

ABRF Microarray Research Group

microRNA Profiling: Platform Comparison

**Don Baldwin – Penn Microarray Facility,
University of Pennsylvania**

Project Goals

- Examine multiple microarray and next-gen sequencer platforms for performance in miRNA profiling
- Provide information on sensitivity, reproducibility, and concordance among platforms
- Make data available for reference in selecting & running miRNA profiling assays

Project History

- Study begun as 2009 MARG project in collaboration with DSRG
- Preliminary report on microarray component of study and single miRNA seq result presented at ABRF 2009
- Over last year, additional sequencing data has been generated on Illumina and ABI SOLiD platforms in MARG member labs

ABRF MicroArray Research Group

microRNA Profiling: Platform Comparison

Study Design & Implementation

Design

- Use commercial total RNA from 2 different human tissues
- Perform miRNA profiling on 4 microarray platforms (Agilent, Affymetrix, Exiqon, and Illumina) and Taqman (ABI) low density arrays in triplicate
- Perform sequencing on 2 deep-sequencing platforms (ABI SOLiD and Illumina GA)
- Analyze data from each platform for detection sensitivity, data reproducibility and profile concordance across platforms

Disclaimer: mention of products or trade names does not constitute and endorsement

Implementation

- Commercial RNA purchased from Ambion: First Choice human total RNA – liver and brain.
- Each platform run in laboratory of different MARG or DSRG member.
- Assays were run according to manufacturer recommendations as implemented in member lab.

Platform	Array
ABI TLDA	Taqman human miRNA A&B v2.0
Affymetrix	GeneChip miRNA
Agilent	Human miRNA
Exiqon	miRCURY LNA miRNA
Illumina	Human miRNA v2 Panel

HT Seq Platform
Illumina GA II
ABI SOLiD

General methods

- RNA aliquots distributed from central site to all member labs performing microarray or sequencing assay.
- No LMW RNA enrichment for microarray platforms; small RNA enrichment for sequencing platforms.
- Triplicate assays were performed simultaneously for each microarray assay method.
- No replicates for sequencing (because of cost), except that Illumina sequencing was performed twice in separate labs.

Data analysis

- Primary data collected and analyzed centrally
- Microarray and Taqman data analyzed with Partek Genomics Suite software at University of Pennsylvania
- Sequencing data aligned and analyzed with GeneSifter (GeoSpiza) software at Oregon Health & Science University

ABRF MicroArray Research Group

microRNA Profiling: Platform Comparison

Microarray Results

Partek Genomic Suite data analyses

Platform	Raw data	Reference	Normalization	Filter
TaqMan	avg delta Ct	4-6 endogenous targets	median shift	
GeneChip	.CEL	all, one channel	RMA	"hsa-"
DASL	AVG Signal	all, one channel	log quantile	"hsa-"
Agilent	TotalGeneSignal	all, one channel	floor, log quantile	
miRCURY	.GPR bkgd corr median	all, dye swap pairs	floor, loess log ratios	"hsa-", flagged

rank order: top 400 expression levels

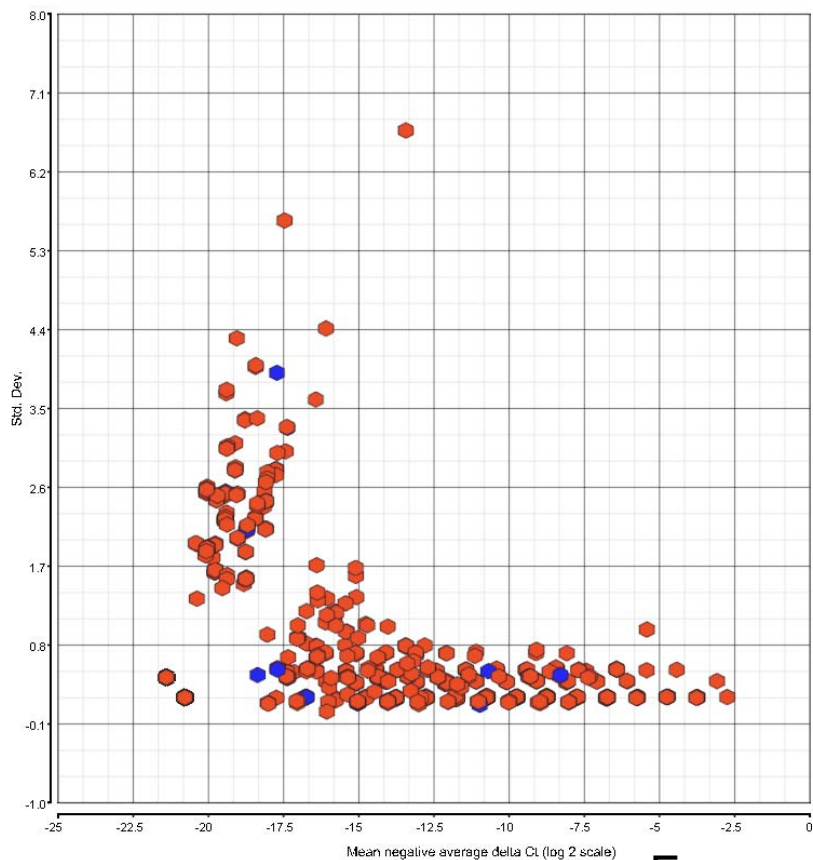
common targets: miRBase accession number

SAM: Significance Analysis for Microarrays, FDR 5%

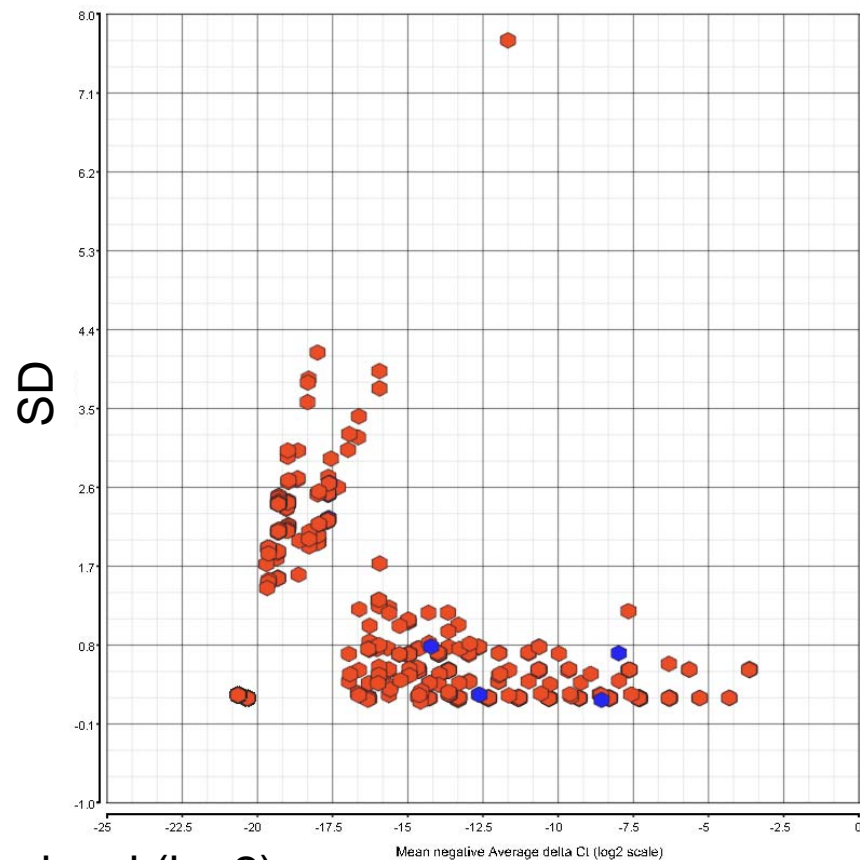
Mean and Standard Deviation of three technical replicates

