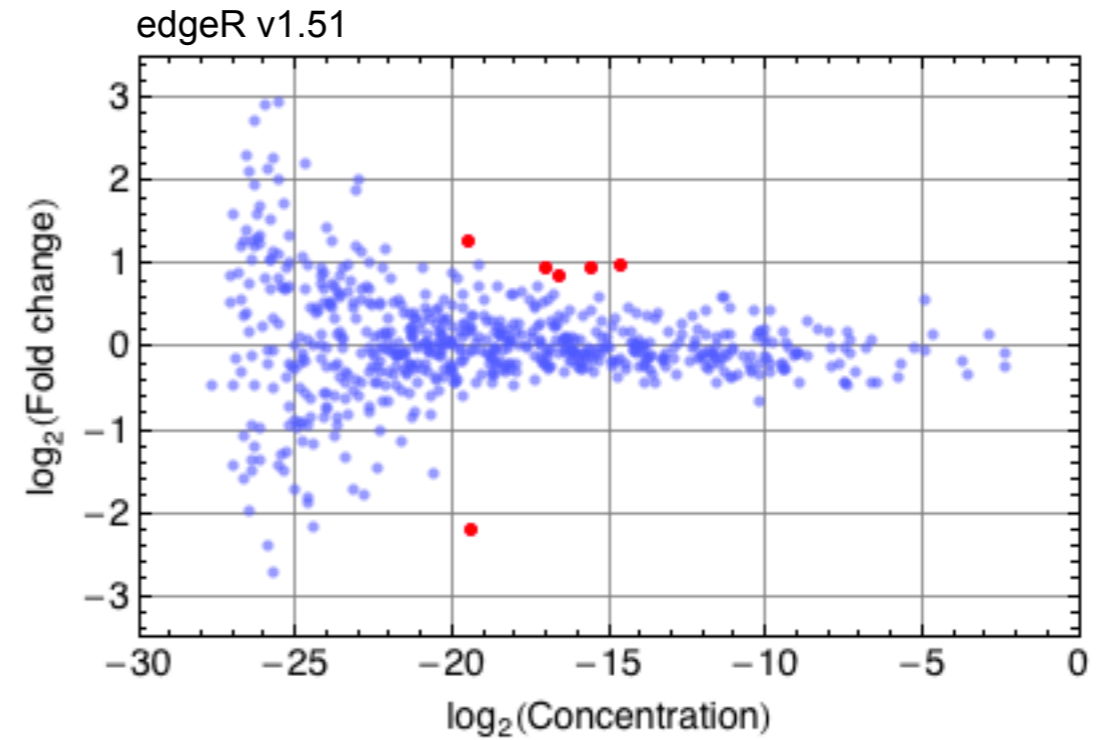
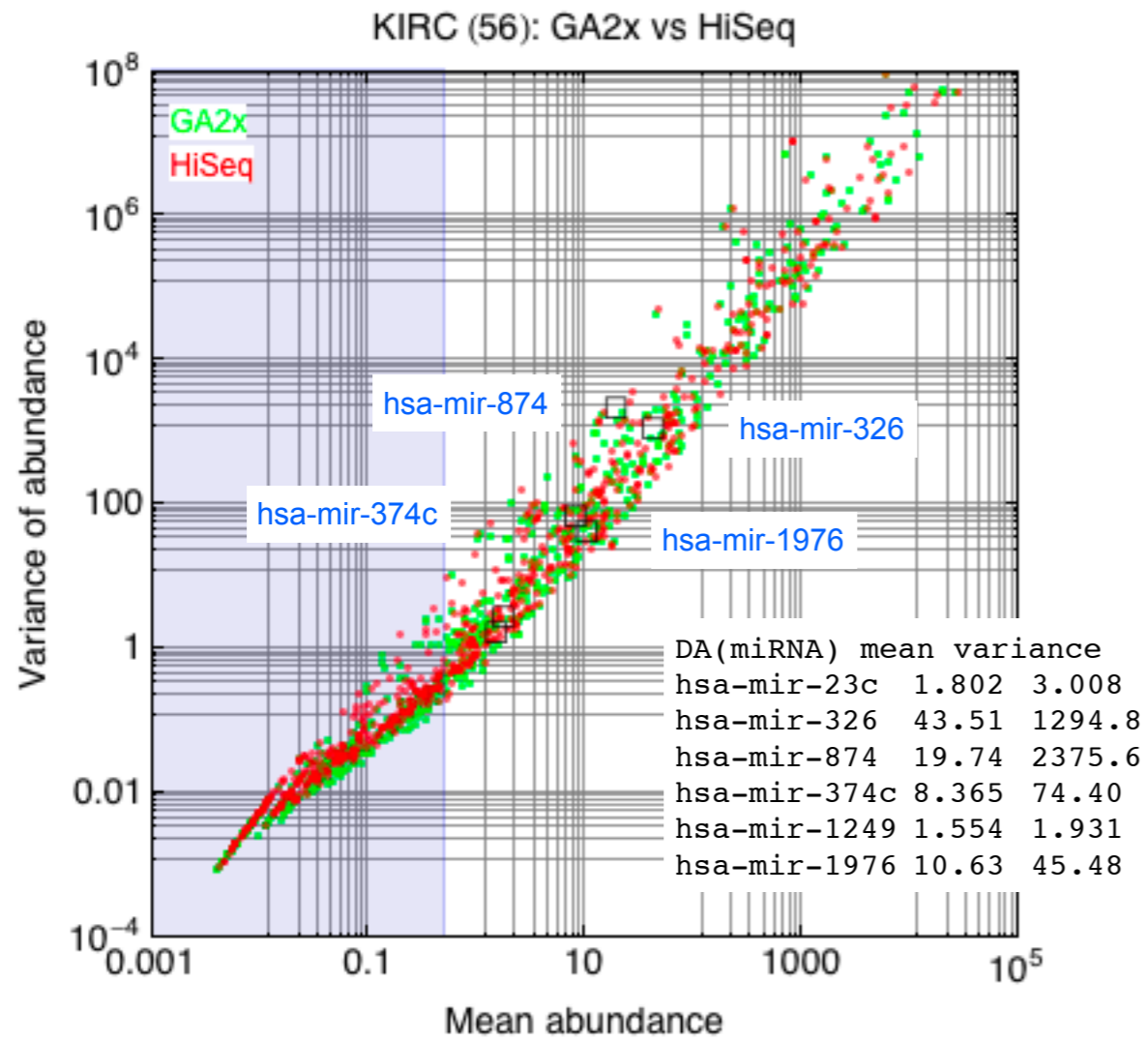


