

Hardy-Weinberg Equilibrium expects genotype frequencies:

$$p^2 + 2pq + q^2 = 1$$

*Bioconductor package, genetics, computes Hardy-Weinberg Equilibrium stats:
HWE.chisq or HWE.exact*

...C|A|T...
G

SNP Analysis

Hardy-Weinberg Equilibrium

- X Chromosome SNPs fail (MAOA and MAOB)
- Other genes are consistent for the “Least” CFS category, except for SLC6A4 SNPs, which weakly fail.
- Certain CRHR1 and NR3C1 SNPs fail for “All” and “Sick” categories but not the “Least” category.

...C | A | T...
 | G |

SNP Analysis

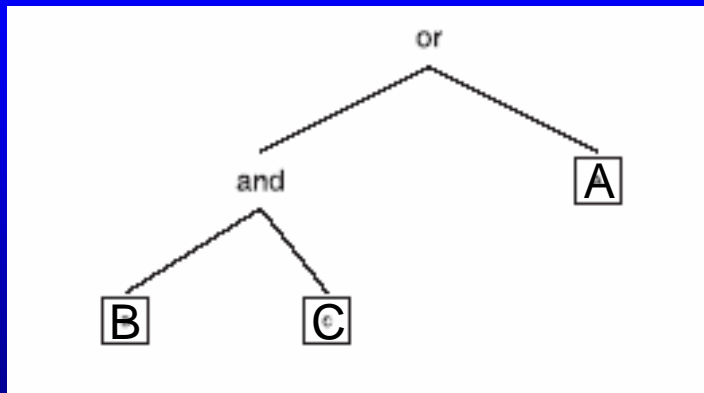
Logic Regression

General Logic Regression Model:

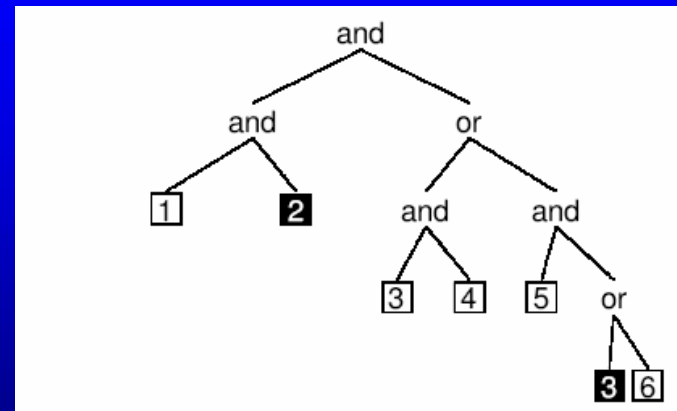
$$Y = \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \dots$$

where L_1 and L_2 are Boolean (0=False, 1=True) expressions which can be represented by logic trees.

$$L = (B \cup C) \cup A$$



A more complicated logic tree



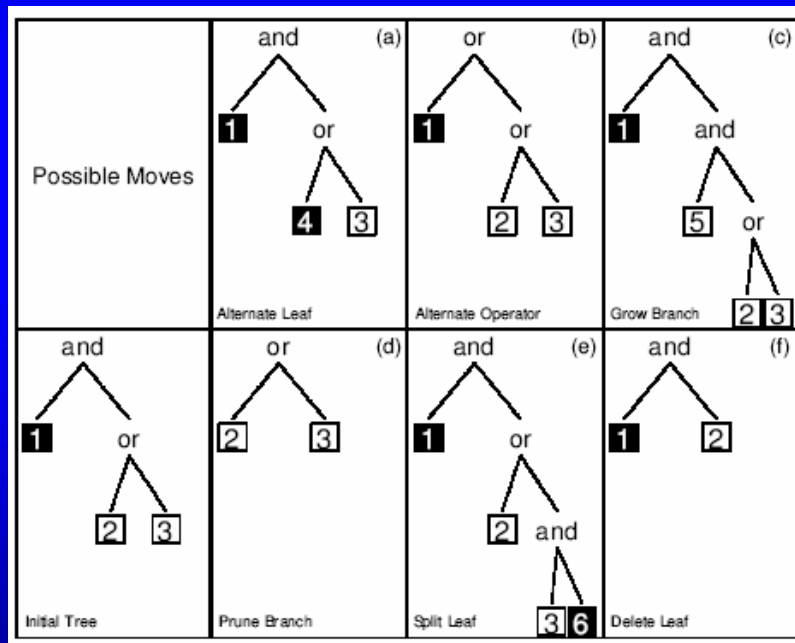
From Ruczinski, et al, (2003), Logic Regression,
Journal of Computational and Graphical Statistics, 12(3), 475-511

...C | A | T...
 | G |

SNP Analysis

Logic Regression

- Ruczinski et al (2003) provide *LogicReg* “R” package
- Uses simulated annealing algorithm to search high-dimensional space, with well-defined move set:



From Ruczinski, et al, (2003),
Logic Regression,
*Journal of Computational and
Graphical Statistics*, 12(3), 475-511.