Applied Biosystems TaqMan Low Density Arrays

Brain



Liver

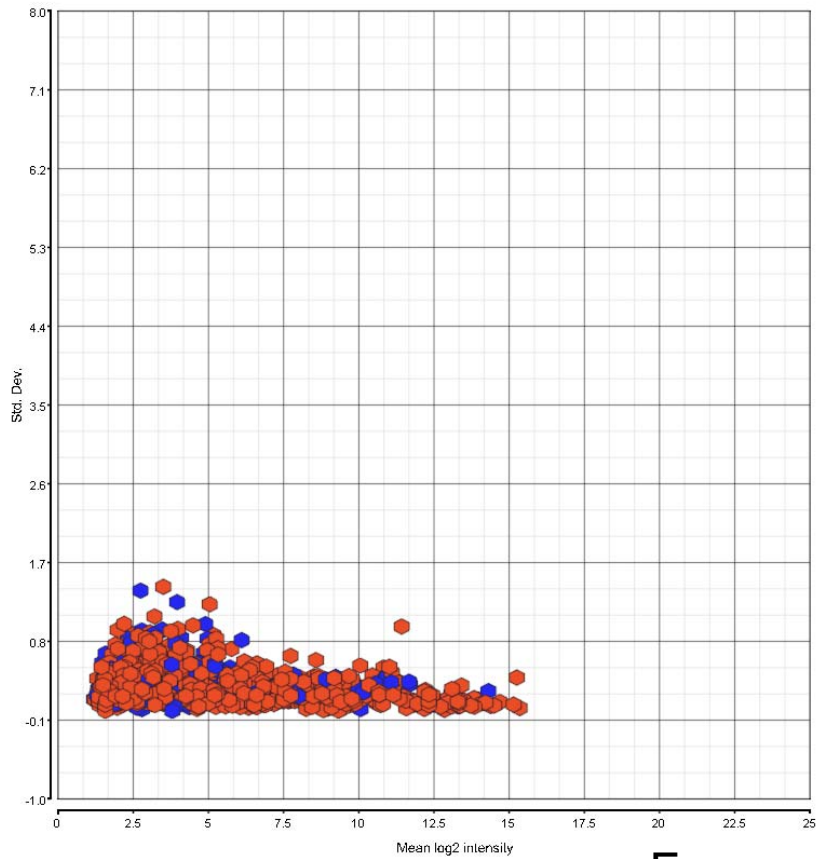


column 3
● common
● unique

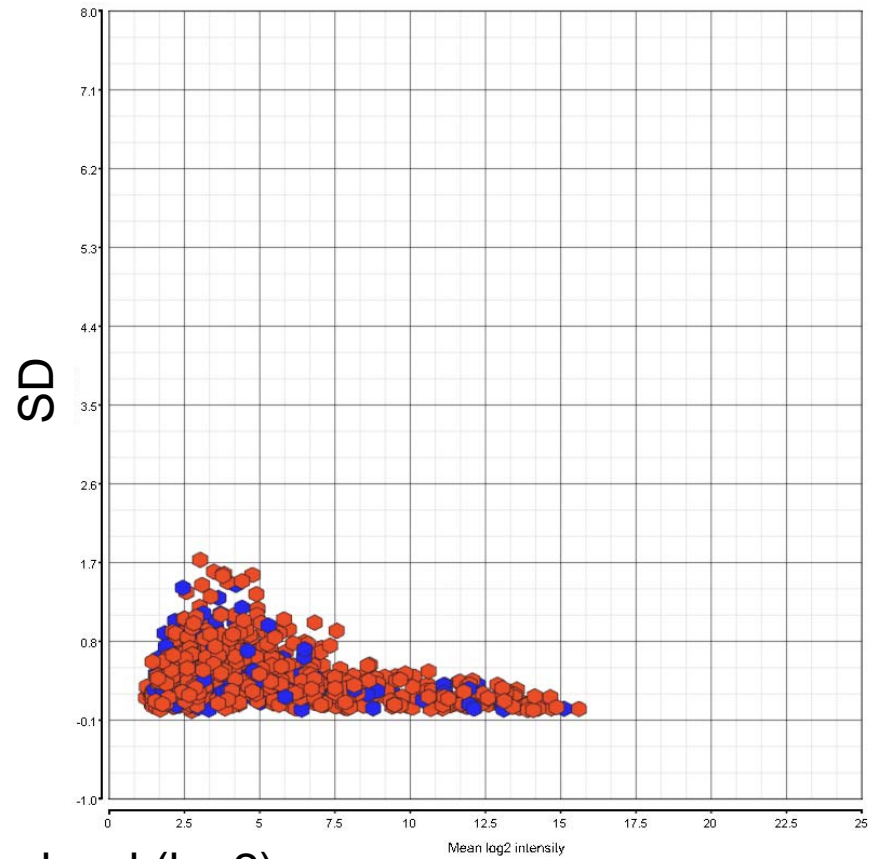
Expression level (log2)