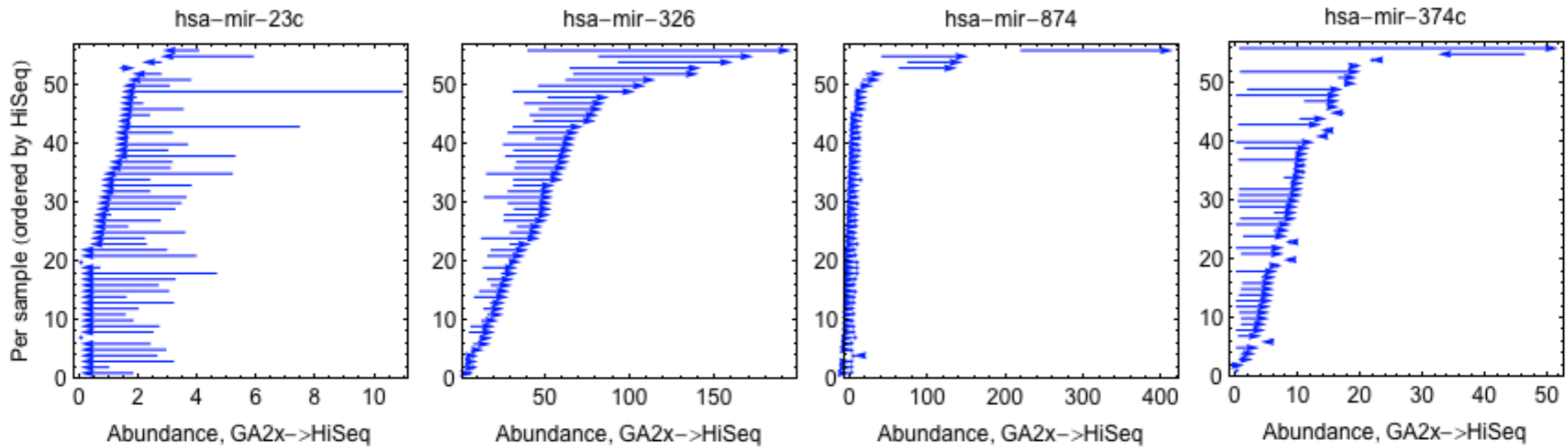
Differentially abundant KIRC/P miRNAs between GA2x and HiSeq

G. Robertson, Andy Chu, Richard Corbett, Samuel Wu
27 May 2011, 21h30

1. Adjusted data: GA2x vs. subsampled HiSeq.
2. Original data: GA2x vs. non-subsampled HiSeq
3. GC content of miRNAs
4. PCA driver miRNAs and GC content



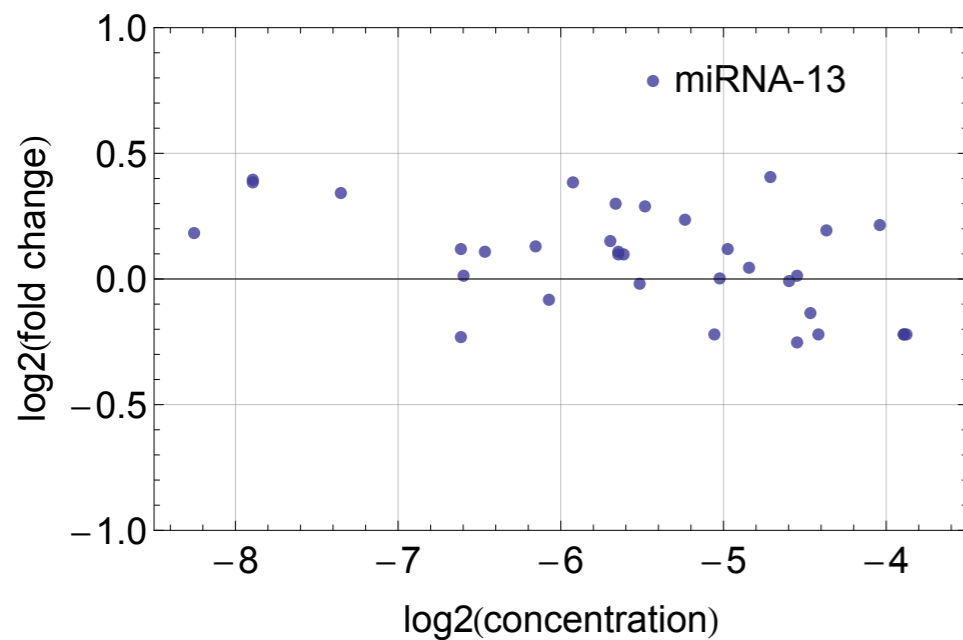
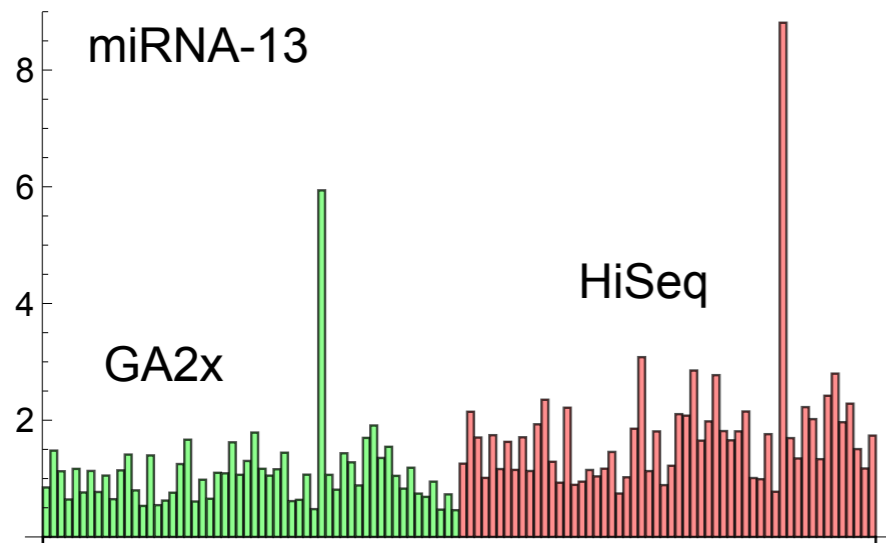
logConc	logFC	P.Value	adj.P.Val
hsa-mir-23c	-19.43	-2.20	8.01e-12 5.91e-09
hsa-mir-326	-14.64	0.991	2.39e-06 0.00088
hsa-mir-874	-15.61	0.961	9.25e-06 0.0023
hsa-mir-374c	-17.01	0.963	2.23e-05 0.00411
hsa-mir-1249	-19.48	1.30	4.74e-05 0.0070
hsa-mir-1976	-16.60	0.854	0.00014 0.017
...			



