약 력

1. 인적사항

	성 명	김 선 영
	소속기관	생명공학연구원
	직 위	선임연구원
	전자메일	kimsy@kribb.re.kr

2. 학력/경력

연 도	학교 / 기관	전공 / 직위	학위 / 비고
1988. 3 - 1992. 2	서울대학교 / 미생물학과	미생물학	학사
1992. 3 - 1994. 2	서울대학교 / 미생물학과	미생물학	석사
1994. 3 - 1998. 8	서울대학교 / 미생물학과	미생물학	박사
1998. 9 - 2001. 6	생명공학연구소/ 세포생물학 실험실	박사 후 연구원	
2001. 7 - 2004.12	Dept. Pathology, Columbia University	박사 후 연구원	
2005. 1 - 현재	생명공학연구원/의약유전체연구센터	선임 연구원	

3. 주요연구실적(개조식, 간단하게)

•***Kim S.Y.** and *Volsky DJ (2005) PAGE: Parametric Analysis of Gene Set Enrichment, *BMC Bioinformatics*, 6:144, *(equal corresponding authors)

•**Kim S.Y.**, Li J, Bentsman G, Brooks AI, Volsky DJ. (2004) Microarray analysis of changes in cellular gene expression induced by productive infection of primary human astrocytes: implications for HAD, *J Neuroimmunol*, 157(1-2):17-26.

•**Kim, S. Y.**, Choi, S. Y., Wei, C., and D. J. Volsky, (2003) Transcriptional regulation of human excitatory amino acid transporter 1 (EAAT1): Cloning of the EAAT1 promoter and characterization of its basal and inducible activity in human astrocyte. *J. Neurochemistry*, 87: 1485-1498

•**Kim,** S. Y., Wei, C., Choi, S.Y., and D. J. Volsky, (2003) Cloning and Characterization of the 3'-untranslated region of the human excitatory amino acid transporter 2 transcript, J. *Neurochemistry*, 86(6): 1458-67

Gene Set and Pathway Analysis of Microarray Data

생명공학연구원 선임연구원 **김 선 영**

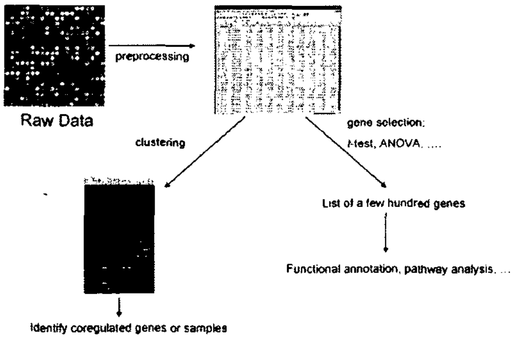
최근의 microarray 기술의 발달로 인해 점점 더 많은 양의 mRNA 발현 데이터가 쌓여 가고 있다. 이제는 데이터를 만드는 단계보다는 데이터로부터 중요한 생물학적 의미를 끌어내는 것이 더욱 중요한 일이 되었다. micorarray 기술이 처음 도입된 이후로, 많은 앨고리즘과 소프트웨어가 개발되어, 실험자들이 microarray 데이터로부터 생물학적 의미를 끌어내는 작업을 도와주어 왔다. 그런데, 이전의 데이터 마이닝 방법들은 거의 예외 없이 전체 데이터로부터 선택된 몇 십, 몇 백 개의 유전자 리스트로부터 출발한다. 그런데, 이러한 방법 (over-representation analysis, ORA로 줄임)은 몇 가지 한계를 가지고 있어서, 최근에는 전체 데이터로부터 의미 있는 유전자 세트 (gene set)를 찾아내는 방법들이 도입되었다. 본 세미나는 이런 방법들, 줄여서 gene set analysis라 함, 에 사용되는 앨고리즘들과 소프트웨어들을 비교, 검토하고자 한다.

Gene Set and Pathway Analysis of Microarray Data

Outline

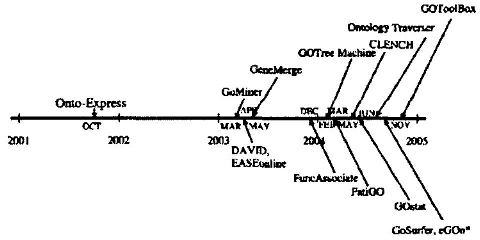
- Existing GO and Pathway Analysis Tools
- Disadvantages of Over-Representation Analysis (ORA)
- Gene Set Analysis or Functional Class Scoring (FCS)
- PAGE (Parametric Analysis of Gene Set Enrichment)
- Comparison of Current Gene Set Analysis Tools

Interpretation of Microarray Data



Raw Data
preprocessing
clustering
gene selection:
t-test, ANOVA, ...
List of a few hundred genes
Functional annotation, pathway analysis, ...
Identify coregulated genes or samples

Evolution history of GO-based functional analysis software



Khatri & Draghici, 2005, Bioinformatics

Review of GO analysis tools: statistical model

Tool	Statistical model	Correction for multiple experiments	GO visualization	Microarrays supported	Time to process 200 genes (s)
Onto-Express	χ², binomial, hypergeometric, Fisher's exact test	Šidák, Holm, Bonferroni, FDR	Flat, Tree	172 commercial arrays (Affymetrix, Super Array, Sigma-Genosys, Clontech, PerkinElmer, Operon, Takara, NIA); can also upload a user-defined list	7, 8, 16, 20
GoMiner	Fisher's exact test	[illegible]	Tree, DAG	Uploads from user	77, 121, 223, 346
DAVID	None	None	Not available	Not applicable	15, 17, 27, 54
EASEonline	Fisher's exact test	[illegible]	Not available	27 arrays (Affymetrix only); can also upload a user-defined list	13, 19, 34, 34
GeneMerge	Hypergeometric	Bonferroni	Flat, no hierarchical structure	Uploads from user	4, 6, 6, 8
FuncAssociate	Fisher's exact test		Not available	Uploads from user	22, 27, 29, 56
GOTM	Hypergeometric	None	Tree	12 arrays (Affymetrix only); uploads from user	59, 60, 152,
FatiGO	Fisher's exact test	Step-down minP, FDR (Benjamini and Hochberg, 1995), FDR (Benjamini and Yekutieli, 2001)	Flat, Tree	Uploads from user	15, 40, 69, 105
CLENCH	Hypergeometric, χ², binomial	None	DAG	Uploads from user	NA
GOstat	χ², Fisher's exact test	FDR, Holm	Not available	Uploads from user	12, 20, 46, 86
GOToolBox	Hypergeometric, binomial, Fisher's exact test	Bonferroni, Holm, Hochberg, Hommel, FDR	Not available	Uploads from user	22, 31, 105, 230
GoSurfer	χ²	q-value	DAG	22 arrays (Affymetrix only); uploads from user	2, 2, 2, 3
Ontology Traverser	Hypergeometric	FDR	Not available	5 arrays (Affymetrix); uploads from user	NA
eGOn	Binomial	None	Tree	Uploads from user	26, 45, 89, 94