Affymetrix microRNA GeneChip

Brain



Liver

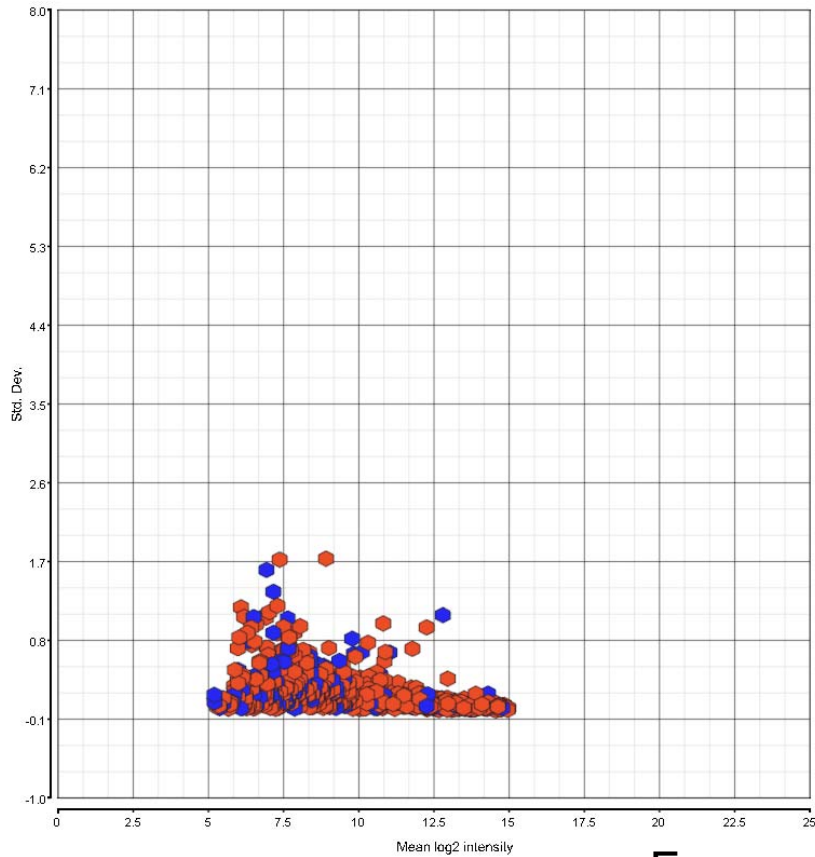


probe type
● common
● unique

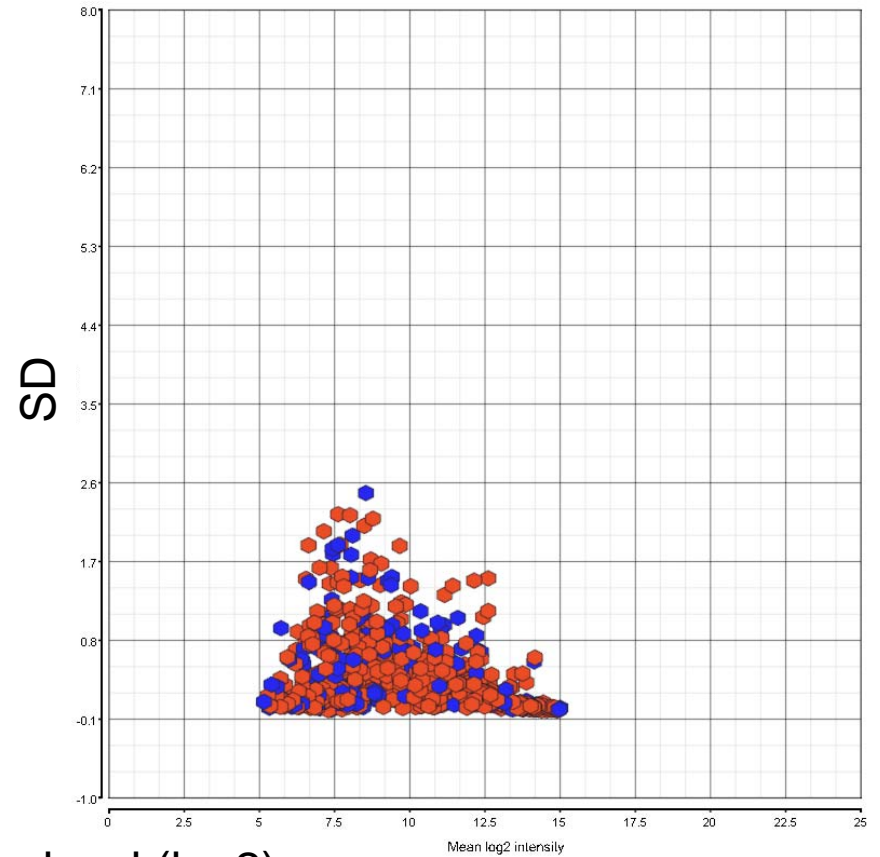
Expression level (log2)

Illumina microRNA DASL assay

Brain



Liver

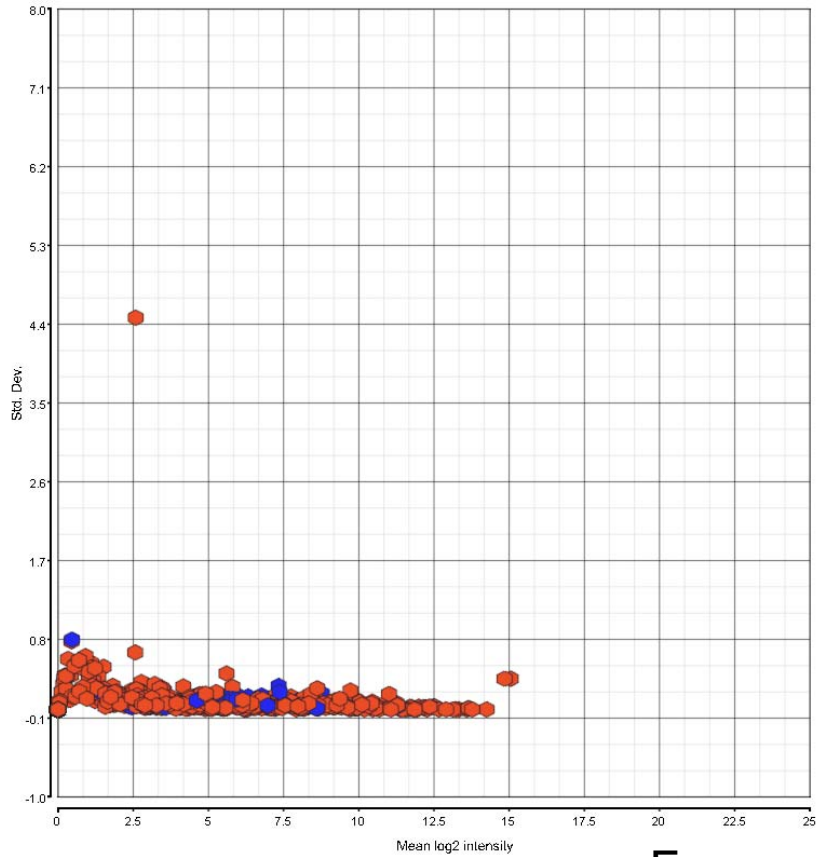


probe type
● common
● unique

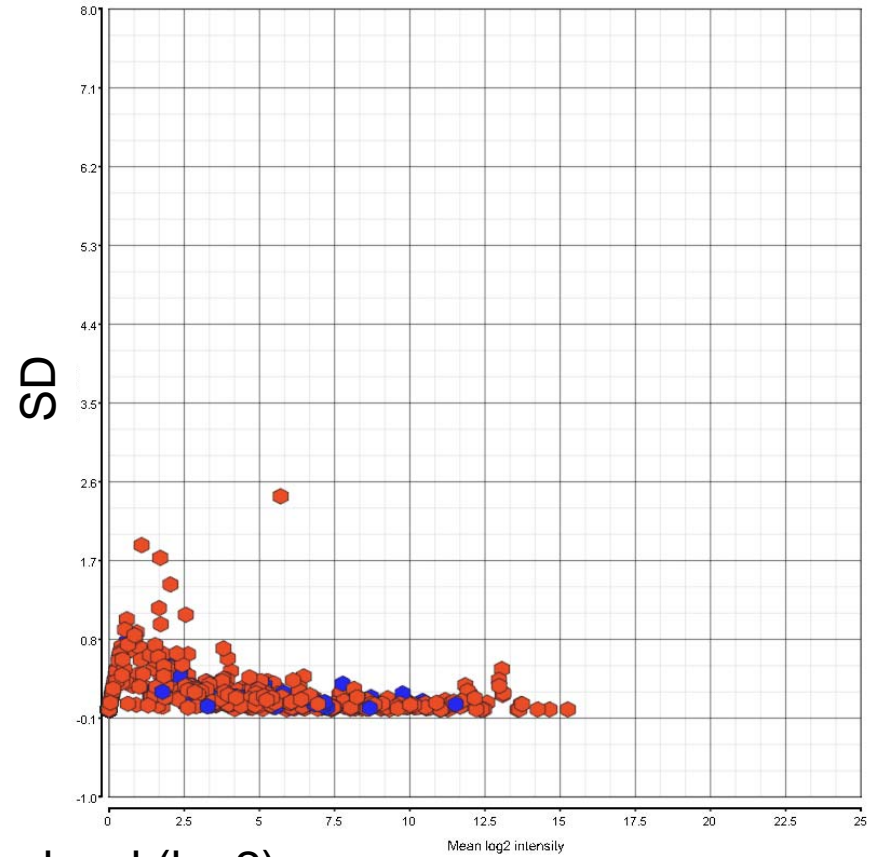
Expression level (log2)