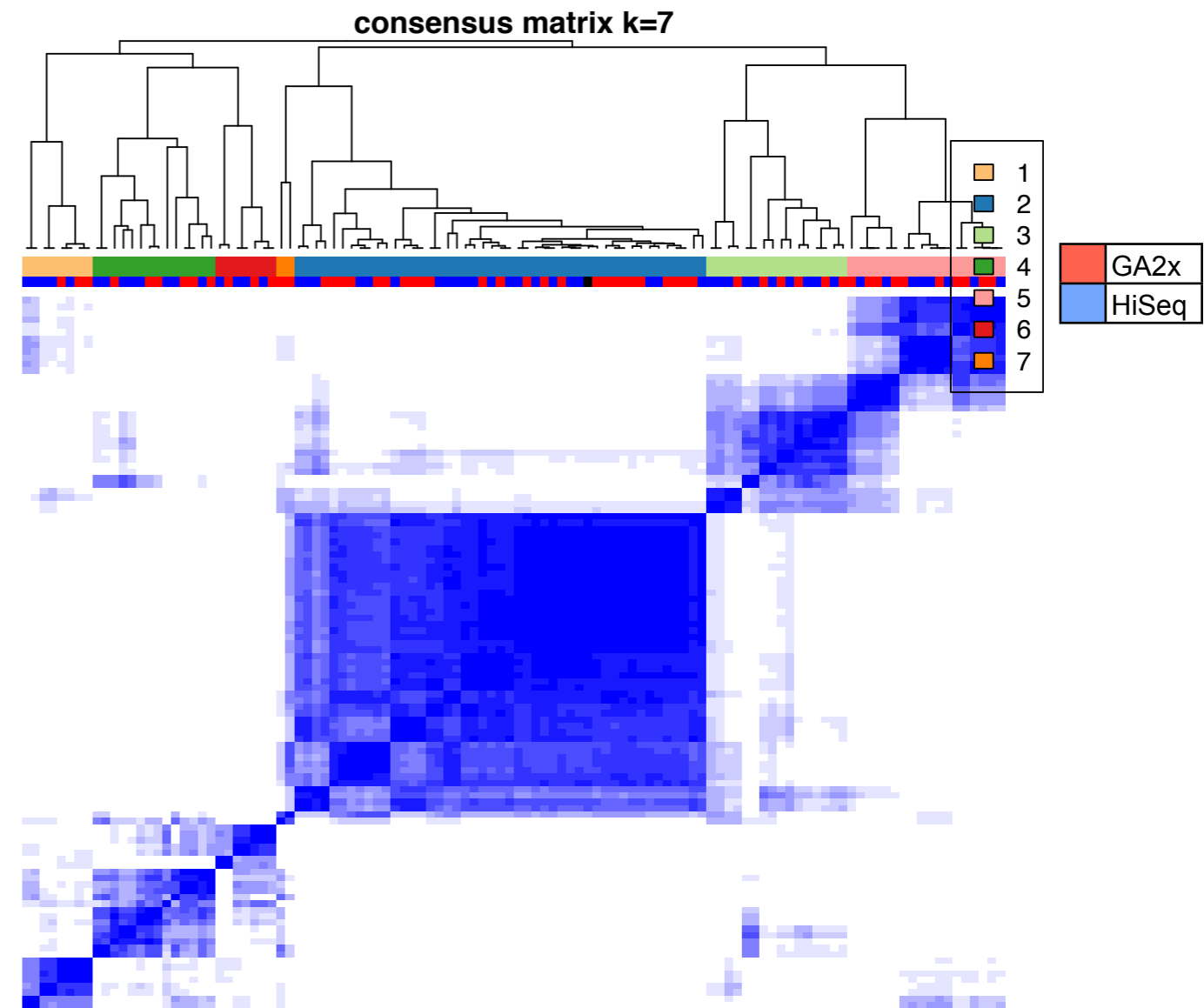
KIRC/P GA2x vs HiSeq, subsample 5, 35 miRNAs

Differential abundance analysis, edgeR v1.5.1
 A larger set of miRNAs should estimate background variance more effectively.