- Proposed move accepted or rejected based on “score” and “temperature”

...C|A|T...
|G|

SNP Analysis

Logic Regression

Recode SNP information as Boolean data:

SNP	Dominant Model	Recessive Model	Genotype
	SNP_1	SNP_2	
1 AA	0	0	Homozygous Reference "Allele 1"
2 AT	1	0	Heterozygous "Both"
3 TT	1	1	Homozygous Variant "Allele 2"

NOT SNP_1 is written as !SNP_1 in logicFS Bioconductor package

...C | A | T...
 | G |

SNP Analysis

Logic Regression

Holger Schwender (2006) published *logicFS* Bioconductor package, which uses Ruczinski's *LogicRec* package

ID	Classifier	SNP1	SNP2	SNP3
1	1	2	2	1
2	1	3	2	1
3	0	3	3	3
4	1	3	1	1
5	0	2	1	3

Convert Patient
SNP data
to Boolean format



ID	Classifier	SNP1_1	SNP1_2	SNP2_1	SNP2_2	SNP3_1	SNP3_2
1	1	1	0	1	0	0	0
2	1	1	1	1	0	0	0
3	0	1	1	1	1	1	1
4	1	1	1	0	0	0	0
5	0	1	0	0	0	1	1

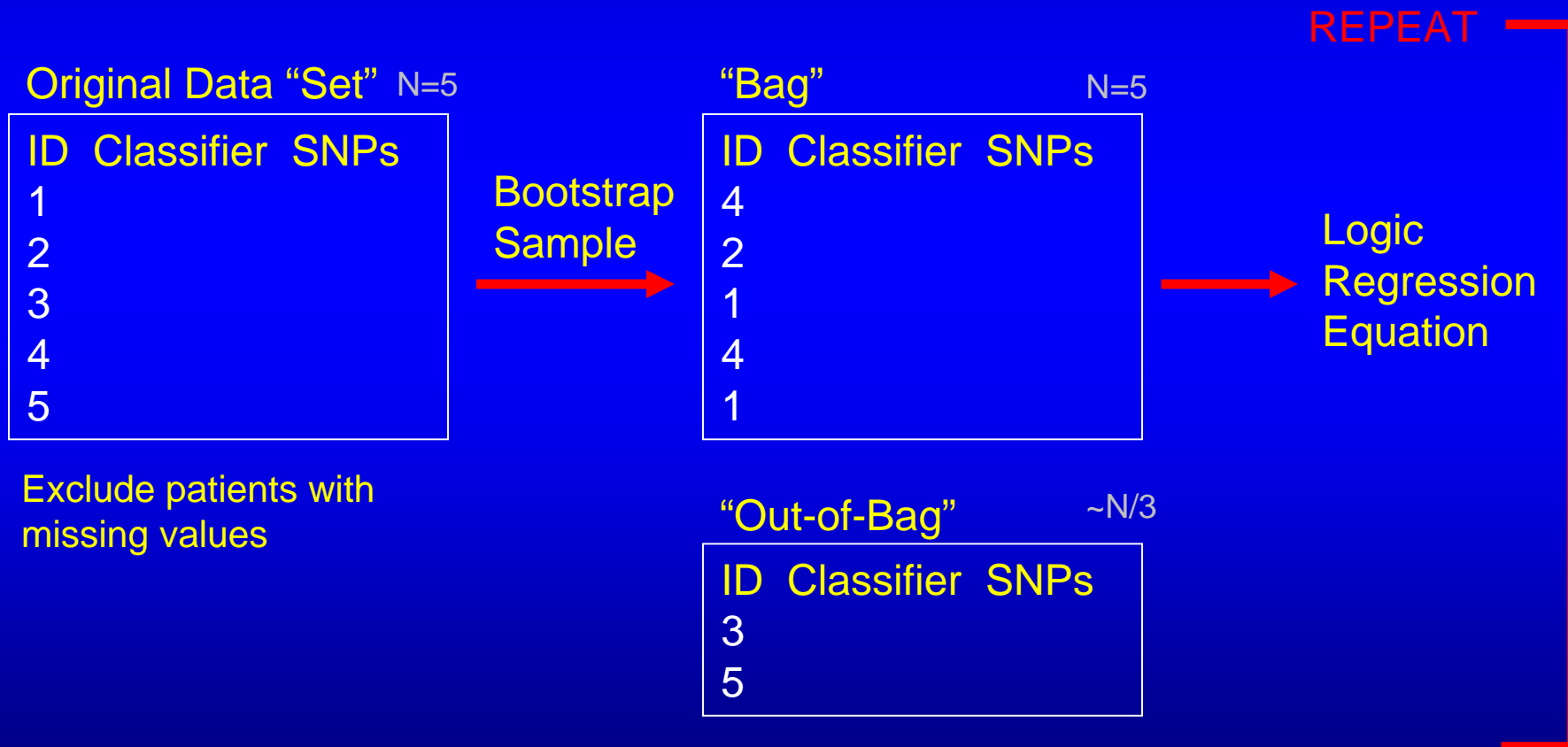
Logic Regression Results: *Classifier = !SNP3_1*
Classifier = !SNP3_2
Classifier = !SNP3_1 & !SNP3_2

...C | A | T...
 | G |

SNP Analysis

Bagged Logic Regression

Schwender's *logicFS* package introduced a bootstrap aggregating, or "bagging," version of logic regression.



Out-of-Bag (OOB) Error Rate Estimated from Regression Equation and OOB Set

...C | A | T...
 | G |

SNP Analysis

Bagged Logic Regression

Schwender's *logicFS* package uses the Quine-McCluskey algorithm to reduce logic regression equations to a minimum disjunctive (OR) normal form.

Regression Equations:

$$Y_{\text{Bag1}} = L_1 \dot{\cup} L_2 \dot{\cup} L_3$$

$$Y_{\text{Bag2}} = L_3$$

$$Y_{\text{Bag3}} = L_1 \dot{\cup} L_3$$

...

where each L is a conjunction (AND) of one or more variables,
e.g., $L_1 = X_1 \dot{\cup} X_3$, $L_2 = X_1 \dot{\cup} X_2 \dot{\cup} X_3$, $L_3 = X_2$

...C|A|T...
G

SNP Analysis

Bagged Logic Regression

Schwender's *logicFS* package uses the Quine-McCluskey algorithm to reduce logic regression equations to a minimum disjunctive (OR) normal form.