Agilent SurePrint glass slides

Brain



Liver

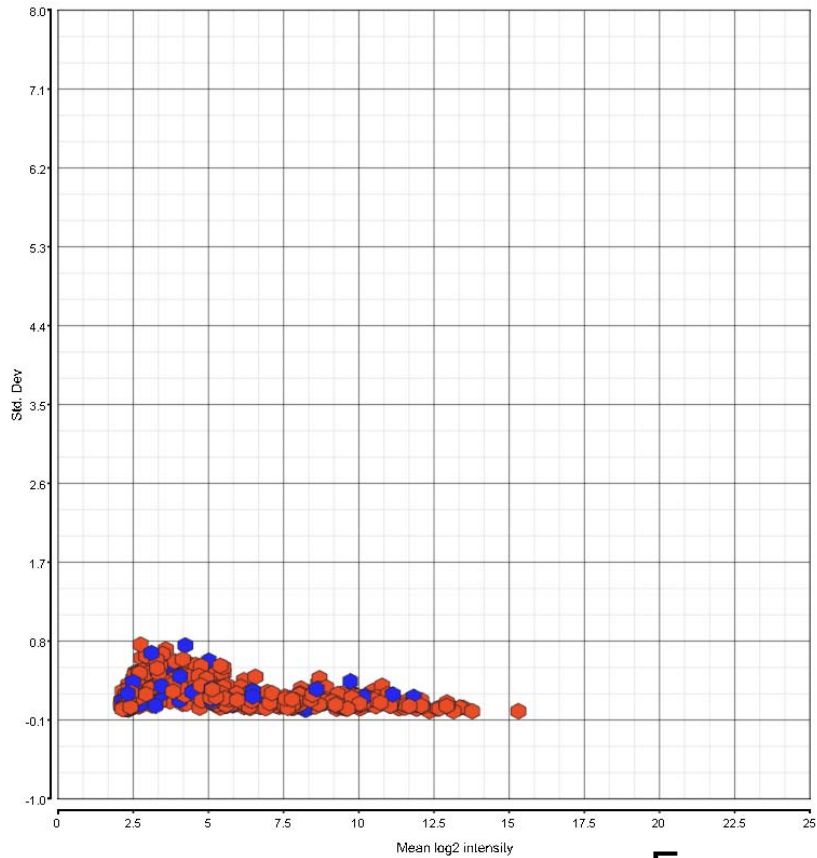


probe type
● common
● unique

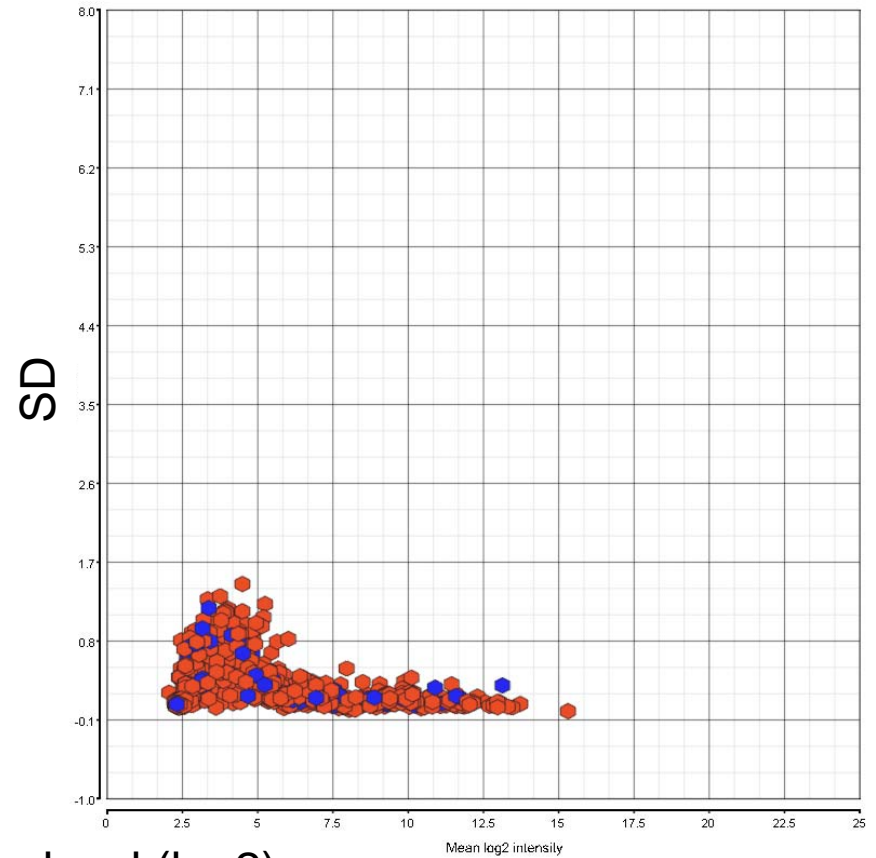
Expression level (log2)

Exiqon miRCURY LNA glass slides

Brain



Liver



probe type
● common
● unique

Expression level (log2)

Log2 dynamic range

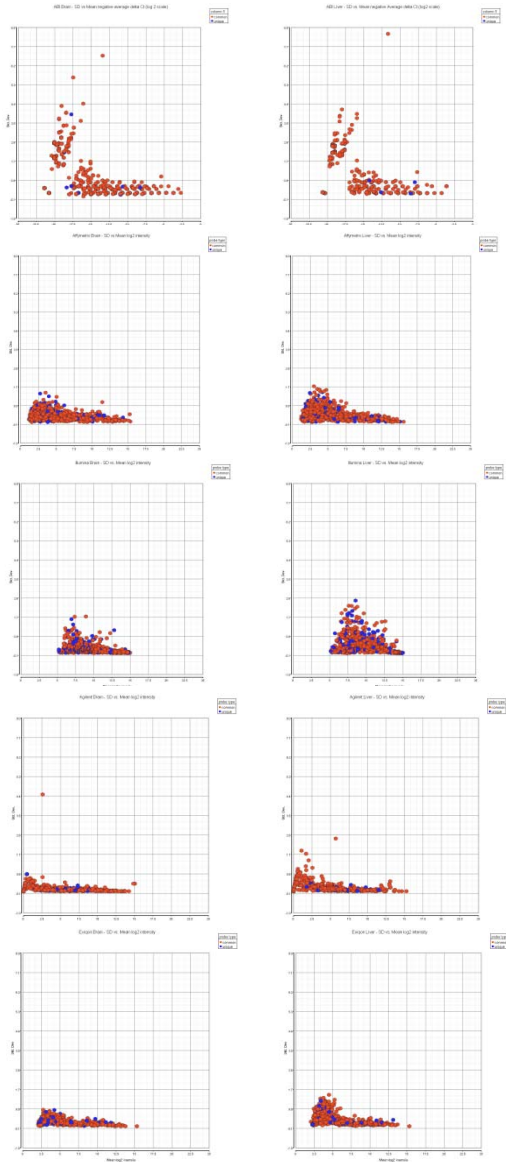
ABI 20

Affymetrix 15

Illumina 10

Agilent 15

Exiqon 13



Correlation coefficients: technical replicates

Table 1. Brain technical replicate correlation (Average Pearson Correlation Coefficients, R)

Platform	Total number of human probes	Normalization method	All probes	Top 400 probes (Brain)	Common human probes (639)
ABI	664	Median shift	0.9449	0.9250	0.9449
Affymetrix	847	RMA	0.9883	0.9877	0.9899
Illumina	858	Quantile	0.9855	0.9864	0.9870
Agilent	723	Quantile	0.9968	0.9990	0.9967
Exiqon (BG subtracted)	739	Loess	0.9924	0.9922	0.9926

Table 2. Liver technical replicate correlation (Average Pearson Correlation Coefficients, R)

Platform	Total number of human probes	Normalization method	All probes	Top 400 probes (Liver)	Common human probes (639)
ABI	664	Median shift	0.9384	0.9277	0.9376
Affymetrix	847	RMA	0.9771	0.9781	0.9785
Illumina	858	Quantile	0.9527	0.9373	0.9562
Agilent	723	Quantile	0.9952	0.9955	0.9952
Exiqon (BG subtracted)	739	Loess	0.9713	0.9697	0.9701

Detection of differential expression: Brain vs. Liver

Table 3. Differentially expressed miRNAs detected from common set of 639 human probes