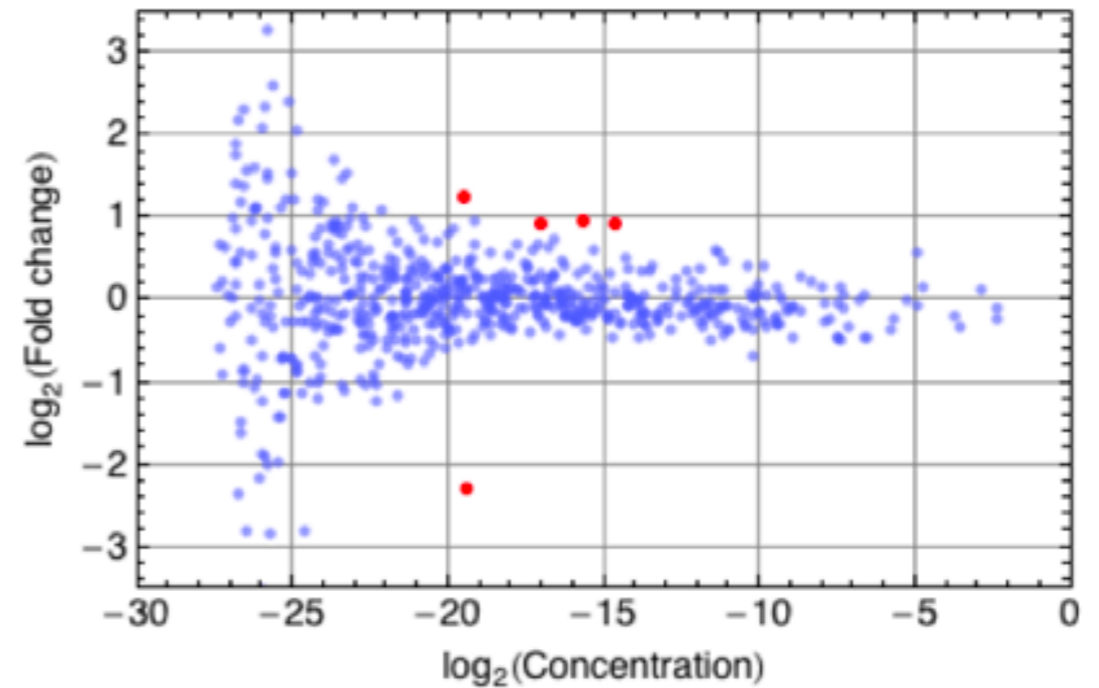
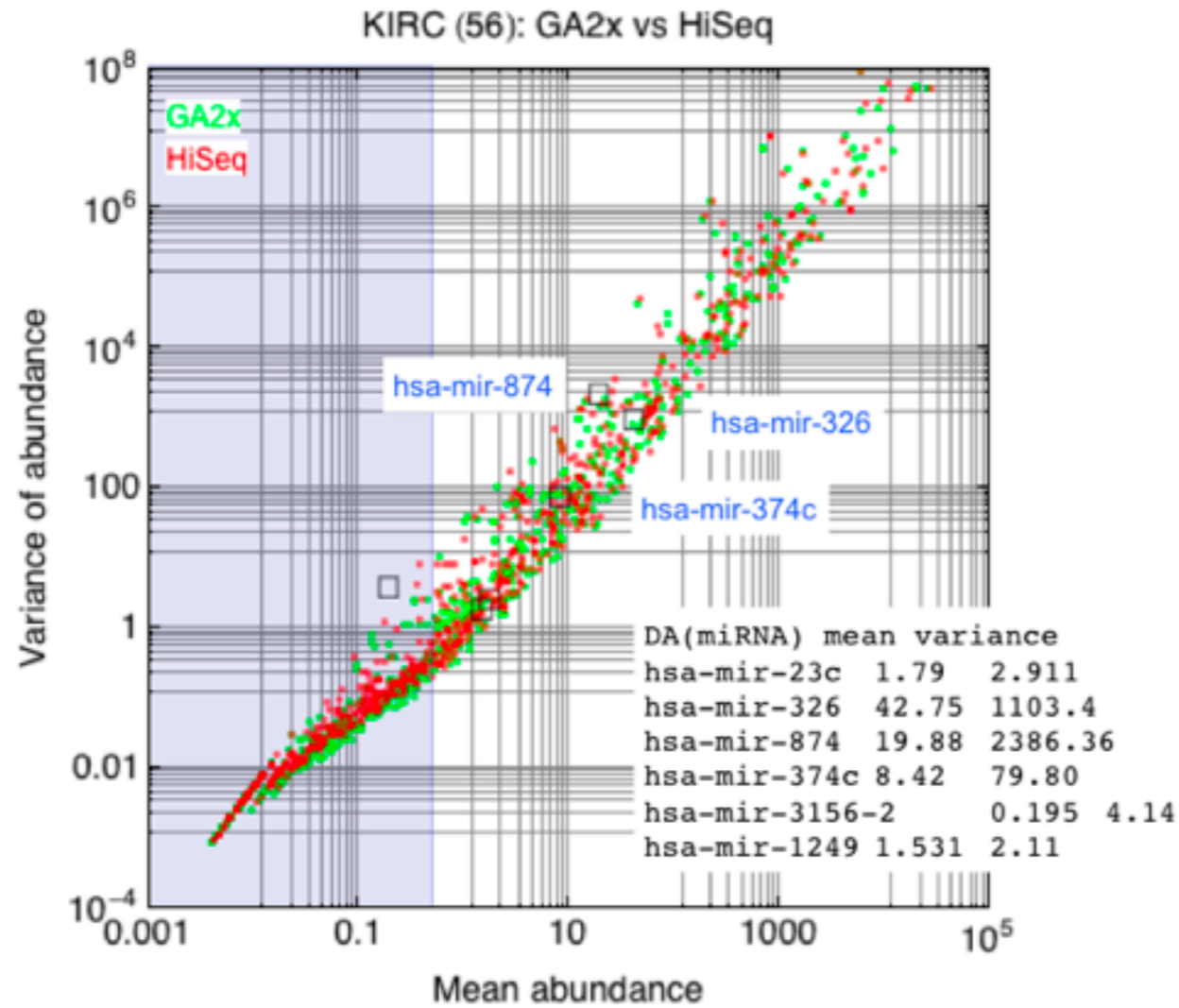
Consensus clustering, Spearman, k-means, 200 iterations
 The two platforms are mixed, i.e. do not clearly separate.



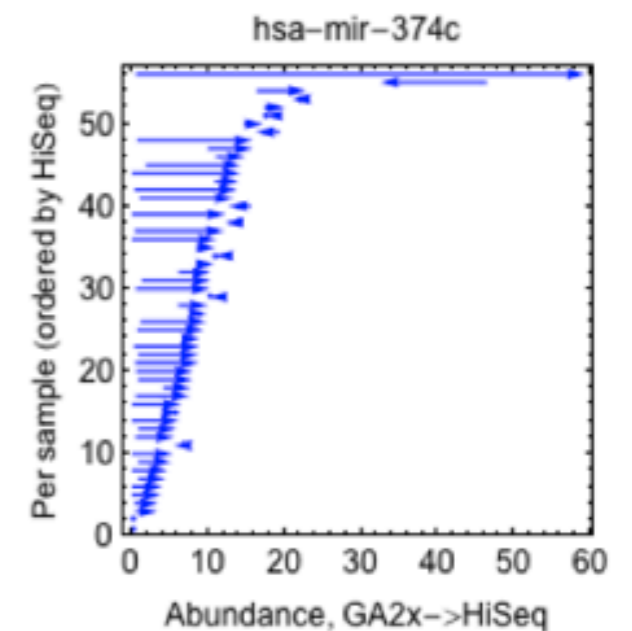
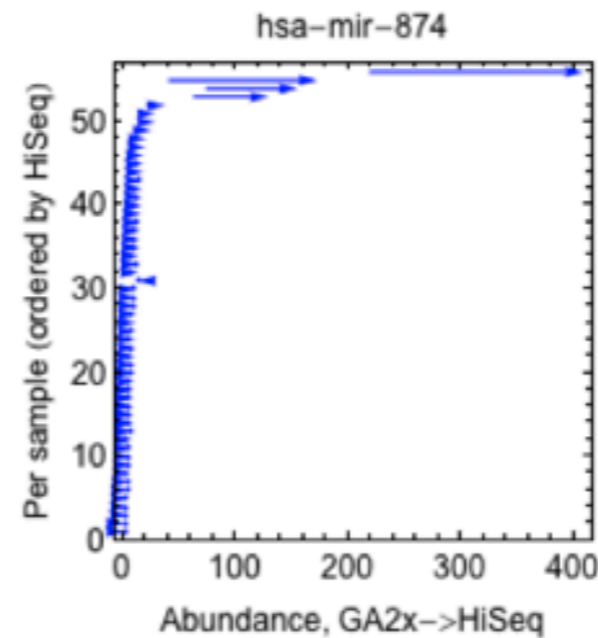
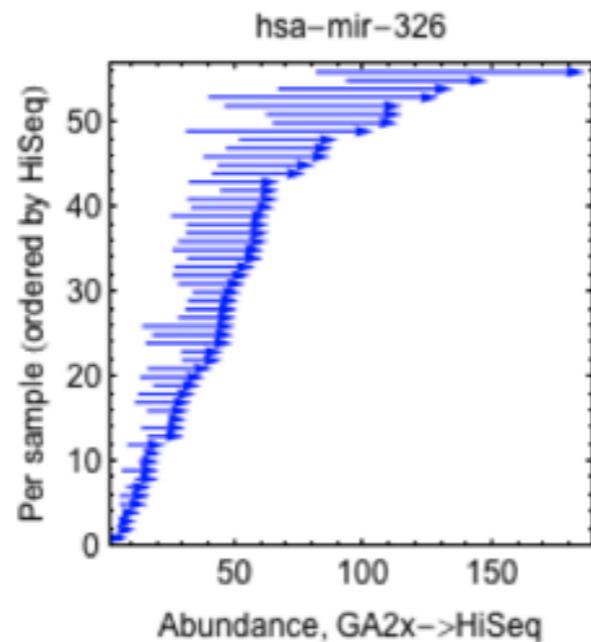
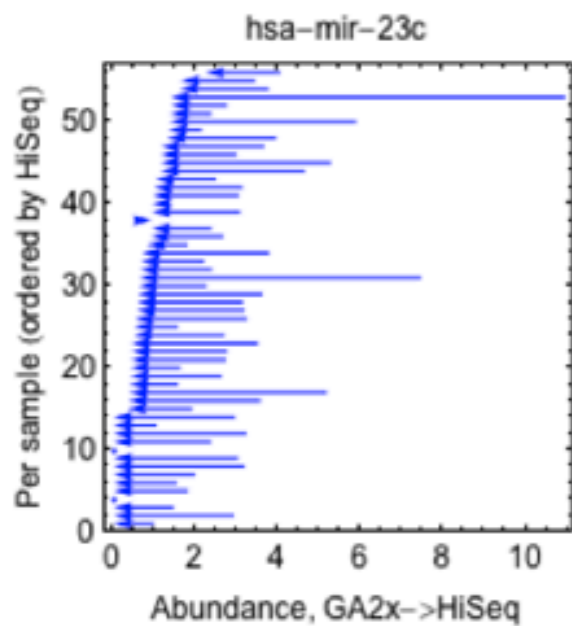
gene	logConc	logFC	P.Value	adj.P.Val
mi-13	-5.4255	0.783499	0.000945101	0.0330785
mi-23	-4.70622	0.406441	0.0291534	0.510185
mi-1	-3.89152	-0.221841	0.120596	0.751885
mi-2	-3.89345	-0.220216	0.120596	0.751885
mi-3	-3.87952	-0.220564	0.122239	0.751885
mi-34	-4.04362	0.211961	0.152072	0.751885
...				