Khatri & Draghici, 2005, Bioinformatics

Review of GO analysis tools: user interface

Tool	Scope of the analysis	Level of abstraction	User interface	Application type	Platform	Supported input IDs
Onto-Express	All GO categories	Fully dynamic; different levels of abstractions in different GO subtrees	Java GUI	Web-based	Any	GenBank, UniGene, Entrez Gene, Affymetrix, Gene symbol
GoMiner	All GO categories	Static global analysis	Java GUI	Stand-alone	Windows only	Organism specific IDs in GO
DAVID	All GO categories	Only lowest level of GO	HTML GUI	Web-based	Any	GenBank, UniGene, Entrez Gene, Affymetrix, RefSeq, UniProt, PIR
EASEonline	All GO categories	User selected, fixed level	HTML GUI	Both	Any	Affymetrix, GenBank, UniGene, Entrez Gene
GeneMerge	One category	Only lowest level of GO	HTML GUI	Both	Any	Only supports organism specific IDs used in GO
FuncAssociate	All GO categories	Only lowest level of GO	HTML GUI	Web-based	Any	[illegible] gene products
GOTM	All GO categories	Only lowest level of GO	HTML GUI	Web-based	Any	Affymetrix, UniGene, [illegible], Swiss-Prot, Entrez Gene
FatiGO	One category	User selected, fixed level and static global analysis	HTML GUI	Web-based	Any	Affymetrix, GenBank
CLENCH	All GO categories	Static global analysis	Command-line input, HTML output	Stand-alone	Windows only	A. thaliana MIPS IDs
GOstat	All GO categories	User selected, fixed level	HTML GUI	Web-based	Any	GenBank, UniGene, Gene symbol, Organism specific IDs in GO
GOToolBox	All GO categories	User selected, fixed level	HTML GUI	Web-based	Any	Only organism specific IDs in GO
GoSurfer	All GO categories	Only lowest level of GO	C/C++ GUI	Stand-alone	Windows only	Affymetrix, UniGene, Entrez Gene
Ontology Traverser	One category	Only lowest level of GO	HTML GUI	Web-based	Any	Affymetrix
eGOn	One category	Only lowest level of GO	HTML GUI	Web-based	Any	GenBank, UniGene, Clone

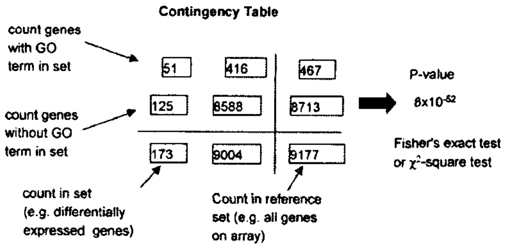
Khatri & Draghici, 2005, Bioinformatics

Existing pathway analysis software

Software	Method	Annotations	Platform	URL	Comments	Refs
CLENCH	Hypergeometric probability, χ^2, binomial distribution	GO	Install (perl)	http://www.personal.psu.edu/[illegible]/Clench/	Species specific, Arabidopsis	[38]
FatiGO (GEPAS)	Fisher Exact test, various multiple testing correction, permutation	GO	Web	http://fatigo.bioinfo.cnio.es/	Part of GEPAS, several organisms	[40, 41]
FunSpec	Hypergeometric distribution, multiple testing correction	GO, [illegible]	Web	http://funspec.med.utoronto.ca/	Yeast only, last updated 2002	[42]
[illegible]	Visual, overlap probability	GO, KEGG, GenMAPP	Install (java)	www.[illegible].com		
GeneXPress	Hypergeometric distribution	GO, KEGG, GenMAPP	Install (java)	http://GeneXPress.stanford.edu	Currently yeast only	[39]
GO-cluster	Pearson's correlation coefficient	GO	Install (Windows)	http://www.[illegible]/go-cluster/		[43]
[illegible]	Visual, search tool	GO	Web (java)	http://[illegible].edu/goproject		[44]
GoMiner	Fisher Exact test, Relative enrichment R_e	GO	Install (java)	http://discover.nci.nih.gov/gominer/		[45]
GOstat	Fisher Exact test, multiple testing correction	GO	Web	http://gostat.wehi.edu.au		[46]
[illegible]	Pearson's χ^2, q-value	GO	Install (Windows)	www.[illegible].harvard.edu/[illegible]/		[47]
MAPPFinder	Standardized difference score (z) from hypergeometric distribution	GO	Install (Windows)	www.genmapp.org	Yeast, mouse and human	[18]
NetAffx GO mining tool	Directed acyclic graph, χ^2, links to AmiGO	GO	Web (log in)	www.affymetrix.com/analysis/[illegible]	Affymetrix data only	[48]
Onto-Express	χ^2, hypergeometric distribution, binomial distribution, multiple testing	GO	Web (log in)	http://vortex.cs.wayne.edu/Projects.html	Part of Onto-Tools	[50]
PathMAPA	Fisher Exact test	[illegible]	Web	http://bioinformatics.[illegible]	Arabidopsis only	[49]
Pathway Miner	Fisher Exact test	Biocarta, GenMAPP, KEGG	Web	http://www.biorag.org/pathway.html		[6]
Pathway Processor	Fisher Exact test	KEGG	Install (java)	http://cgr.harvard.edu/[illegible].html	Yeast only but adaptable	[51]

Curtis et al., 2005. Trends Biotech.

ORA (Over-Representation Analysis)



Contingency Table
count genes with GO term in set
count genes without GO term in set
51
416
467
125
8588
8713
173
9004
9177
count in set (e.g. differentially expressed genes)
Count in reference set (e.g. all genes on array)
P-value
8x10^-52
Fisher's exact test or χ²-square test

Problems of Over-Representation Analysis (ORA)

•Arbitrary Cut-off (or threshold)

•Actual values are ignored after gene selection

•Many small changed, but important genes are ignored.

Problem 1. Arbitrary Cut-off

Hypothetical example   Cut-off: $|\log2(FC)| > 1$, p-value < 0.05

Gene	FC (log2)	P-value (*t*-test)	Selected?
Gene1	3	0.001	Yes
Gene2	0.02	0.98	No
Gene3	1.01	0.049	Yes
Gene4	0.99	0.002	No
Gene5	1.01	0.051	No
Gene6	4.2	0.051	No

Problem 2: Actual values are ignored after selection

Gene	FC (log2)	P-value (*t*-test)
Gene1	3	0.001
Gene2	1.1	0.02
Gene3	1.01	0.049
Gene4	3.99	0.00002
Gene5	1.01	0.01
Gene6	4.2	0.00001

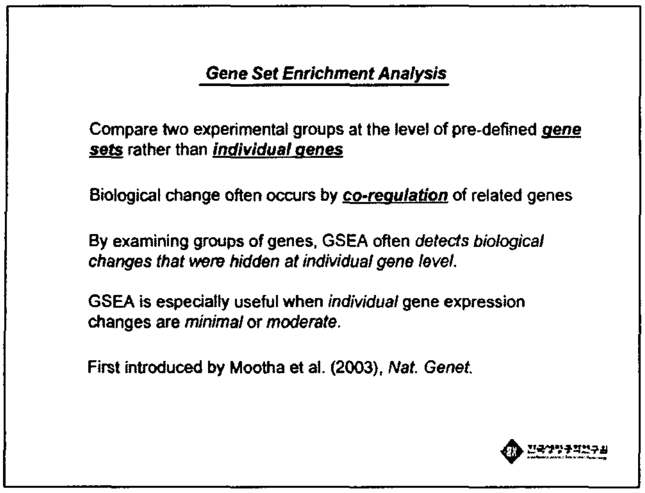
↓

Gene1, Gene2, Gene3, Gene4, Gene5, Gene6... AS USER INPUT

Problem 3: If No or Few genes pass selection?