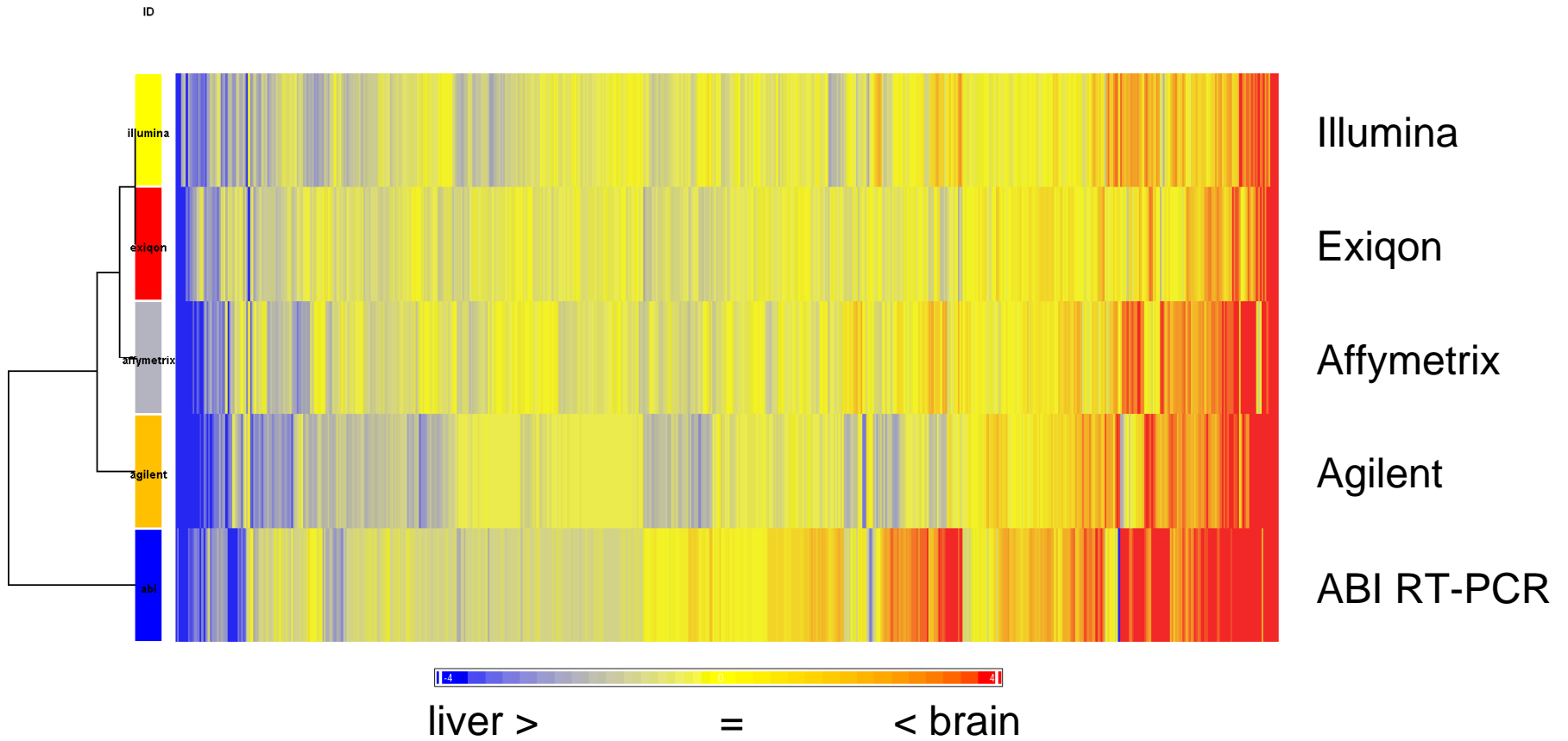
Platform	Significantly different miRNAs (FDR ≤ 5)	miRNAs with two-fold or greater difference	miRNAs with significant difference of $>2x$	Maximum negative fold-change detected	Maximum positive fold-change detected
ABI	206	338	206	-57586	2607
Affymetrix	305	243	238	-620	5153
Illumina	347	248	247	-109	339
Agilent	347	357	340	-11345	3955
Exiqon (BG subtracted)	362	184	184	-2929	734

Concordance for detection of differential expression: common target set

Table 4. Pearson correlation coefficients for differentially expressed miRNAs

Platform	ABI	Affymetrix	Illumina	Agilent	Exiqon	Mean concordance	Median concordance
ABI	1	0.7570	0.7426	0.6999	0.6699	0.7739	0.7426
Affymetrix	0.7570	1	0.7411	0.7580	0.7163	0.7945	0.7570
Illumina	0.7426	0.7411	1	0.7381	0.6701	0.7784	0.7411
Agilent	0.6999	0.7580	0.7381	1	0.8304	0.8053	0.7580
Exiqon	0.6699	0.7163	0.6701	0.8304	1	0.7773	0.7163

Hierarchical clustering: all human miRNAs log2 ratio



ABRF MicroArray Research Group

microRNA Profiling: Platform Comparison

Sequencing Results

GeneSifter (GeoSpiza) analysis pipeline

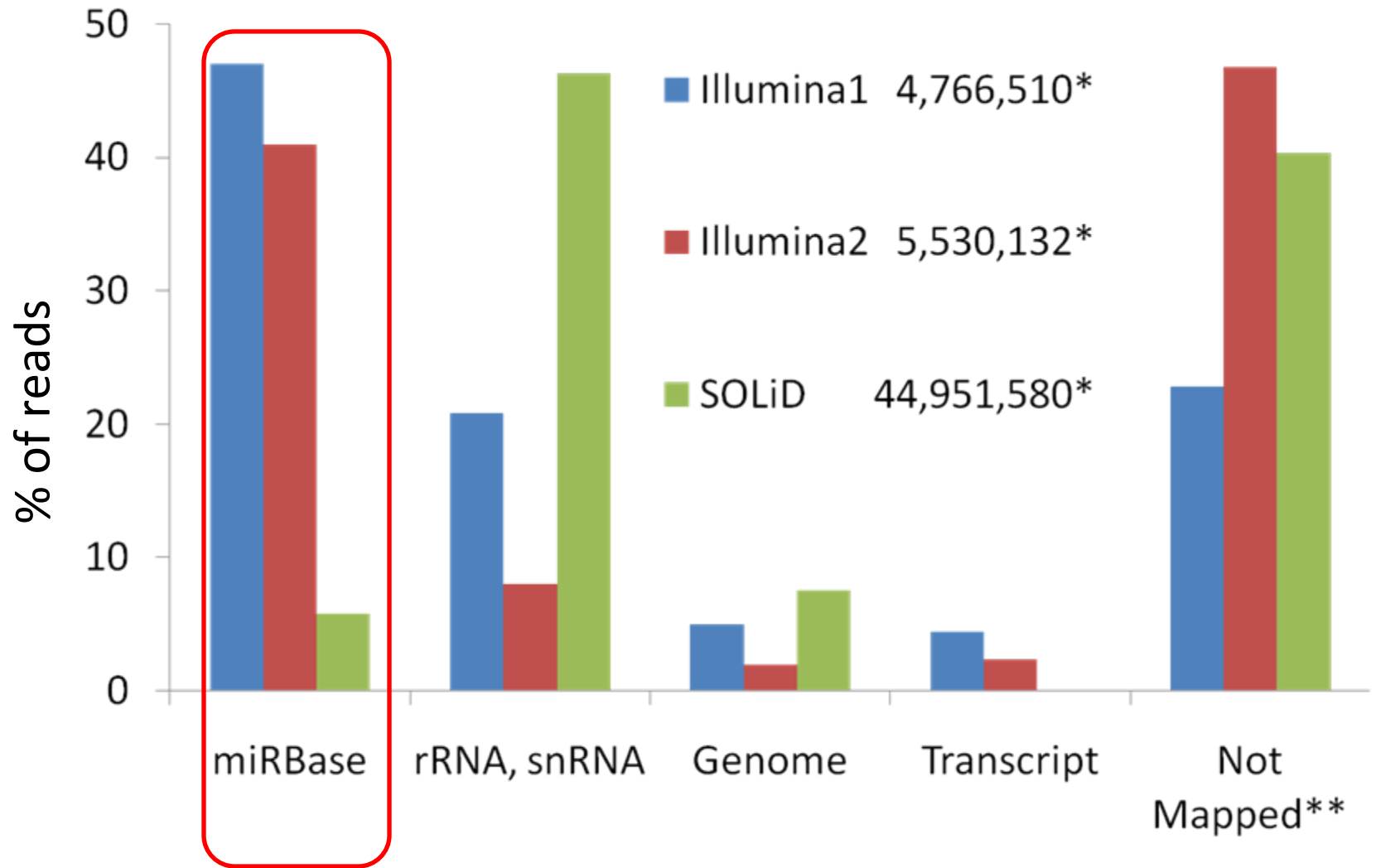
	<u>Illumina</u>	<u>SOLiD</u>
Input File:	FASTQ	csfasta
Alignment algorithm:	bowtie	mapreads