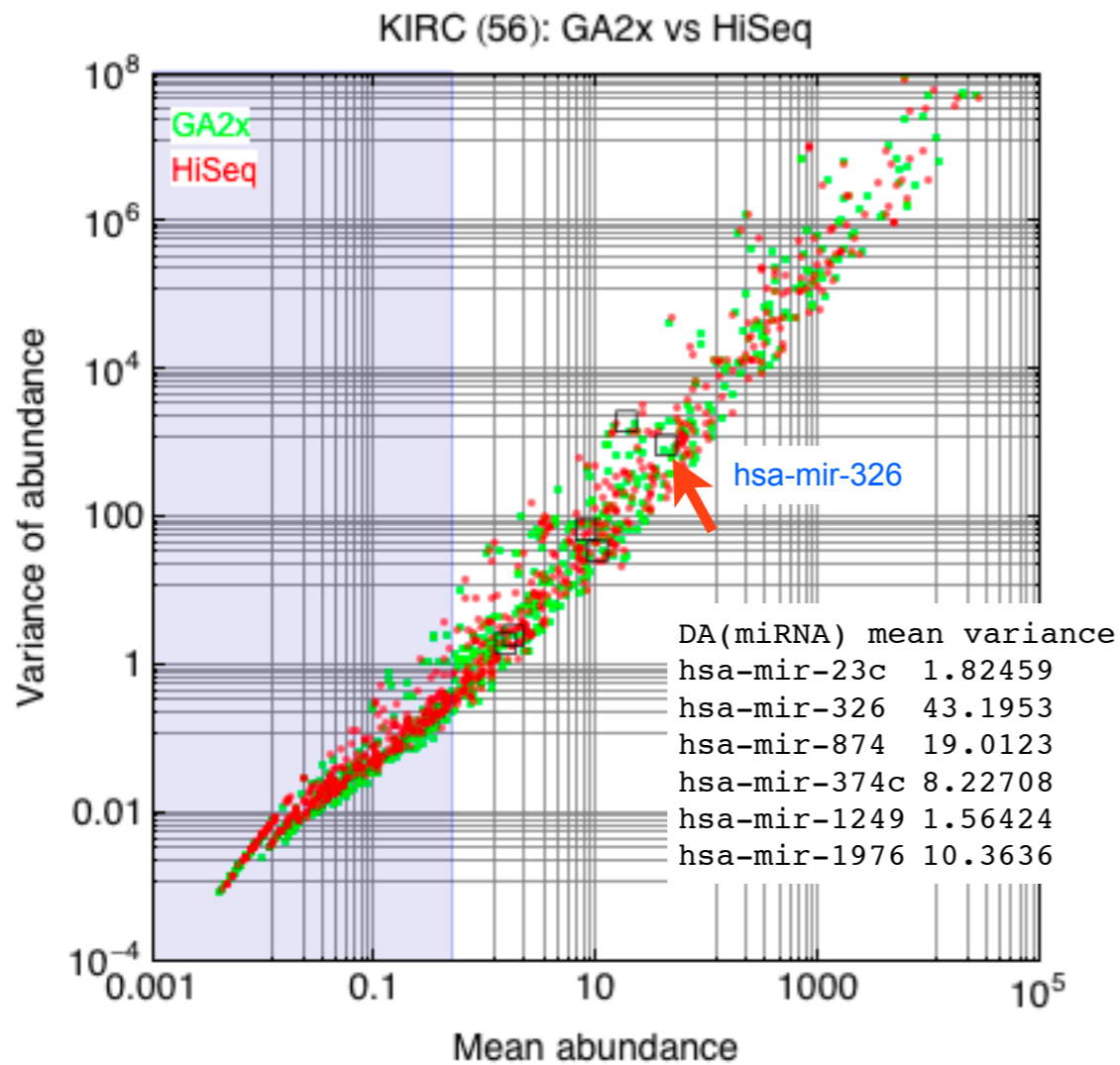
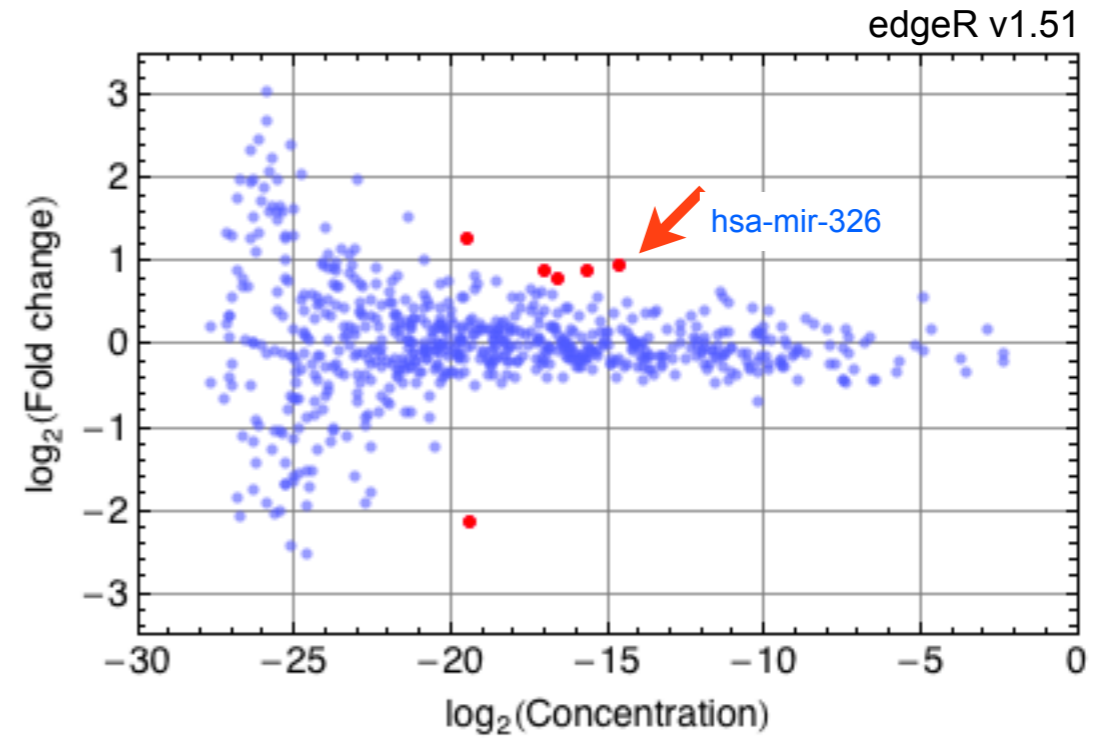
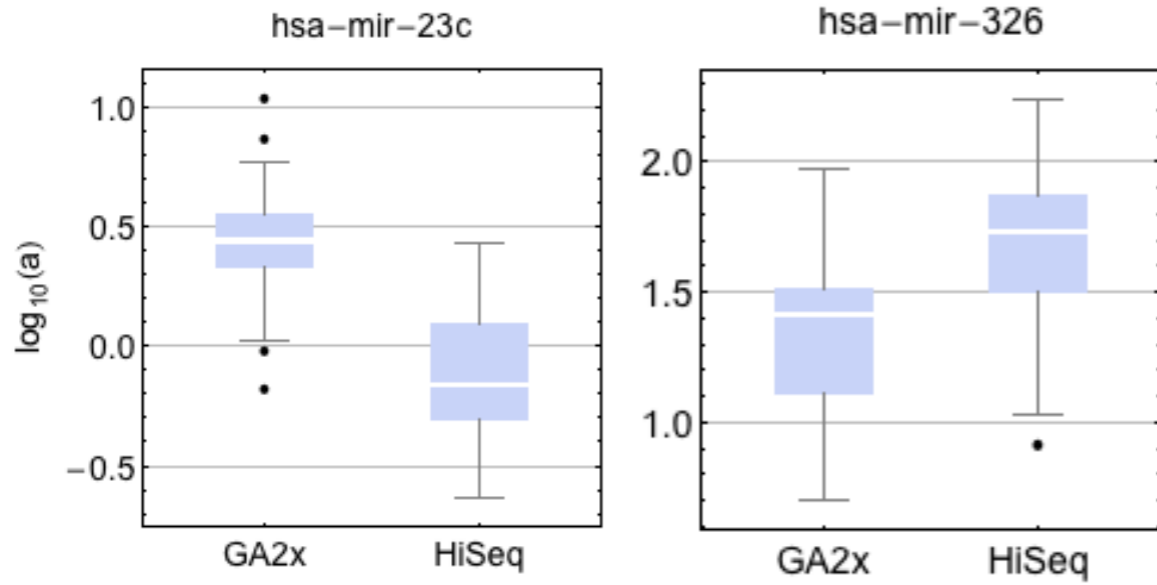
Samuel Wu