No genes to input into statistical test.

Data mining becomes impossible.



Gene Set Enrichment Analysis

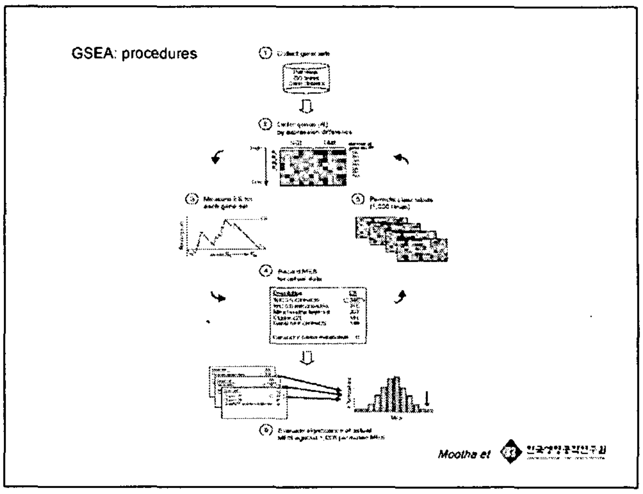
Compare two experimental groups at the level of pre-defined ***gene sets*** rather than ***individual genes***

Biological change often occurs by ***co-regulation*** of related genes

By examining groups of genes, GSEA often *detects biological changes that were hidden at individual gene level.*

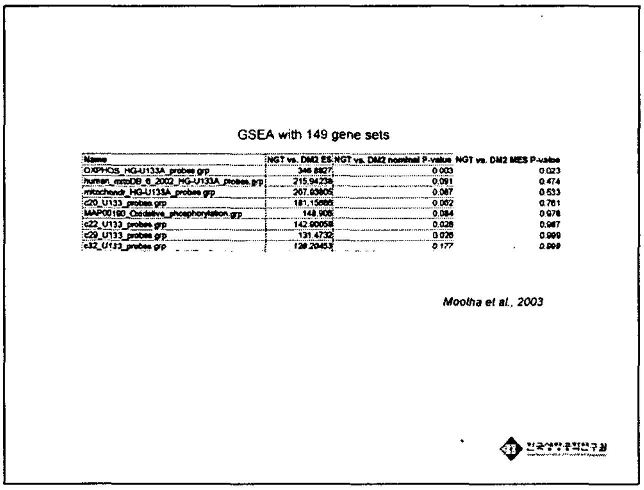
GSEA is especially useful when *individual* gene expression changes are *minimal* or *moderate.*

First introduced by Mootha et al. (2003), *Nat. Genet.*



GSEA: procedures

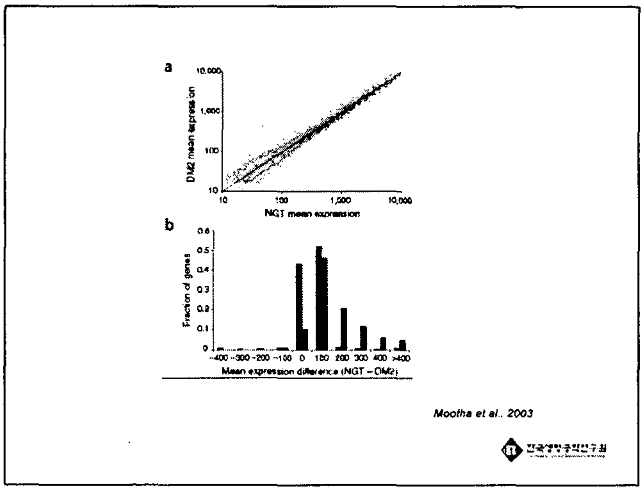
Mootha et



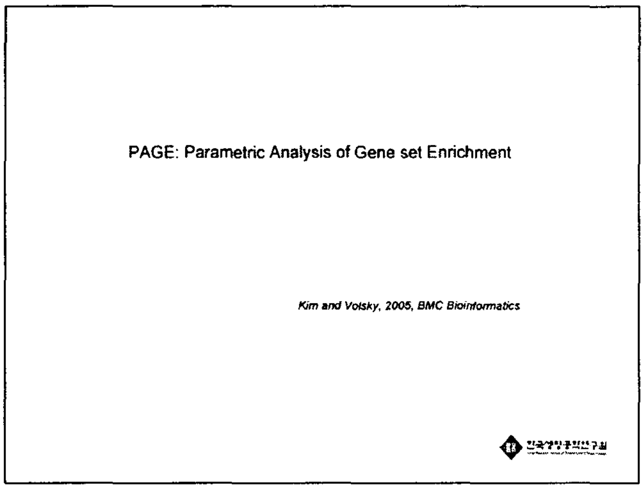
GSEA with 149 gene sets

Name	NGT vs. DM2 ES	NGT vs. DM2 nominal P-value	NGT vs. DM2 MES P-value
OXPHOS_HG-U133A_probes.grp	346.8827	0.003	0.023
human_mitoDB_6_2002_HG-U133A_probes.grp	215.94238	0.091	0.474
mitochondr_HG-U133A_probes.grp	207.93805	0.087	0.533
c20_U133_probes.grp	181.15686	0.062	0.761
MAP00190_Oxidative_phosphorylation.grp	148.906	0.084	0.976
c22_U133_probes.grp	142.90058	0.028	0.987
c29_U133_probes.grp	131.4732	0.026	0.999
c32_U133_probes.grp	128.20453	0.177	0.999