Regression Equations:

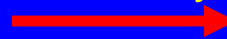
$$Y_{\text{Bag1}} = L_1 \dot{\cup} L_2 \dot{\cup} L_3$$

$$Y_{\text{Bag2}} = L_3$$

$$Y_{\text{Bag3}} = L_1 \dot{\cup} L_3$$

...

Summary



Disjunct	Count
L_3	3
L_1	2
L_2	1

where each L is a conjunction (AND) of one or more variables, e.g., $L_1 = X_1 \dot{\cup} X_3$, $L_2 = X_1 \dot{\cup} X_2 \dot{\cup} X_3$, $L_3 = X_2$

Aggregate results by disjunctive term. Compute proportion and "importance" score.

...C | A | T...
 | G |

SNP Results

Bagged Logic Regression

Out-of-Bag Error Rate Summary [%]			Classifier %
Comparison	Run 1	Run 2	
Worst vs Least	29.7	29.7	W: 35.9
Middle vs Least	55.7	50.0	L: 53.4
Worst vs Middle	40.0	37.1	M: 32.8
Sick vs Least	43.2	42.3	L: 36.9
“Random”	54.0	54.9	50.0

Run 1: 25,000 iterations for simulated annealing, 500 “bags”
Run 2: 50,000 iterations for simulated annealing, 750 “bags”
Exploratory technique for now. “Best” parameters not clear.

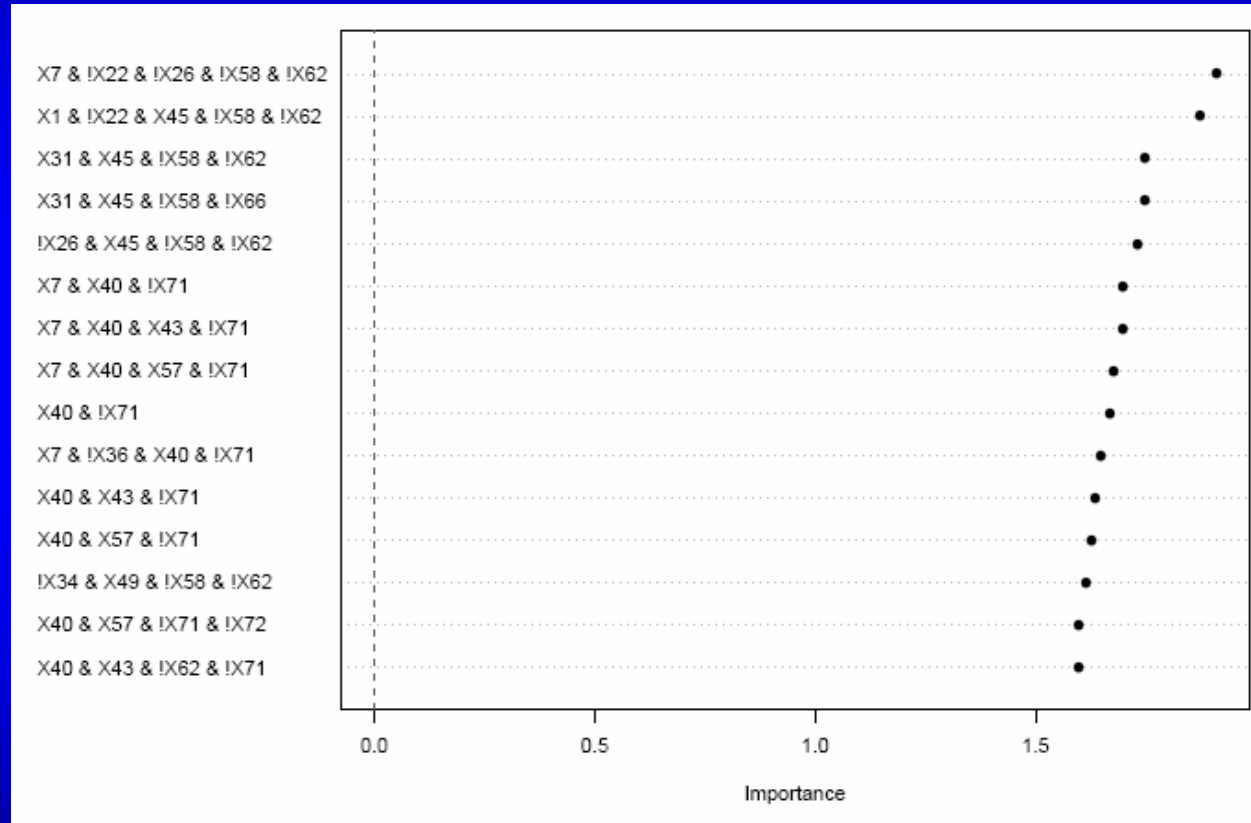
...C|A|T...
G

SNP Results

Bagged Logic Regression

“Importance” of Worst vs Least Disjuncts

1
2
3
4
5



- 1 TPH2.hCV8376042_1 & !TH.hCV243542_2 & !COMT.hCV11804650_2 & !CRHR2.hCV15960586_2 & !NR3C1.hCV11159943_2
- 2 TPH2.hCV15836061_1 & !TH.hCV243542_2 & CRHR1.hCV2544836_1 & !CRHR2.hCV15960586_2 & !NR3C1.hCV11159943_2
- 3 COMT.hCV2538747_1 & CRHR1.hCV2544836_1 & !CRHR2.hCV15960586_2 & !NR3C1.hCV11159943_2
- 4 COMT.hCV2538747_1 & CRHR1.hCV2544836_1 & !CRHR2.hCV15960586_2 & !NR3C1.hCV8950998_2
- 5 !COMT.hCV11804650_2 & CRHR1.hCV2544836_1 & !CRHR2.hCV15960586_2 & !NR3C1.hCV11159943_2