Alignment Flow:

1. Input file (raw reads) aligned to genome of reference species.
2. Reads mapping to rRNA, tRNA, mtRNA, etc. define “filtered” reads.
3. Reads mapping to miRBase genome coordinates define known miRNAs.
4. Reads mapping outside any of the defined genome references are intergenic.
5. Reads mapping to multiple genomic coordinates set aside as “non-uniquely mapped”.
6. Reads not mapping at all set aside as “Not mapped”.

Normalized counts: reads per million mapped miRNAs

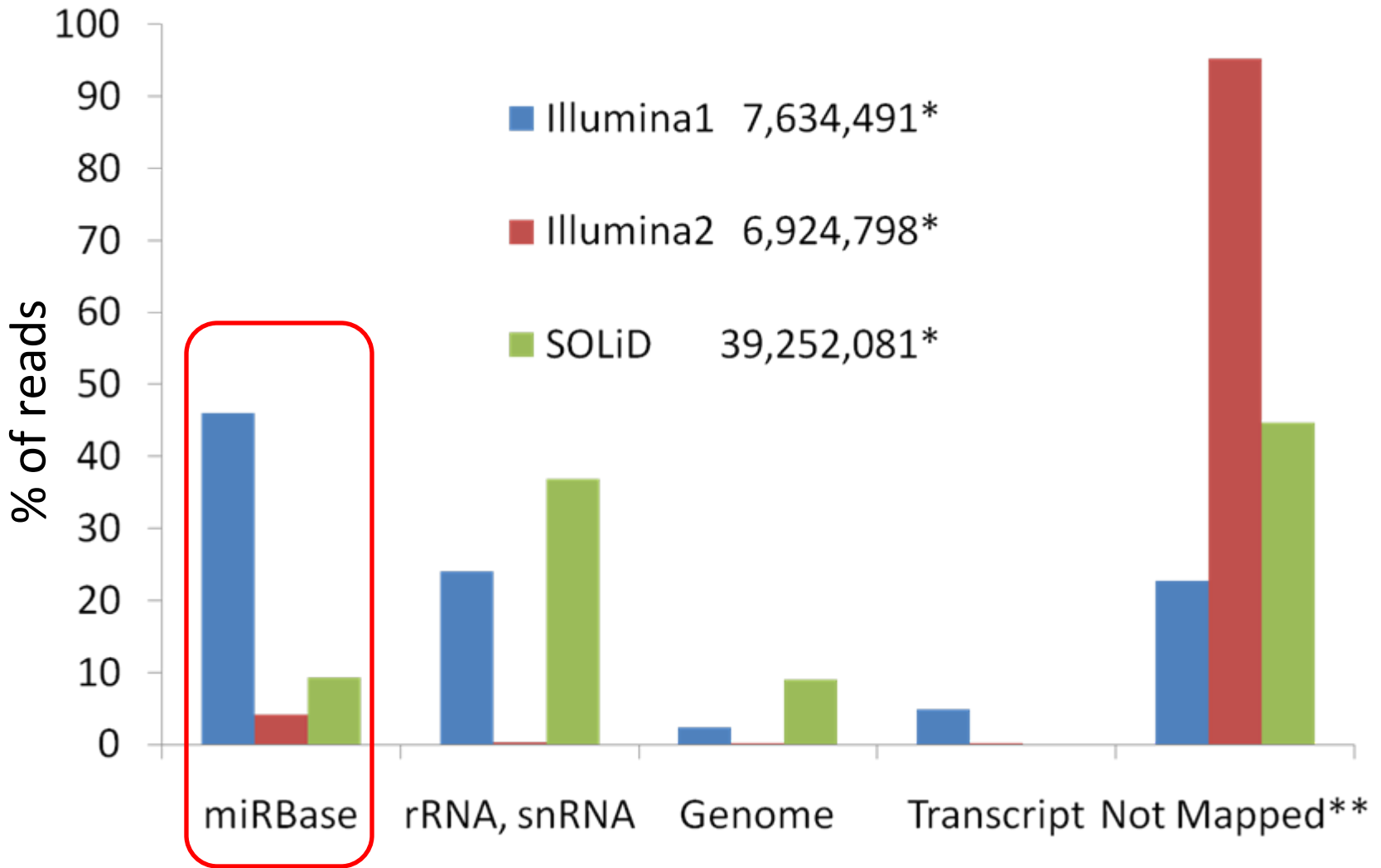
Distribution of Reads (Brain)



* Total # of reads

** Adapter or not mapped

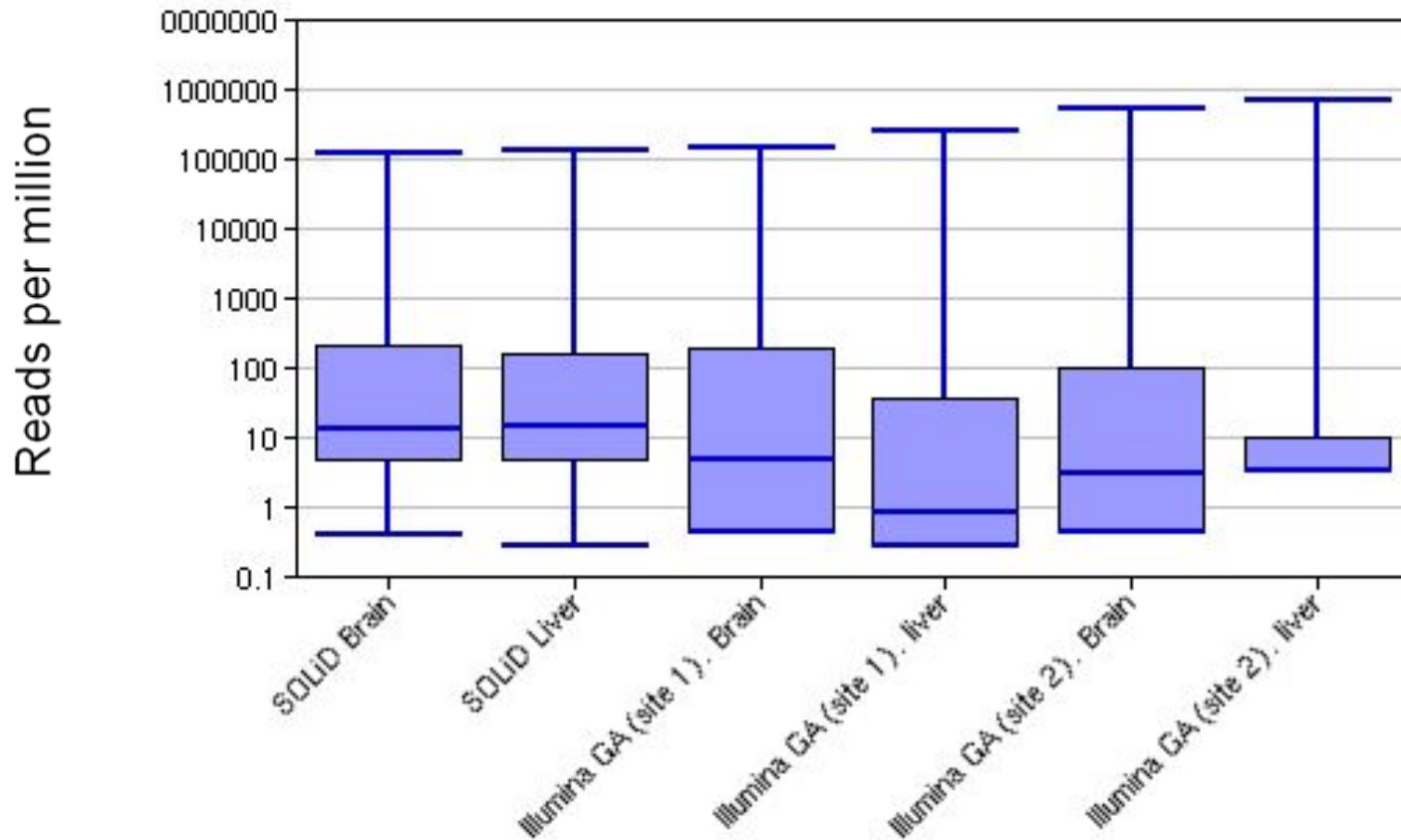
Distribution of Reads (Liver)