Mootha et al., 2003

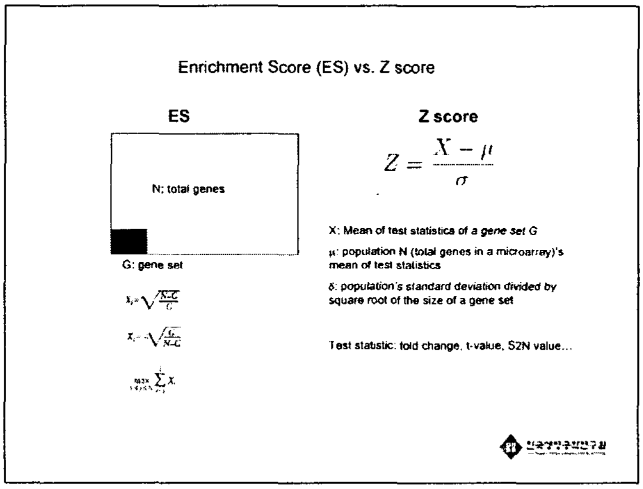


Mootha et al., 2003



PAGE: Parametric Analysis of Gene set Enrichment

Kim and Volsky, 2005, BMC Bioinformatics



Enrichment Score (ES) vs. Z score

ES

N: total genes

G: gene set

$X_i = \sqrt{\frac{N-G}{G}}$

$X_i = -\sqrt{\frac{G}{N-G}}$

$\max_{1 \le j \le N} \sum_{i=1}^{j} X_i$

Z score

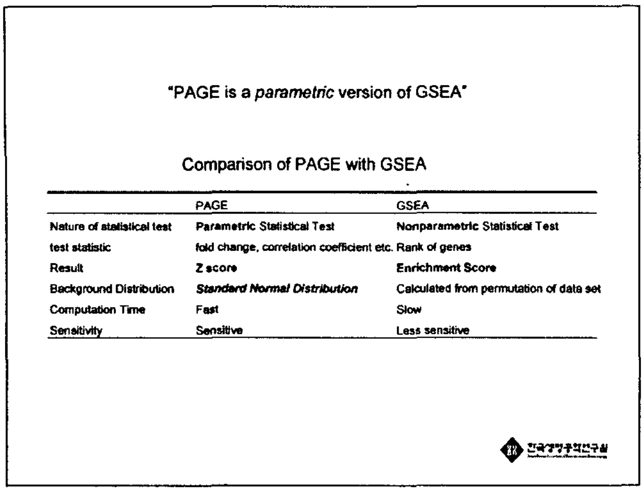
$$Z = \frac{X - \mu}{\sigma}$$

X: Mean of test statistics of a *gene set* G

μ: population N (total genes in a microarray)'s mean of test statistics

δ: population's standard deviation divided by square root of the size of a gene set

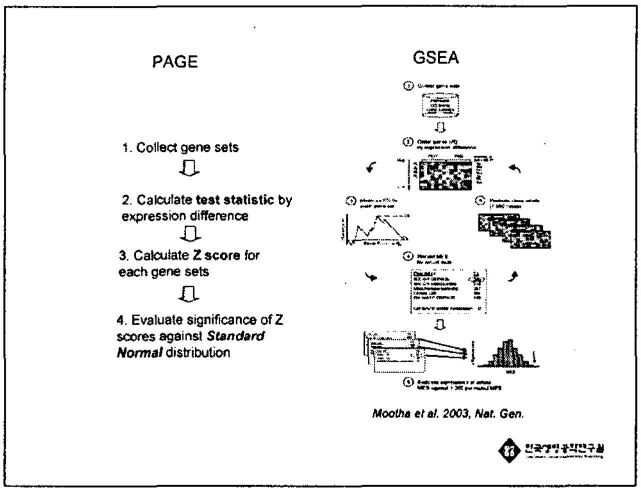
Test statistic: fold change, t-value, S2N value...



"PAGE is a *parametric* version of GSEA"

Comparison of PAGE with GSEA

	PAGE	GSEA
Nature of statistical test	Parametric Statistical Test	Nonparametric Statistical Test
test statistic	fold change, correlation coefficient etc.	Rank of genes
Result	Z score	Enrichment Score
Background Distribution	***Standard Normal Distribution***	Calculated from permutation of data set
Computation Time	Fast	Slow
Sensitivity	Sensitive	Less sensitive

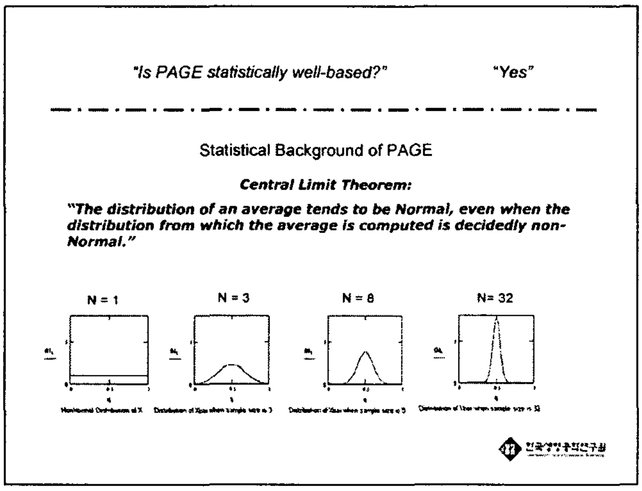


PAGE

1. Collect gene sets
2. Calculate test statistic by expression difference
3. Calculate Z score for each gene sets
4. Evaluate significance of Z scores against ***Standard Normal*** distribution

GSEA

Mootha et al. 2003, Nat. Gen.

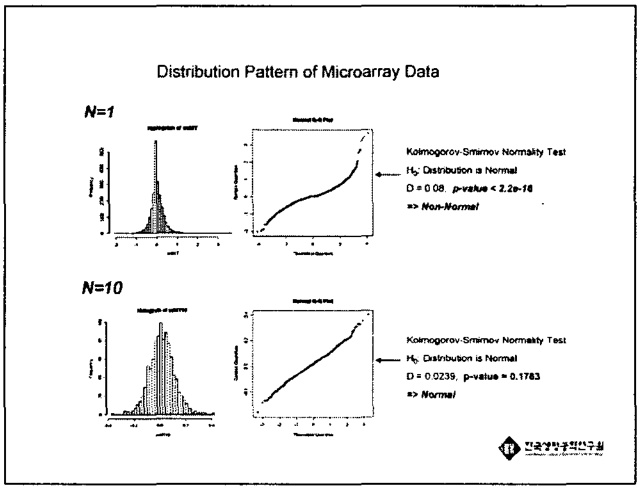


"Is PAGE statistically well-based?" *"Yes"*

Statistical Background of PAGE

Central Limit Theorem:

"The distribution of an average tends to be Normal, even when the distribution from which the average is computed is decidedly non-Normal."