...C|A|T...
|G|

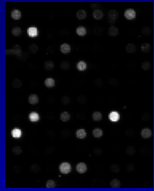
SNP Results

Bagged Logic Regression

Top Two Disjuncts from Run 2: Worst vs Least

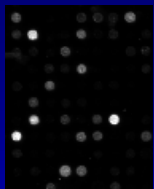
1	TPH2.hCV8376042_1	&	
	!TH.hCV243542_2	&	Single Disjunct matches
	!COMT.hCV11804650_2	&	73% of Worst/Least patients
	!CRHR2.hCV15960586_2	&	
	!NR3C1.hCV11159943_2		
<hr/>			
2	TPH2.hCV15836061_1	&	Single Disjunct matches
	!TH.hCV243542_2	&	75% of Worst/Least patients
	CRHR1.hCV2544836_1	&	0 (64%): 39/41 (95%)
	!CRHR2.hCV15960586_2	&	1 (36%): 9/23 (39%)
	!NR3C1.hCV11159943_2		correct: 48/64 (75%)

Goertzel, et al, Pharmacogenomics (2006), Importance of genes based on SNPs: NR3C1, TPH2, COMT, CRHR2, CRHR1, NRC1, TH, POMC, 5HTT



Microarray Analysis

- Scale/log transform Gene Expression Data
- Apply Kruskal-Wallis Test: Worst, Middle, Least
Reject null hypothesis for $p \leq 0.05$
- Apply Wilcoxon-Mann-Whitney tests:
Worst-Least, Middle-Least, Worst-Middle
- Apply Dunn-Sidák Family-Wise Error Rate
p-value adjustment
- Apply Benjamini & Hochberg multiple test
correction applied separately to each category of
genes
Reject null hypothesis for p-value significant at FDR level of 0.05

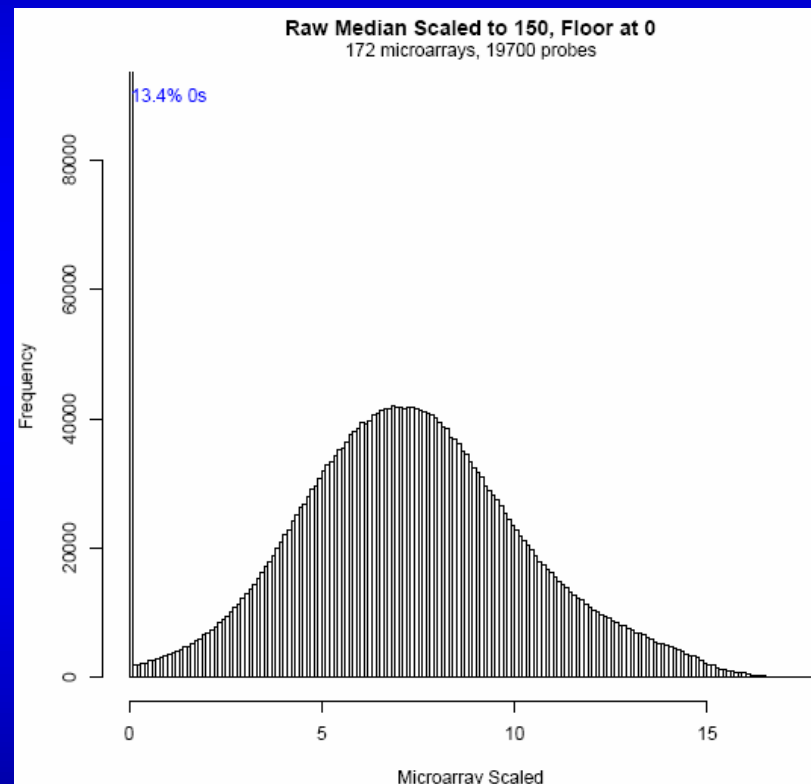
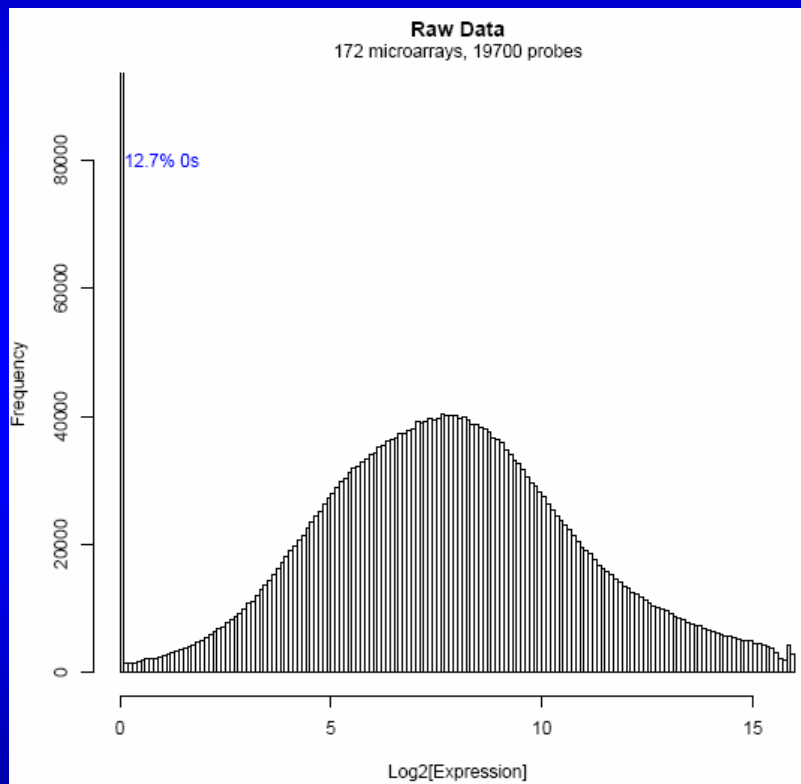