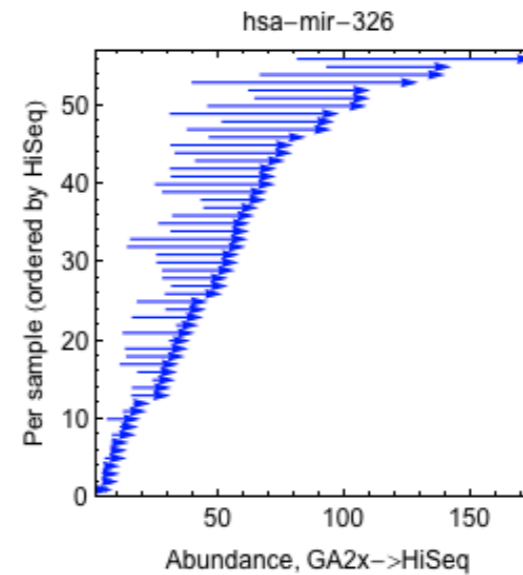
logConc	logFC	P.Value	adj.P.Val
hsa-mir-23c	-19.46	-2.278	2.19e-12 1.604e-09
hsa-mir-326	-14.65	0.944	6.76e-06 0.00194
hsa-mir-874	-15.65	0.964	7.95e-06 0.0019
hsa-mir-374c	-17.00	0.939	3.33e-05 0.0061
hsa-mir-3156-2	-24.36	5.884	4.54e-05 0.0066
hsa-mir-1249	-19.49	1.247	8.14e-05 0.0099
...			