Distribution Pattern of Microarray Data

N=1

Kolmogorov-Smirnov Normality Test

H_0: Distribution is Normal

D = 0.08, ***p-value < 2.2e-16***

=> ***Non-Normal***

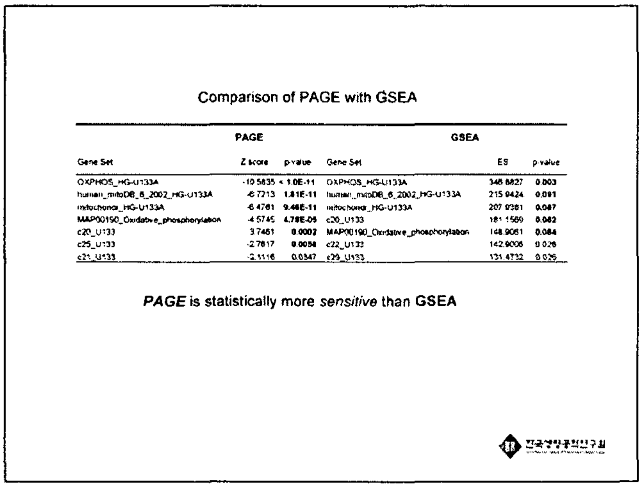
N=10

Kolmogorov-Smirnov Normality Test

H_0: Distribution is Normal

D = 0.0239, **p-value = 0.1783**

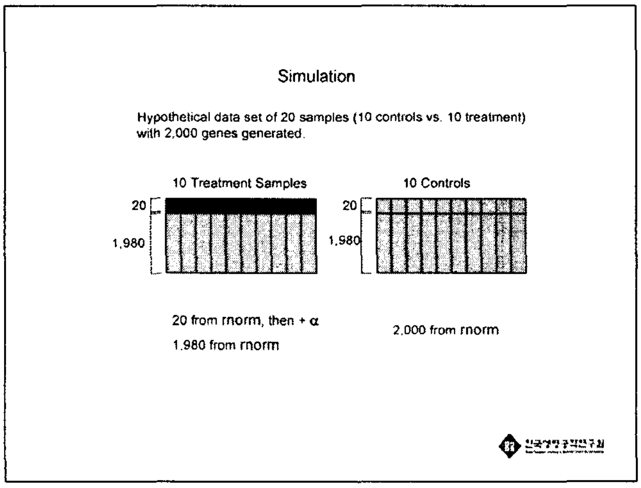
=> ***Normal***



Comparison of PAGE with GSEA

PAGE			GSEA		
Gene Set	Z score	p-value	Gene Set	ES	p-value
OXPHOS_HG-U133A	-10.5835	< 1.0E-11	OXPHOS_HG-U133A	348.8827	0.003
human_mitoDB_6_2002_HG-U133A	-6.7213	1.81E-11	human_mitoDB_6_2002_HG-U133A	215.9424	0.091
mitochondr_HG-U133A	-6.4761	9.46E-11	mitochondr_HG-U133A	207.9381	0.087
MAP00190_Oxidative_phosphorylation	-4.5745	4.78E-06	c20_U133	181.1569	0.082
c20_U133	3.7461	0.0002	MAP00190_Oxidative_phosphorylation	148.9061	0.084
c25_U133	-2.7617	0.0058	c22_U133	142.9006	0.026
c21_U133	-2.1116	0.0347	c29_U133	131.4732	0.026

PAGE is statistically more *sensitive* than GSEA



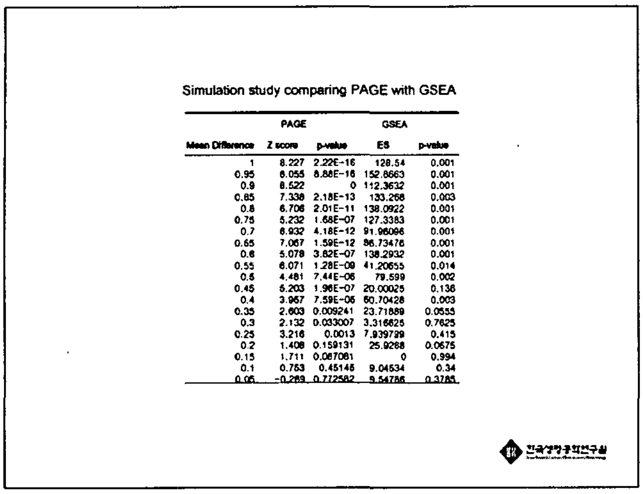
Simulation

Hypothetical data set of 20 samples (10 controls vs. 10 treatment) with 2,000 genes generated.

20 from rnorm, then + α

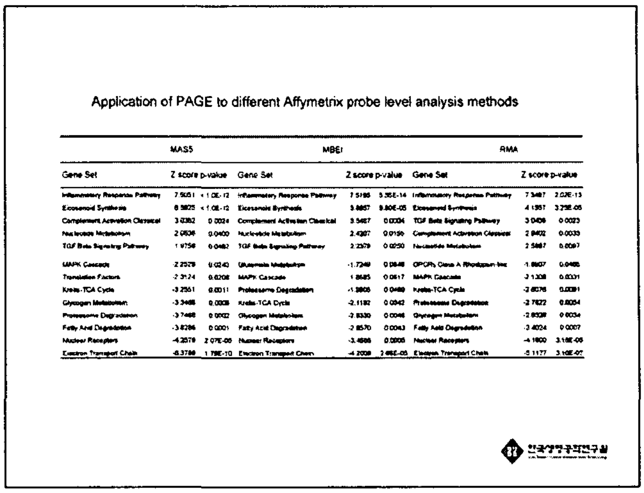
1,980 from rnorm

2,000 from rnorm



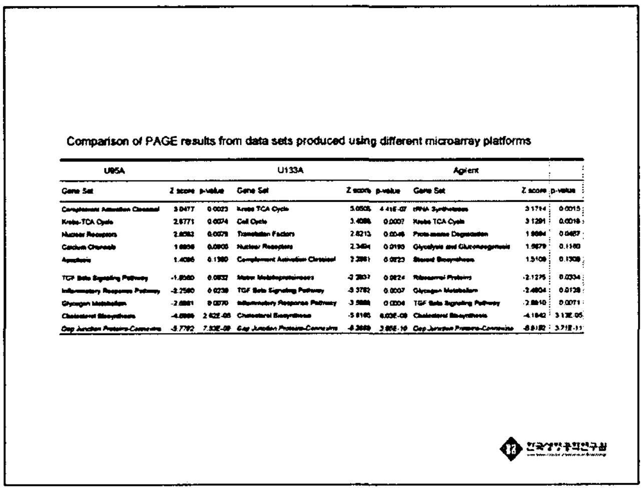
Simulation study comparing PAGE with GSEA

	PAGE		GSEA	
Mean Difference	Z score	p-value	ES	p-value
1	8.227	2.22E-16	128.54	0.001
0.95	8.055	8.88E-16	152.8663	0.001
0.9	8.522	0	112.3632	0.001
0.85	7.338	2.18E-13	133.268	0.003
0.8	6.706	2.01E-11	138.0922	0.001
0.75	5.232	1.68E-07	127.3383	0.001
0.7	6.932	4.18E-12	91.96096	0.001
0.65	7.067	1.59E-12	86.73476	0.001
0.6	5.078	3.82E-07	138.2932	0.001
0.55	6.071	1.28E-09	41.20655	0.014
0.5	4.481	7.44E-06	79.599	0.002
0.45	5.203	1.96E-07	20.00025	0.136
0.4	3.957	7.59E-05	60.70428	0.003
0.35	2.603	0.009241	23.71889	0.0555
0.3	2.132	0.033007	3.316625	0.7625
0.25	3.216	0.0013	7.939799	0.415
0.2	1.408	0.159131	25.9268	0.0675
0.15	1.711	0.087081	0	0.994
0.1	0.763	0.45146	9.04534	0.34
0.05	-0.289	0.772582	9.54786	0.3785