Microarray Analysis

Scaling by Array

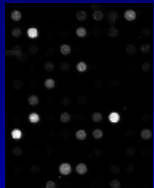
Raw Data: 172 arrays x 19,700 probes

Median of Each Array Scaled to 150



172 patients: 26 Worst, 53 Middle, 44 Least, 49 Excluded

Like in Affymetrix analysis, the array median was scaled to 150. Floor set at 0.

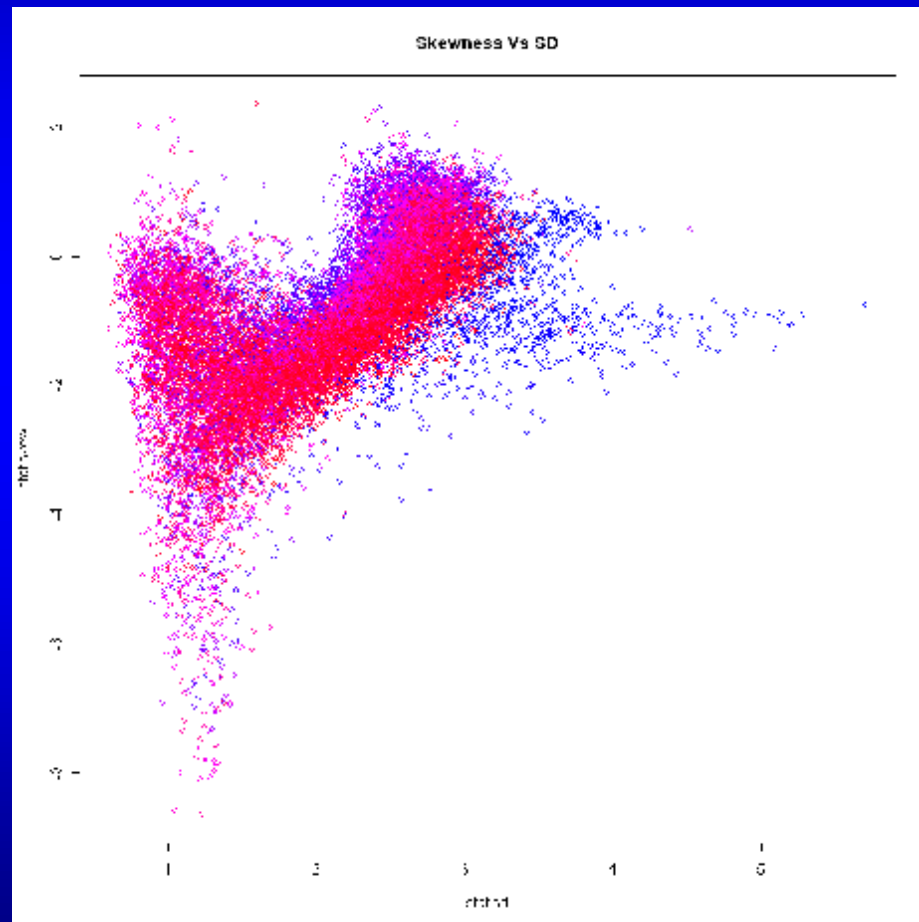
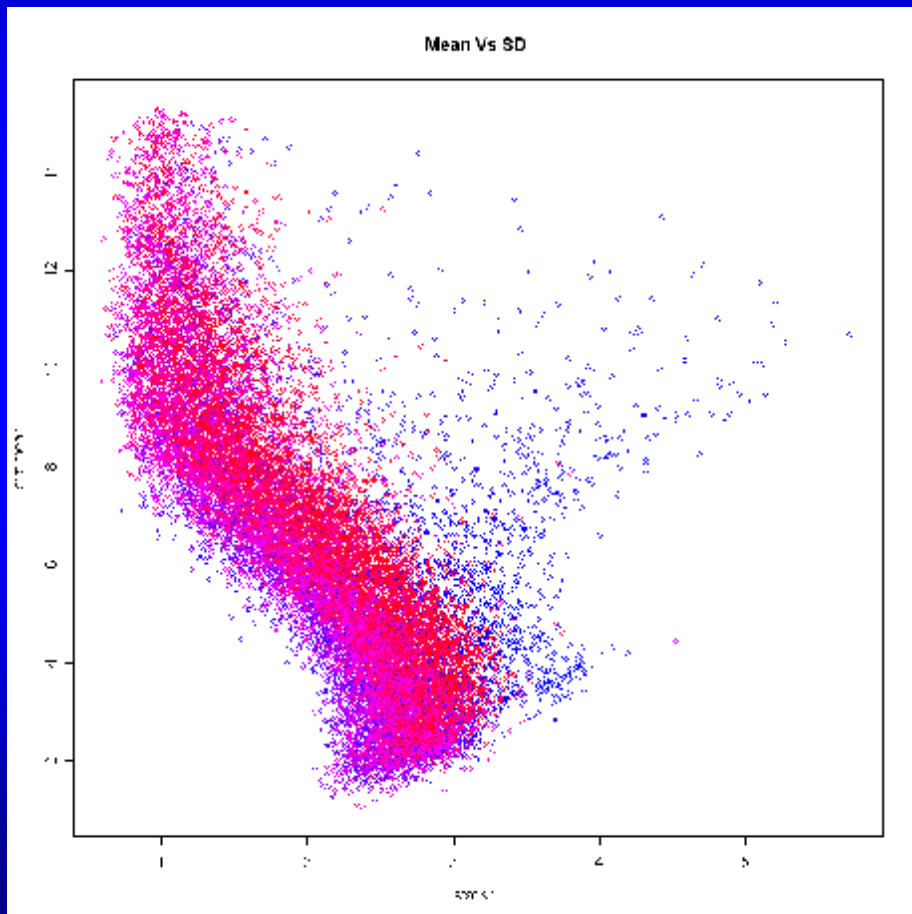