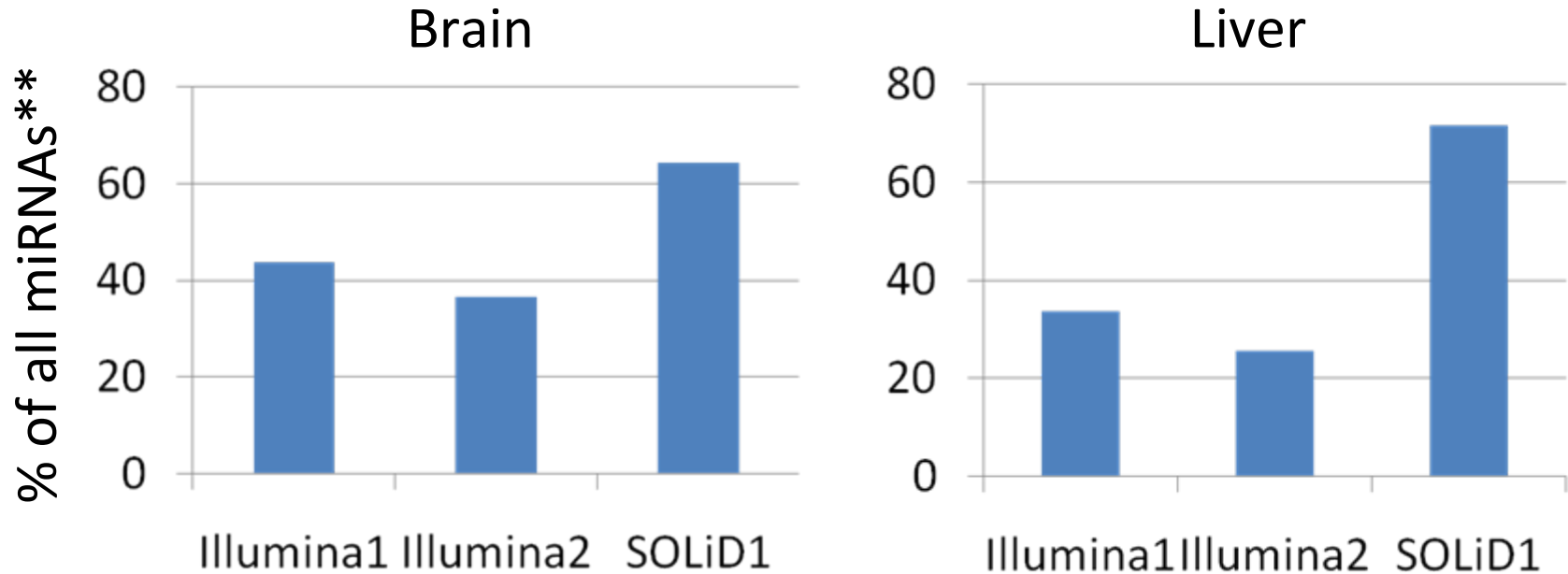
* Total # of reads

** Adapter or not mapped

Expression range: miRNA mapped reads



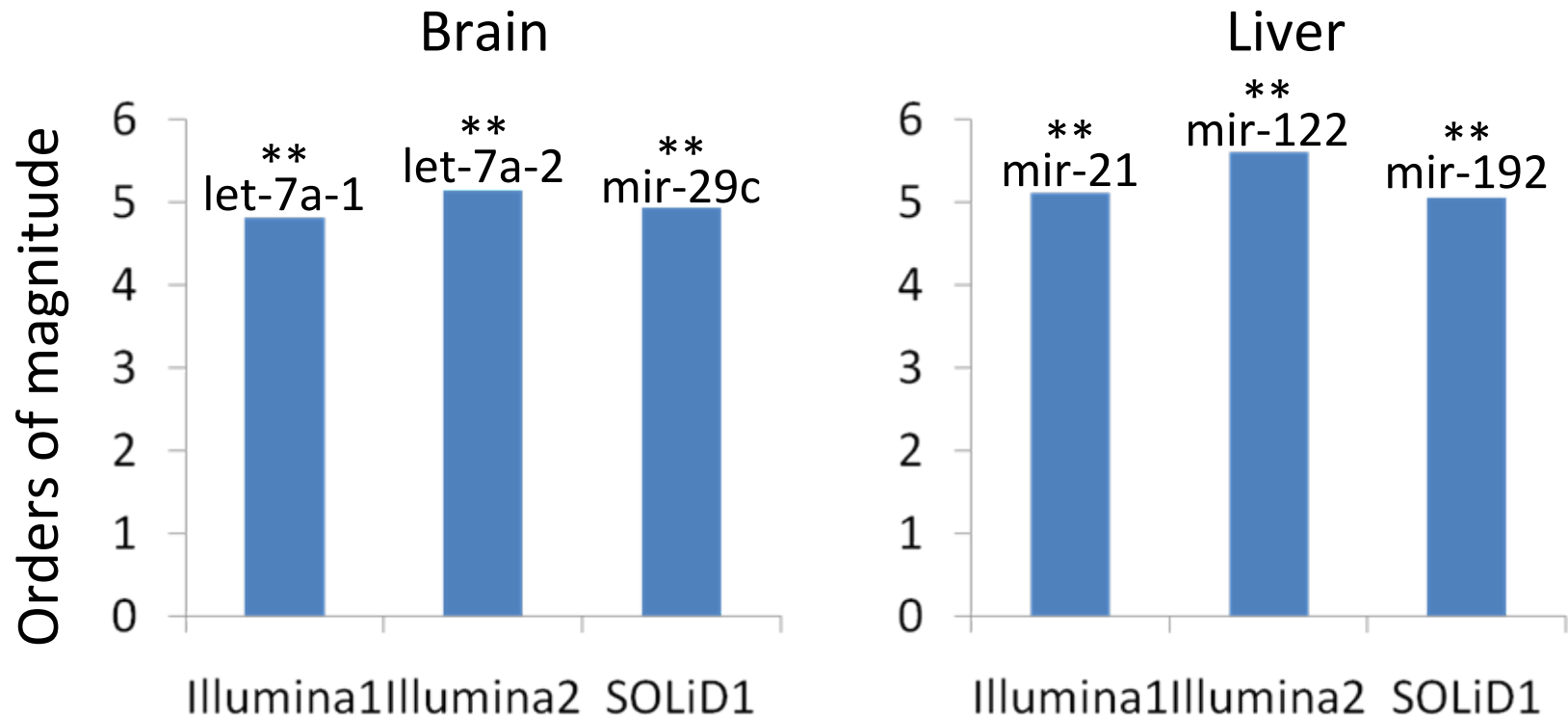
microRNAs Detected*



* >5 reads after normalization

** All = 705 different miRNAs

Dynamic Range of miRNA Reads*



* Magnitude of difference between highest and lowest expressed miRNA

** miRNA measured as highest expressed

Microarray & Next Gen Seq miRNA profiling summary

- All platforms tested are effective in detecting miRNA transcripts
- Intraplatform reproducibility generally high
- Differential expression detection among MA platforms was similar ($r=0.65$ to 0.84)
- Concordance between MA and next-gen seq (analysis in progress)
- Dynamic detection range much greater for next gen sequencing ($\sim 2-3 \log_{10}$ more than MA)

Summary: Assay requirements as implemented in MARG study

	ABI Taqman RT PCR	Affymetrix arrays	Agilent arrays	Exiqon LNA arrays	Illumina arrays	ABI SOLid seq	Illumina GA seq
major equipment:	7900HT PCR system	GeneChip scanner & fluidic stations	glass slide scanner	glass slide scanner	BeadArray station	Sequencer & sample prep instruments	Sequencer & cluster station
total RNA	500 ng	1 ug ^a	200ng ^a	200 ^a	200 ng	500 ng	1 ug/5 ug
time*	6 hrs	1.5 days	2 days	2 days	2 days	2 weeks	1 week
ease-of-use	XXX	XX(X)	XX	XX	XXX	X	X
cost per sample**	\$400	\$250	\$250	\$340/ 2 channel slide	\$200	~\$1300	~\$1000

*from total RNA to primary data

**reagents & supplies (including array); labor not included

^a current recommended minimum inputs ~ 100 ng

MARG project labs 2009 & 2010

- Don Baldwin – University of Pennsylvania
- Chris Harrington – Oregon Health & Science University
- Susan Hester – Environmental Protection Agency, NC
- Herbert Auer – Institute for Research in Biomedicine, Spain
- Wei Wang – Cornell University
- Nadereh Jafari – Northwestern University
- Steve Potter – Cincinnati Children's Hospital

DSRG project labs 2009

- Peter Schweitzer – Cornell University

Other MARG members 2010

- Nalini Raghavachari – NHLBI Genomics Core Facility, NIH
- Natalia Reyero – Jackson State University

Acknowledgements: We thank GeoSpiza and Partek for their assistance

ABRF MicroArray Research Group

microRNA Synthetic Reference Project

Outline

Latin Square design

Standard	Set 1	Set 2	Set 3	Set 4	Set 5 (majority)
A	0.5x	0.1x	0.01x	0.001x	1x
B	0.001x	0.5x	0.1x	0.01x	1x