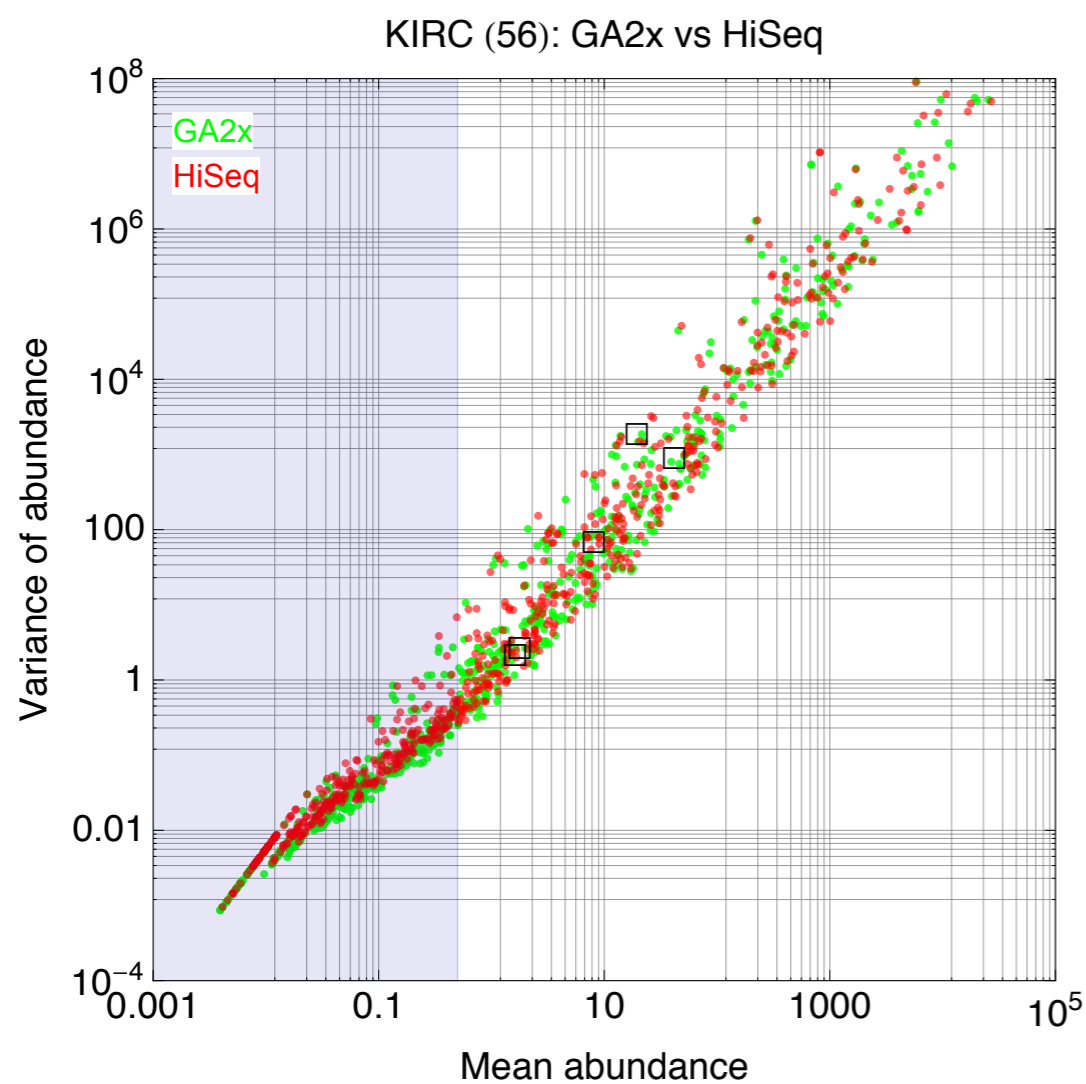
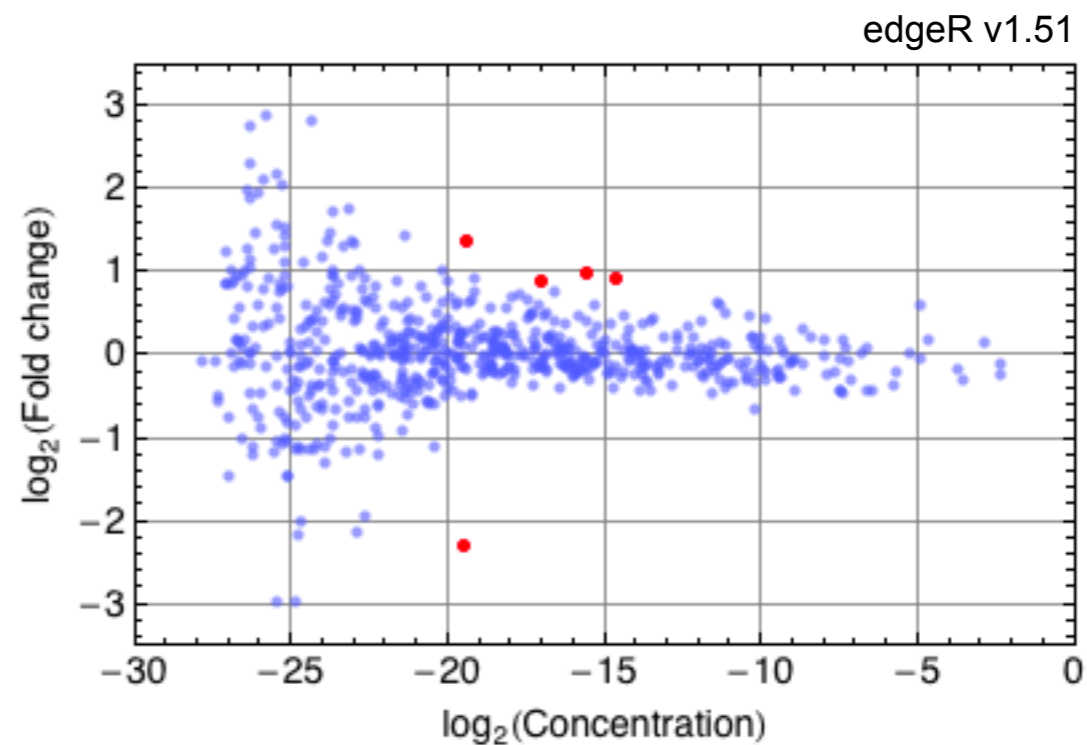
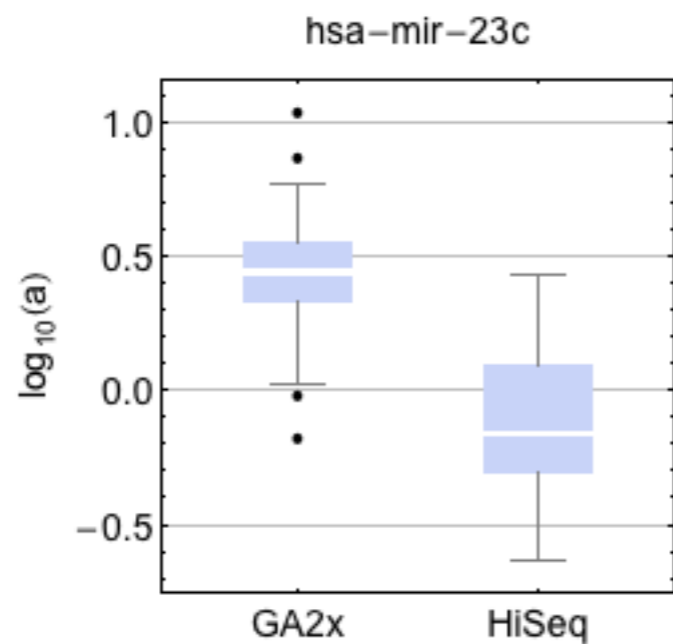
KIRC/P GA2x vs HiSeq, subsample 2, 735 miRNAs