Microarray Analysis

Statistics by Probe

Mean vs Standard Deviation

Skewness vs Standard Deviation

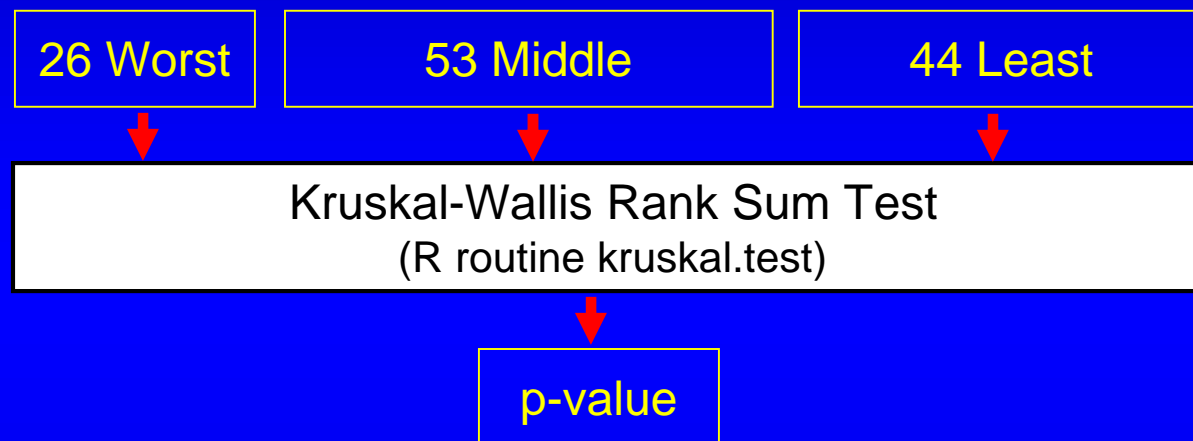


Heteroskedacity: Non-parametric statistics should be used



Microarray Analysis

Kruskal-Wallis Nonparametric ANOVA:
Selects Probes With Differences Among Three Groups



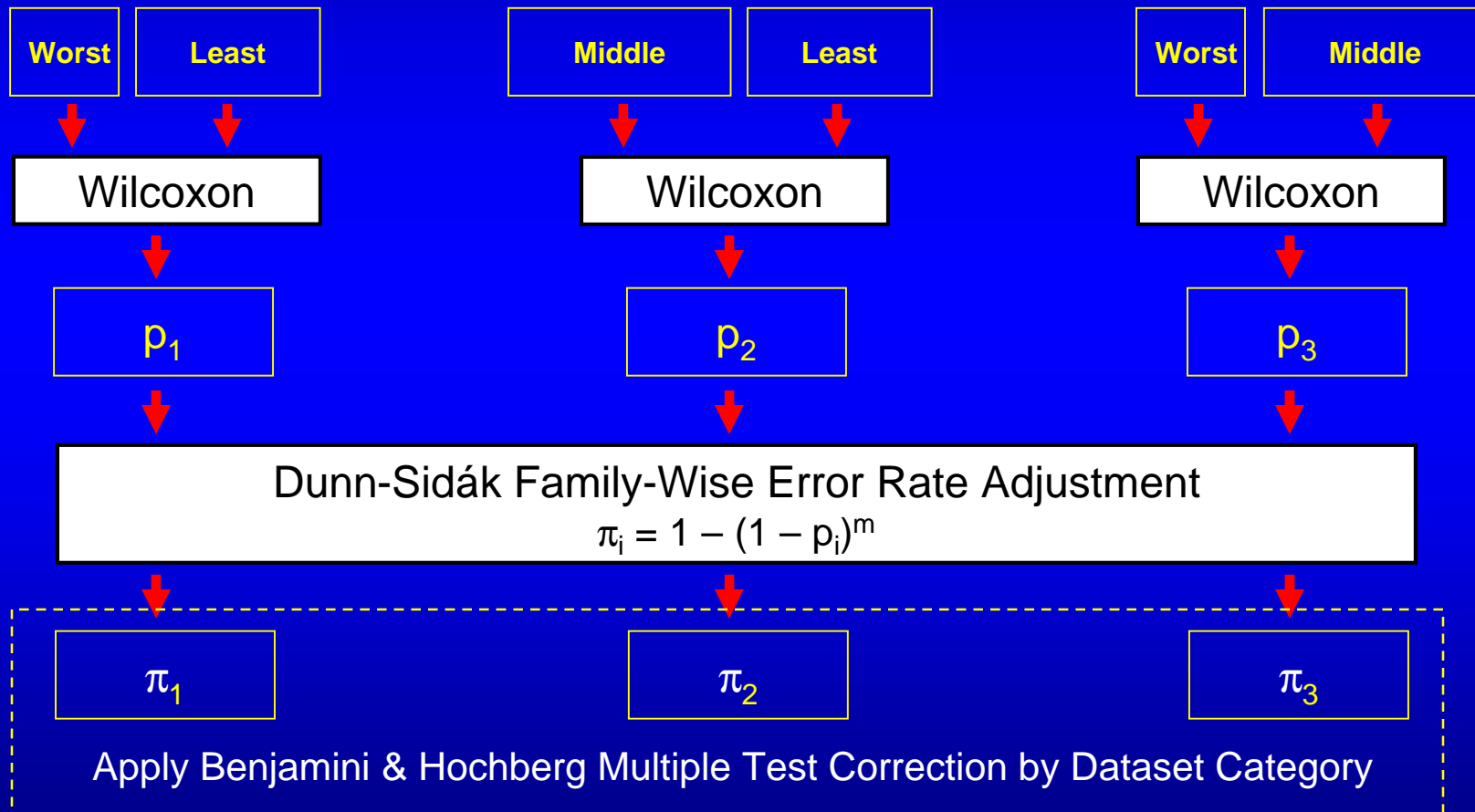
If $p\text{-value} \leq 0.05$ reject null hypothesis that all are in the same group and accept alternate hypothesis there is a difference in at least one comparison.