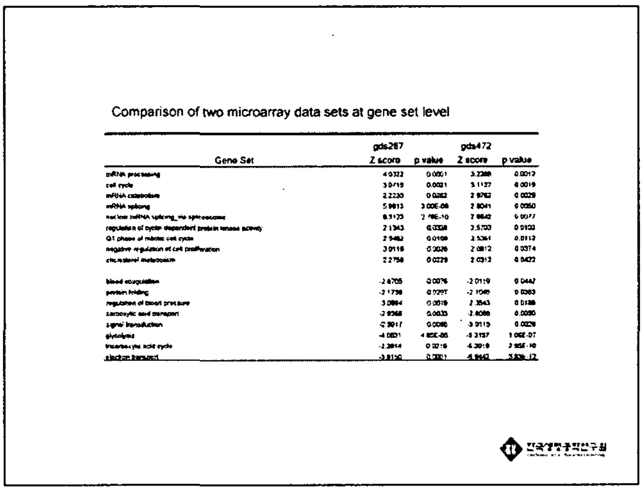
Application of PAGE to different Affymetrix probe level analysis methods

MAS5			MBEI			RMA		
Gene Set	Z score	p-value	Gene Set	Z score	p-value	Gene Set	Z score	p-value
Inflammatory Response Pathway	7.5051	< 1.0E-12	Inflammatory Response Pathway	7.5196	5.35E-14	Inflammatory Response Pathway	7.3487	2.02E-13
Eicosanoid Synthesis	6.5825	< 1.0E-12	Eicosanoid Synthesis	5.8857	9.80E-05	Eicosanoid Synthesis	4.1357	3.29E-05
Complement Activation Classical	3.0382	0.0024	Complement Activation Classical	3.5487	0.0004	TGF Beta Signaling Pathway	3.0408	0.0023
Nucleotide Metabolism	2.0536	0.0400	Nucleotide Metabolism	2.4207	0.0156	Complement Activation Classical	2.9402	0.0033
TGF Beta Signaling Pathway	1.9756	0.0482	TGF Beta Signaling Pathway	2.2379	0.0250	Nucleotide Metabolism	2.5887	0.0097
MAPK Cascade	-2.2579	0.0240	Glutamate Metabolism	-1.7249	0.0846	GPCRs Class A Rhodopsin-like	-1.9807	0.0466
Translation Factors	-2.3124	0.0208	MAPK Cascade	-1.8685	0.0617	MAPK Cascade	-3.1308	0.0031
Krebs-TCA Cycle	-3.2561	0.0011	Proteasome Degradation	-1.9808	0.0480	Krebs-TCA Cycle	-2.6076	0.0091
Glycogen Metabolism	-3.3468	0.0008	Krebs-TCA Cycle	-2.1182	0.0342	Proteasome Degradation	-2.7822	0.0054
Proteasome Degradation	-3.7468	0.0002	Glycogen Metabolism	-2.8330	0.0046	Glycogen Metabolism	-2.9328	0.0034
Fatty Acid Degradation	-3.8286	0.0001	Fatty Acid Degradation	-2.8570	0.0043	Fatty Acid Degradation	-3.4024	0.0007
Nuclear Receptors	-4.2579	2.07E-05	Nuclear Receptors	-3.4866	0.0005	Nuclear Receptors	-4.1800	3.18E-05
Electron Transport Chain	-6.3786	1.79E-10	Electron Transport Chain	-4.2008	2.66E-05	Electron Transport Chain	-5.1177	3.10E-07



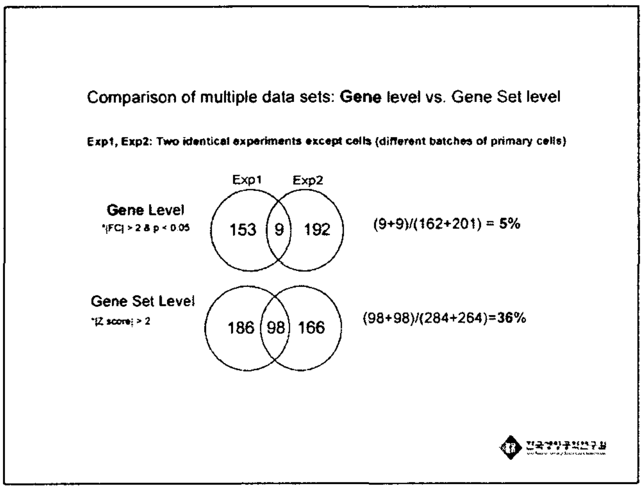
Comparison of PAGE results from data sets produced using different microarray platforms

U95A			U133A			Agilent		
Gene Set	Z score	p-value	Gene Set	Z score	p-value	Gene Set	Z score	p-value
Complement Activation Classical	3.0477	0.0023	Krebs TCA Cycle	5.0605	4.41E-07	tRNA Synthetases	3.1714	0.0015
Krebs-TCA Cycle	2.6771	0.0074	Cell Cycle	3.4086	0.0007	Krebs TCA Cycle	3.1291	0.0018
Nuclear Receptors	2.6562	0.0079	Translation Factors	2.8213	0.0048	Proteasome Degradation	1.9804	0.0487
Calcium Channels	1.6806	0.0906	Nuclear Receptors	2.3484	0.0190	Glycolysis and Gluconeogenesis	1.5879	0.1160
Apoptosis	1.4086	0.1590	Complement Activation Classical	2.2881	0.0223	Steroid Biosynthesis	1.5109	0.1308
TGF Beta Signaling Pathway	-1.8560	0.0632	Matrix Metalloproteinases	-2.2837	0.0224	Ribosomal Proteins	-2.1275	0.0334
Inflammatory Response Pathway	-2.2580	0.0239	TGF Beta Signaling Pathway	-3.3782	0.0007	Glycogen Metabolism	-2.4804	0.0128
Glycogen Metabolism	-2.6881	0.0070	Inflammatory Response Pathway	-3.5888	0.0004	TGF Beta Signaling Pathway	-2.6910	0.0071
Cholesterol Biosynthesis	-4.6866	2.62E-06	Cholesterol Biosynthesis	-5.8165	6.03E-09	Cholesterol Biosynthesis	-4.1842	3.13E-05
Gap Junction Proteins-Connexins	-5.7792	7.50E-09	Gap Junction Proteins-Connexins	-6.2889	3.20E-10	Gap Junction Proteins-Connexins	-6.6182	3.71E-11