C	0.01x	0.001x	0.5x	0.1x	1x
D	0.1x	0.01x	0.001x	0.5x	1x
E	1x	1x	1x	1x	1x

Index by sequence composition

Randomize Key	Name	CLUSTALW
1	hsa-miR-664*ACUGGCUAGGGAAAAUGAUUGGAU.....
2	hsa-miR-577UAGAUAAAAUAUUGGUACCTUG.....
3	hsa-miR-576-3pAAGAUGUGGAAAA.AUUGGAAUC.....
4	hsa-miR-208aAUAAGACGAGCAAAAAGCUUGU.....
5	hsa-miR-208bAUAAGACGAACAAAAGGUUUGU.....
6	hsa-miR-559UAAAG.UAAAUAUGCACCAAAA.....
7	hsa-miR-9*AUAAGCUAGAU...ACCGAAAGU.....
8	hsa-miR-142-5pCAUAAAAG.UAGAAA.GCACUACU.....
9	hsa-miR-335UCAAGAGCAAUAACGAAAAAUGU.....
10	hsa-miR-620AUGGAGAUAGAUUAGAAAU.....
11	hsa-miR-516a-5pUUCUCGAGGAAAGAAGCACUUC.....
12	hsa-miR-516bAUCUGGAGGUAAGAAGCACUUU.....
13	hsa-miR-1255aAGGAUGAGCAAAGAAAGUAGAUU.....
14	hsa-miR-1255bCGGAUGAGCAAAGAAAGUGGUU.....
15	hsa-miR-765UGGAGGAG.AAGGAAGGUGAUG.....
16	hsa-miR-483-5pAAGACGGGAGGAAAGAAGGGAG.....
17	hsa-miR-1UGGAAUGUAAAAGAAGUAUGUAU.....
▪	▪	▪
▪	▪	▪
▪	▪	▪

Choose set members

Supplemented with related miRNAs

Research
Randomizer: 4 sets of
10 unique numbers,
range 1-878

Set 1	Set 2	Set 3	Set 4
3	37	5	26
175	132	46	104
193	135	214	246
200	240	237	296
259	245	310	473
531	292	409	485
652	311	414	790
809	594	496	797
830	680	629	846
877	843	665	859

Decimal key	Name	Sequence	Set
121.1	hsa-miR-130bCAGUGCAAUGAUGAAAGGGCAU.....	1
121.2	hsa-miR-130aCAGUGCAAUGUUAAAAGGGCAU.....	2
121.3	hsa-miR-301bCAGUGCAAUGAUUUGUCAAAAGC.....	3
121.4	hsa-miR-301aCAGUGCAAUGAUUUGUCAAAAGC.....	4
133.1	hsa-miR-376aAUCAUAGAGGAAAUCCACGU.....	1
133.2	hsa-miR-376cAACAUAGAGGAAAUCCACGU.....	2
304.1	hsa-miR-181dAACAUUCAUUGUUGUCGGUGGGU.....	3
304.2	hsa-miR-181bAACAUUCAUUGCUGUCGGUGGGU.....	4
309.1	hsa-let-7fUGAGGUAGUAGAUUGUAUAGUU.....	1
309.2	hsa-let-7aUGAGGUAGUAGGUUGUAUAGUU.....	2
309.3	hsa-let-7eUGAGGUAGGAGGUUGUAUAGUU.....	3
309.4	hsa-let-7bUGAGGUAGUAGGUUGUGUGUU.....	4
309.5	hsa-let-7cUGAGGUAGUAGGUUGUAUGUU.....	1
309.6	hsa-miR-98UGAGGUAGUAAGUUGUAUUGUU.....	2
309.7	hsa-let-7dAGAGGUAGUAGGUUGCAUAGUU.....	3
309.8	hsa-let-7gUGAGGUAGUAGUUUGUACAGUU.....	4
309.9	hsa-let-7iUGAGGUAGUAGUUUGUCUGUU.....	1
348.1	hsa-miR-520bAAAGUGCUUCC..UUUUAGAGGG.....	2
348.2	hsa-miR-520c-3pAAAGUGCUUCC..UUUUAGAGGGU.....	3
348.3	hsa-miR-520fAAGUGCUUCC..UUUUAGAGGGUU.....	4
860.1	hsa-miR-15bUAGCAGCAUCAUGGUUUACA.....	1
860.2	hsa-miR-15aUAGCAGCACAAUUGGUUUUGUG.....	2
871.1	hsa-miR-23aAUCACAUUGCCAGGGAUUUCC.....	3
871.2	hsa-miR-23bAUCACAUUGCCAGGGAUUUACC.....	4

Synthetic reference applications