381 Hattori probes → 13 pass Kruskal-Wallis
1914 CDC PNI probes → 55 pass Kruskal-Wallis

But which pairs of comparisons have significant differences?

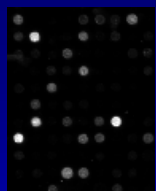
Microarray Analysis

Wilcoxon-Mann-Whitney Nonparametric Test of Two Samples
(R routine `wilcox.test`)



Apply to Each Probe Passing Kruskal-Wallis

Which probes pass all these statistical tests?



Microarray Results: Differential Expression

Hattori Affective Disorder Genes (13 probes of 382: 3.4%)

Group	Worst-Least	Middle-Least	Worst-Middle
1 Neurotransmission	2	4	-
2 Neuroendocrine	-	3	-
3 Intracellular Signaling	-	3	-
4.2 Circadian rhythm	-	1	-
5.2 Schizophrenia	-	1	-
TOTAL	2	12	0

CDC PNI Genes (55 probes of 1914, 2.9%)

System	Worst-Least	Middle-Least	Worst-Middle
Endocrine	6	5	1
Immune	4	12	3
Neuronal	4	4	-
Other	3	22	1
TOTAL	17	43	5

64 probes were identified (4 are in both datasets)



Microarray Results

Differential Expression: 8 Genes: Worst vs. Least (Only)

Dataset	Gene	Group / Description	KEGG Pathway
CDC PNI	EPHB2	Neuronal - Other Neuronal Function / ephrin receptor B2: large erk/cek5 tyrosine kinase; erk	Axon guidance
Hattori	GRIK3	1.3 Neurotransmission System – Amino Acid / ionotropic kainate 3 Glutamate Receptor	Neuroactive ligand-receptor interaction
CDC PNI	IL23R	Immune - Cytokine/Chemokine Receptors / interleukin-23 receptor	Cytokine-cytokine receptor interaction; Jak-STAT signaling
CDC PNI	NR5A2	Endocrine - Hormone Receptor / nuclear receptor subfamily 5, group A, member 2	Maturity onset diabetes of the young
CDC PNI	PMCHL1	Endocrine - Hormones / pro-melanin-concentrating hormone-like 1 protein	Neuroactive ligand-receptor interaction
CDC PNI	RTN4	Other - Other Neuroendocrine Function / brain my043 protein; reticulon 4	-
CDC PNI	SEMA3C	Neuronal - Other Neuronal Function / sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	Axon guidance
CDC PNI	TPO	Endocrine - Hormone Metabolism / thyroid peroxidase isoform 2/3	Cytokine-cytokine receptor interaction; Jak-STAT signaling, Tyrosine metabolism + 6 others

CFS Pathways from *Pharmacogenomics* (2006) 7(3) reported by Fang, et al or Whistler et al
 64 Probes: 8 Worst vs. Least (Only), 5 Worst vs. Middle, 51 Worst or Middle vs. Least



Microarray Results

Differential Expression: 5 Genes: Worst vs. Middle

Dataset	Gene	Group / Description	KEGG Pathway
CDC PNI	C6	Immune - Complement Component / complement component 6 (People with C6 deficiency are prone to bacterial infection.)	Complement and coagulation cascades
CDC PNI	CARD10	Immune - Apoptosis / caspase recruitment domain family, member 10	-
CDC PNI	CISH	Immune - Regulated by Cytokines / cytokine-inducible sh2-containing protein	Jak-STAT signaling pathway
CDC PNI	FURIN	Other - Other / furin preproprotein	Notch Signaling Pathway; Post-translational modification of proteins; TGF-beta signaling pathway
CDC PNI	IDE	Endocrine - Regulates Hormone Activity / insulin-degrading enzyme	Alzheimer's disease

CFS Pathways from *Pharmacogenomics* (2006) 7(3) reported by Whistler et al



Microarray Results

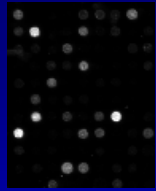
Differential Expression: 50 Genes: Worst-or-Middle vs. Least