Comparison of two microarray data sets at gene set level

	gds287		gds472	
Gene Set	Z score	p value	Z score	p value
mRNA processing	4.0322	0.0001	3.2288	0.0012
cell cycle	3.0719	0.0021	3.1127	0.0019
mRNA catabolism	2.2230	0.0262	2.9782	0.0029
mRNA splicing	5.8813	3.00E-08	2.8041	0.0050
nuclear mRNA splicing_via spliceosome	6.3123	2.76E-10	2.6642	0.0077
regulation of cyclin dependent protein kinase activity	2.1343	0.0328	2.5700	0.0102
G1 phase of mitotic cell cycle	2.5482	0.0108	2.5361	0.0112
negative regulation of cell proliferation	3.0118	0.0026	2.0812	0.0374
cholesterol metabolism	2.2758	0.0229	2.0312	0.0422
blood coagulation	-2.6705	0.0076	-2.0119	0.0442
protein folding	-2.1738	0.0297	-2.1045	0.0363
regulation of blood pressure	-3.0864	0.0019	2.3543	0.0186
carboxylic acid transport	-2.9368	0.0033	-2.8086	0.0050
signal transduction	-2.5917	0.0096	-3.0115	0.0026
glycolysis	-4.0621	4.86E-05	-5.3157	1.06E-07
tricarboxylic acid cycle	-2.3814	0.0016	-6.3019	2.95E-10
electron transport	-3.9150	0.0001	-6.9442	3.80E-12



Comparison of multiple data sets: **Gene** level vs. Gene Set level

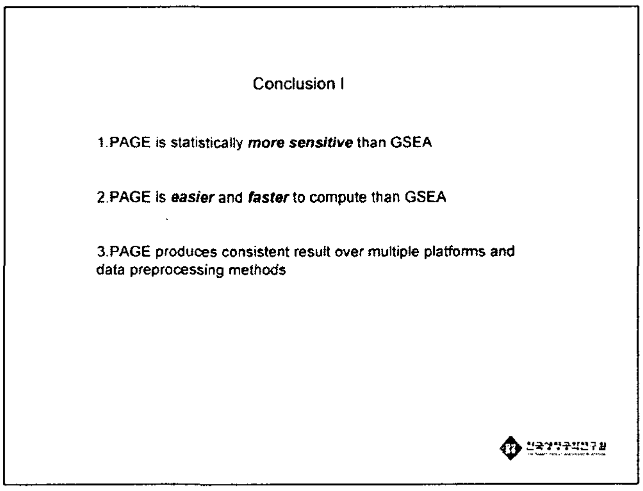
Exp1, Exp2: Two identical experiments except cells (different batches of primary cells)

Gene Level
*|FC| > 2 & p < 0.05

(9+9)/(162+201) = **5%**

Gene Set Level
*|Z score| > 2

(98+98)/(284+264)=**36%**

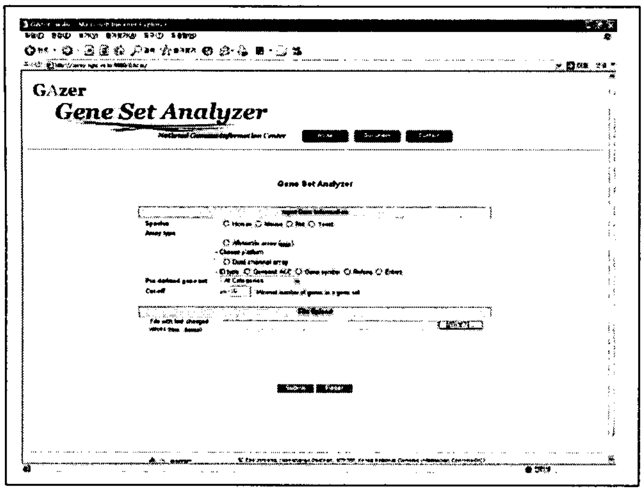


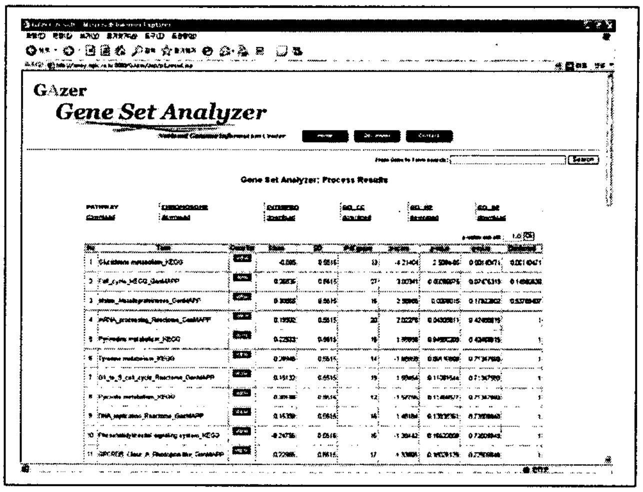
Conclusion I

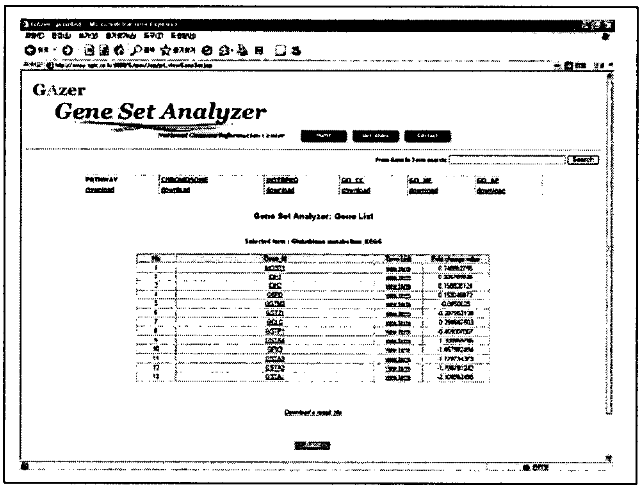
1.PAGE is statistically ***more sensitive*** than GSEA

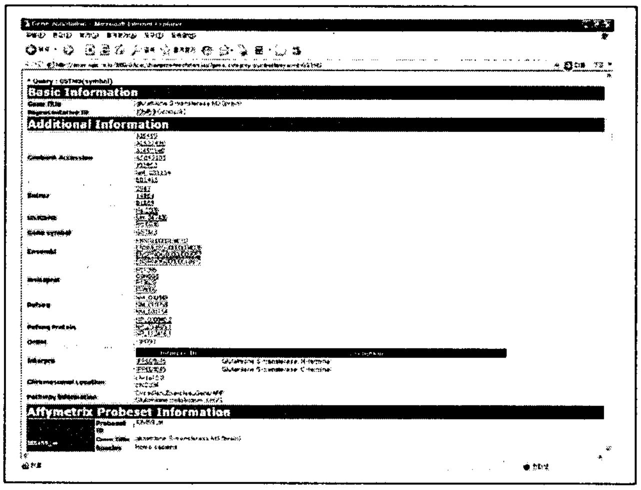
2.PAGE is ***easier*** and ***faster*** to compute than GSEA

3.PAGE produces consistent result over multiple platforms and data preprocessing methods

