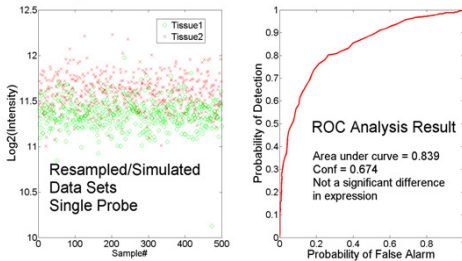


PURPOSE

To test a new methodology for identification of genomic biomarkers in **microarray experiments with small number of replicates**

METHODS: Numerical Analysis

- Non-parametric estimate of noise distribution is derived through resampling of analyzed data set
 - At least two replicates are required per condition
 - Noise distributions are not normal
- The differentially expressed genes (transcripts) are identified with ROC (Receiver Operating Characteristic) analysis of enhanced data sets



The approach viability is shown through:

- Comparisons with three 'benchmark' methods: RankProd, SAM, MAANOVA
- Knowledge-based analysis of results, identification of tissue-specific genes

METHODS: Test Data

Mouse Gene Atlas (Genomics Institute of the Novartis Research Foundation, GNF, San Diego, CA):

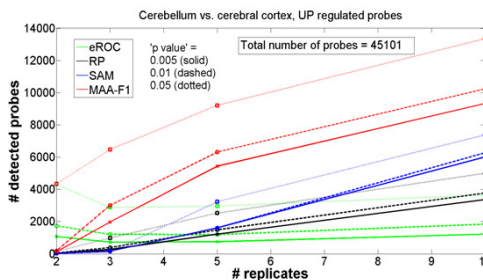
- Expression data for 78 normal mouse tissues obtained with Affymetrix mouse array (MOE430.2)
 - Two experimental repeats per tissue.
- Simulated data produced using the eROC resampling
 - To study the effects of a larger number of experimental repeats
- Lists of differentially expressed probes were produced by each of the methods for pair-wise comparisons of tissues.
 - Tissues with diverse levels of similarity were compared.

RESULTS: Probe sets detected by eROC depend weakly on the number of repeats

Number of detected probes varies significantly between the methods
When similar detection thresholds are used
Number of probes declared as up/down regulated depends strongly on the number of repeats for all benchmark methods.

- Significant increase with the number of experimental repeats, up to a large fraction of the total number of probes

eROC detects similar number of probes in experiments with different numbers of replicates



eROC and RankProd show high consistency in detection of expressed probe sets

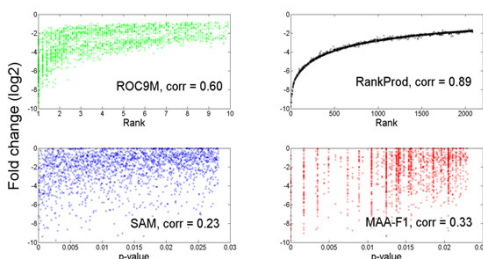
- Fraction of overlapping detected probes for experiments with different numbers of repeats (2-10)*

eROC: 0.8 RankProd: 0.9

SAM: 0.5 MAANOVA: 0.5

* averages for top 1000-2000 probes

RESULTS: Correlation between rank and fold change



Average correlation (Pearson) between measured fold change and method-specific probe rank:

eROC	RankProd	SAM	MAANOVA
0.60	0.88	0.25	0.33 / 0.42

- Std \approx 0.07, 10 experiments, top 2000 probes

RESULTS: Detection of tissue specific probes

eROC and RankProd exceed in detection of tissue specific probes

We determine the percent of known tissue specific genes ranked in top 2000 by different methods. 2 arrays/tissue, tissue pairwise comparisons, several data sources.

DAVID Bioinformatics Resources 6.7 (NIAID)

Tissue pair	eROC	RankProd	SAM	MAA	Affy array
Liver	37.3	40.5	29.3	32.1	15.5
Cerebellum	14.0	12.6	12.3	13.5	6.8
Pancreas	7.9	7.7	5.2	6.8	2.9
Cerebellum	16.6	15.2	13.5	15.8	6.8
Kidney	24.1	33.5	19.3	13.9	9.8
Cerebellum	13.9	13.5	11.9	14.7	6.8
Lung	10.0	10.7	10.4	9.8	5.9
Cerebellum	12.7	13.7	12.8	14.0	6.8
Testis	44.6	47.8	29.2	44.7	15.0
Cerebellum	11.6	12.2	10.5	8.9	6.8

ALLEN MOUSE BRAIN ATLAS *in situ* hybridization data: cerebral cortex and cerebellum specific probes

Tissues	# probes	eROC	RP	SAM	MAA
Cortex vs. Cerebellum	171	53.8	49.7	40.9	51.5
Cortex vs. Hippocampus	139	55.4	55.4	36.0	47.5
Cortex vs. Olfactory Bulb	171	49.1	49.7	30.4	42.7
Cortex vs. Hypothalamus	173	46.8	45.7	28.3	35.3
Average		51.3	50.1	33.9	44.2
Cerebellum vs. Cortex	395	46.1	45.6	34.4	42.8
Cerebellum vs. Hypothalamus	408	42.9	43.4	33.3	39.2
Cerebellum vs. Hippocampus	365	42.5	43.8	32.3	37.8
Cerebellum vs. Olfactory Bulb	386	39.1	40.9	31.9	37.6
Average		42.6	43.4	33.0	39.3
Cerebellum vs. Kidney	420	24.8	24.5	15.0	23.1
Cerebellum vs. Lung	420	24.0	23.6	15.7	20.7
Cerebellum vs. Testis	420	23.3	22.9	11.0	19.3
Cerebellum vs. Liver	420	22.6	21.9	14.0	17.9
Average		23.7	23.2	13.9	20.2

Liver specific genes from oligonucleotide microarray analysis of human tissues (Hsiao LL, Physiol Genomics 21:95-6, 2001)

Tissues	#probes	eROC	RankProd	SAM	MAA-FS
Liver / Cerebellum	337	75.1	69.4	43.3	57.6
Liver / Hypothalamus	337	73.3	69.7	40.4	60.8
Liver / Hippocampus	337	75.1	70.6	40.7	67.1
Liver / Olfactory Bulb	337	74.8	70.6	39.2	61.7
Average		74.6	70.1	40.9	61.8

CONCLUSIONS

eROC is useful for reliable detection of differentially expressed probes in microarray experiments with a small number of replicates (two or more)

- Applicable to all microarray and other molecular expression platforms
- Non-parametric approaches (eROC and RankProd) perform better than parametric methods (SAM, MAANOVA).
 - RankProd ranking is highly correlated with fold change
 - RankProd detects small number of probes in experiments with a few replicates.