ANXA13	DUSP10	IGFBP5	NCOA2	SERPINA6
ATF3	DUSP16	IL18BP	NFKBIL2	SLC1A1
BTK	DUSP22	IL6ST	NPFF	SLC6A7
BTN3A1	EFNA4	INSIG1	NRG1	STAT2
CARD10	EPS15	MAP2K6	NTRK2	TBXAS1
CCL25	FOS	MAPK8IP3	OPRM1	TCF4
CDC37	FYN	MDM2	PDYN / PTPNS1	TLR10
CHGA	GNAS	MR1	PIP5K2A	TNFSF13
CHRM1	HSD11B1	MS4A3	PPARD	TRPM2
CRP	HSPD1	NCOA1	PSMB8	ZNF14

In CFS Pathways from *Pharmacogenomics* (2006) 7(3) reported by Fang, et al

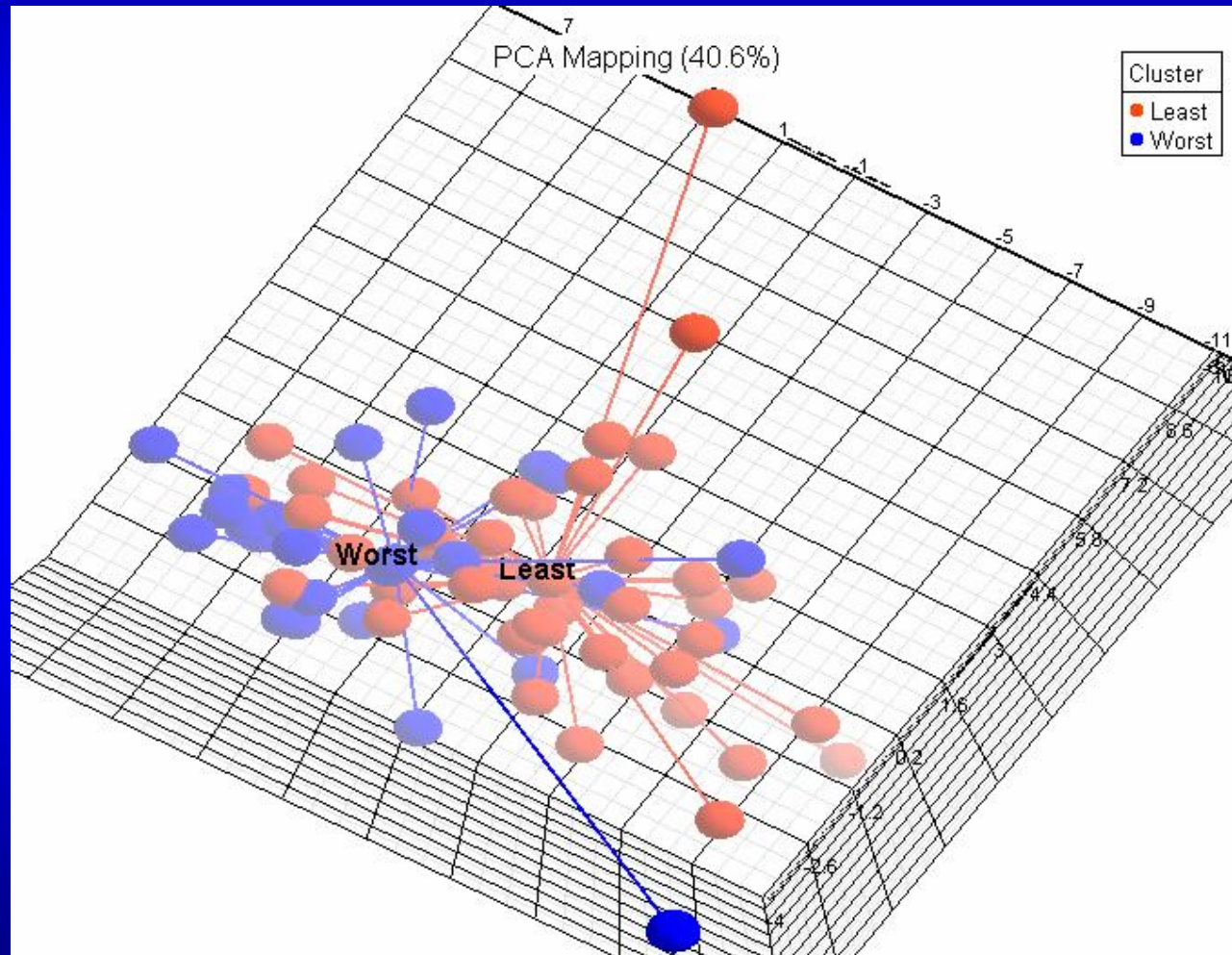
See details in online supplement

<http://research.stowers-institute.org/efg/2006/CAMDA>



Microarray Results

Principal Components Analysis for Genes Passing Kruskal-Wallis Test



Analysis performed with Partek Genomics Suite

Conclusions

- Bootstrap Aggregating (Bagged) Logic Regression is a new technique that may be useful in analyzing SNP associations.
- Bagged Logic Regression identified “Worst-Least” CFS SNP genes consistent with exhaustive search by Goertzel, et al (2006).
- “Interesting” SNPs did not show statistically significant gene expression differences.
- Eight differentially expressed genes distinguish between Worst and Least states; five distinguish between Worst and Middle states.
- Unclear why there were so many more differentially expressed genes (50) between Worst/Middle and Least states.
- Affective Disorder/Immune System Gene Expression and SNP data may be better in disease state classification than subjective clinical data, but further validation is needed.

Acknowledgements



Stowers Institute for Medical Research

Suzanne D. Vernon

Centers for Disease Control and Prevention
Psycho-Neuroendocrine-Immune (PNI) Database

Holger Schwender

University of Dortmund
Help Using *logicFS* Bioconductor Package