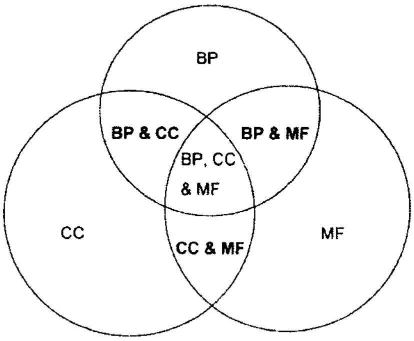


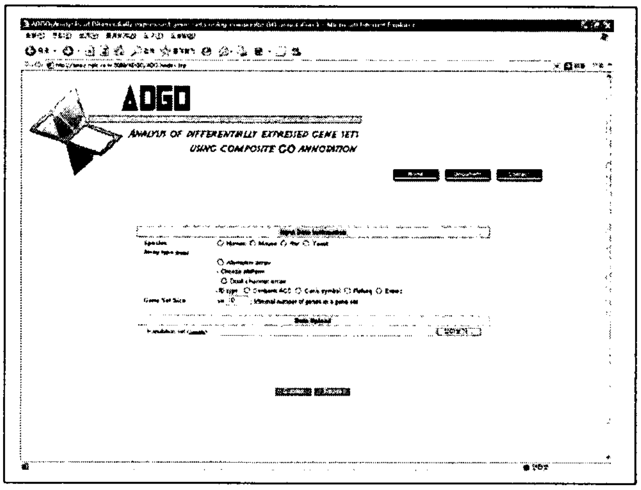


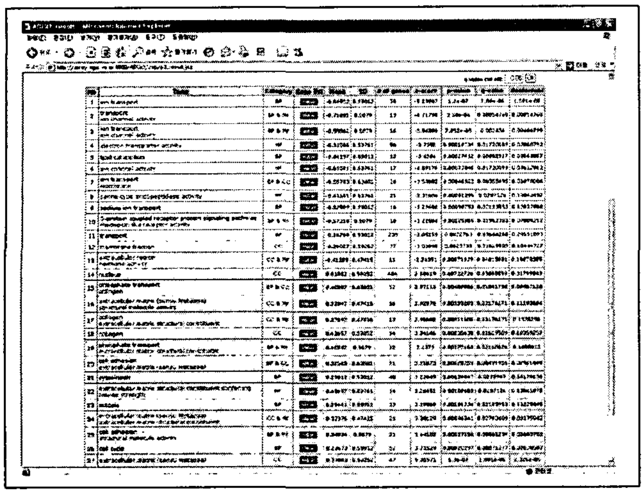


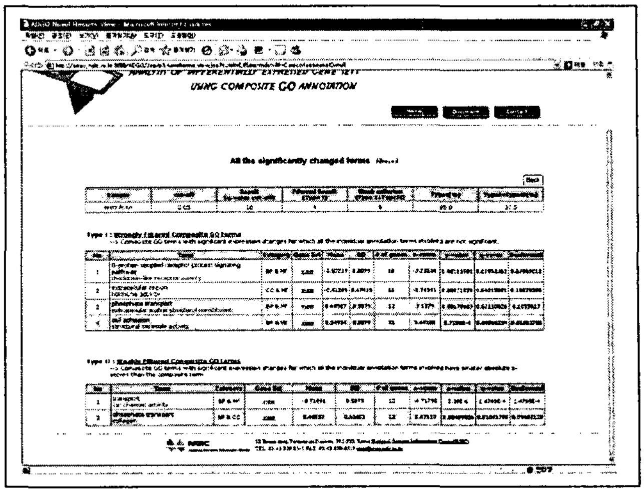


ADGO: Analysis of Differentially Expressed Gene Sets Using Composite GO Annotation



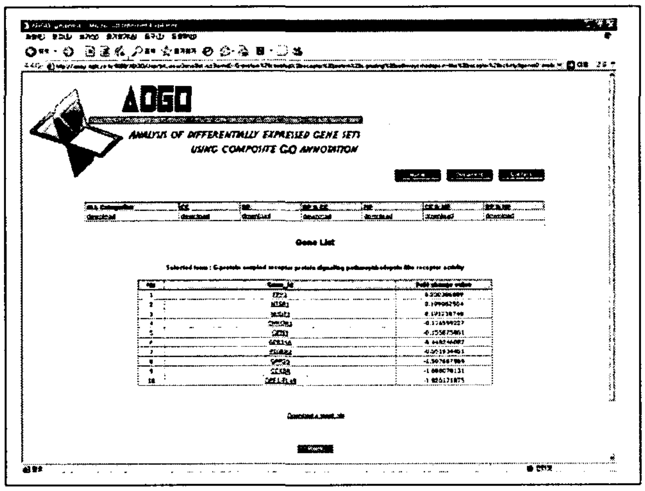

ADGO
ANALYSIS OF DIFFERENTIALLY EXPRESSED GENE SETS USING COMPOSITE GO ANNOTATION
Species
Gene Set Size
Data Upload





USING COMPOSITE GO ANNOTATION

All the significantly changed terms



ADGO

ANALYSIS OF DIFFERENTIALLY EXPRESSED GENE SETS USING COMPOSITE GO ANNOTATION

Gene List

Recent Works in FCS (Functional Class Scoring) Methods

Pavlidis, P. et al., 2002, *Pac. Symp. Biocomput.* 474-485	
Mootha et al., 2003, *Nat. Gen.* 34: 267-273	
Breslin et al., 2004, *BMC Bioinformatics*, 5: 193˙	=> Catmap
Al-Shahrour et al., 2005, *Bioinformatics*, 21: 2988-93	
Tu et al., 2005, *Biotechniques*, 2: 277-83	=> MEGO
Kim & Volsky, 2005, *BMC Bioinformatics*, 6: 144	=> PAGE, GAzer
Boorsman et al., 2005, *Nuc. Acids Res.*, 33	=> T-Profiler
Tian et al., 2005, *PNAS*, 38: 13544-9	
Subramanian et al., 2005, *PNAS*, 38: 15545-50	=> GSEA
Lee et al., 2005, BMC *Bioinformatics*, 6: 269	=> ErmineJ

Comparison of Gene Set Analysis Softwares

	PAGE	GSEA	ErmineJ	MEGO	Catmap	T-profiler
Used Statistics	Fold Change	Rank	P-value	Fold change	P-value	T-statistic
Statistical test	Z-test	permutation	permutation		permutation	t-test
Speed	Fast	Slow	Moderate	Fast	Slow	Fast
Standalone software	YES	YES	YES	YES	YES	NO
GUI	NO	YES	YES	YES	NO	
Web server	YES	NO	NO	NO	NO	YES
Organism	H, M, R, Y	H, M	H, M, R	H,		Y
Gene Sets	GO, Pathways, Chromosomes	GO, Pathways, Chromosomes	GO, Pathways, Chromosomes	GO, Pathways, Chromosomes	GO, Pathways, Chromosomes	GO, Pathways, Chromosomes

ACKNOWLEDGEMENTS

NGIC (National Genome Information Center)

In-Sun Chu
Dougu Nam
Sang-Bae Kim
Sang-Cheol Kim
Seon-Kyu Kim
Seong-Jin Yang
Hyun-Goo Woo

Columbia University

David J Volsky