gene	logConc	logFC	P.Value	adj.P.Val
hsa-mir-23c	-19.40	-2.133	1.69e-11	1.24e-08
hsa-mir-326	-14.65	0.975	3.10e-06	0.00114
hsa-mir-874	-15.701	0.906	2.55e-05	0.00625
hsa-mir-374c	-17.033	0.911	5.31e-05	0.00813
hsa-mir-1249	-19.485	1.280	5.53e-05	0.00813
hsa-mir-1976	-16.611	0.797	0.00037	0.0450
...				



KIRC/P GA2x vs HiSeq, subsample 1, 735 miRNAs

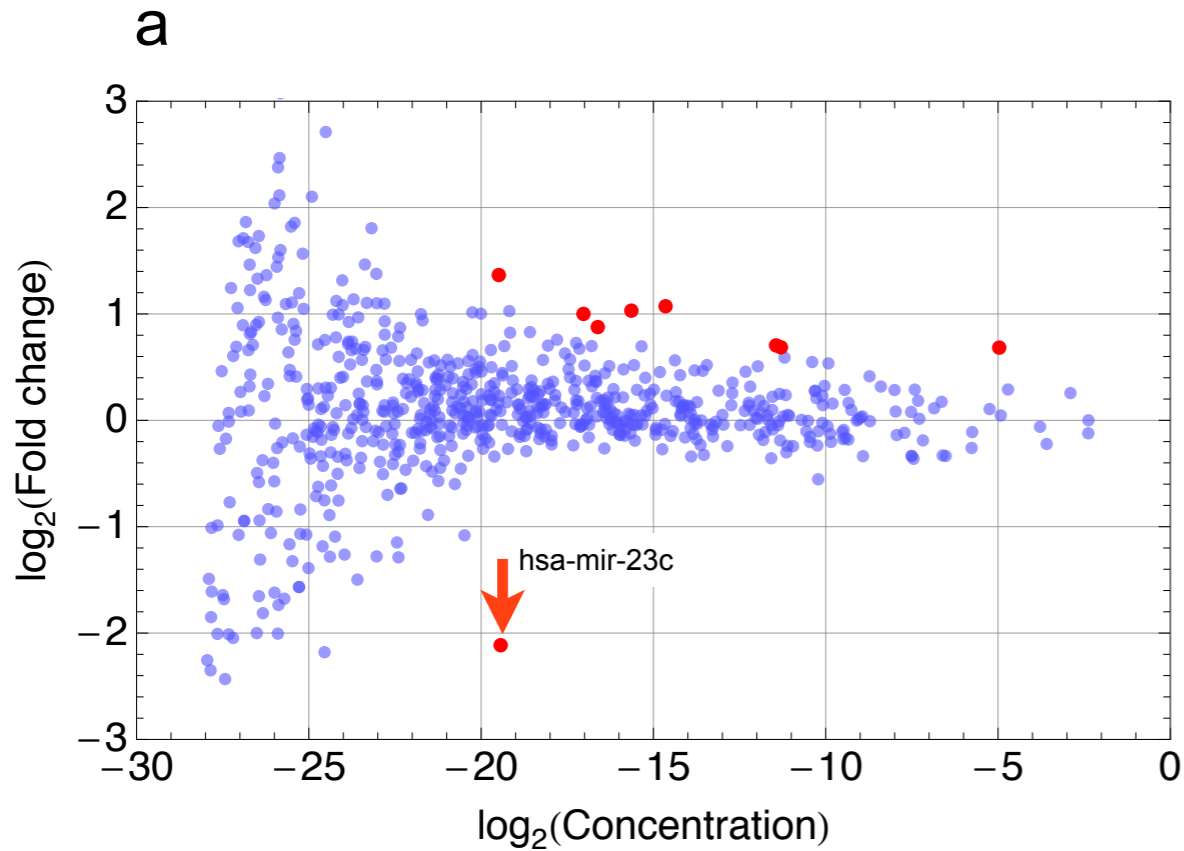
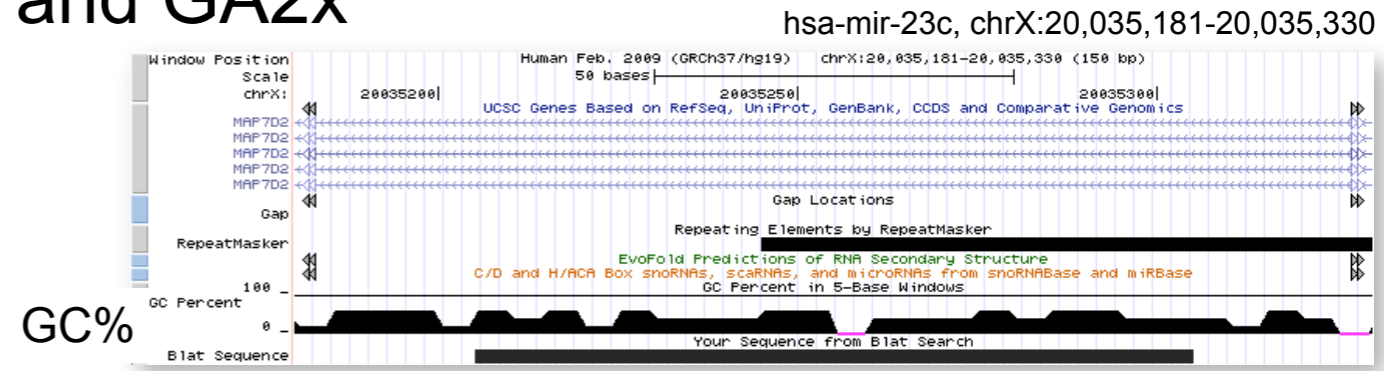


gene	logConc	logFC	P.Value	adj.P.Val
hsa-mir-23c	-19.477	-2.29	2.88e-12	2.12e-09
hsa-mir-874	-15.620	0.984	5.62e-06	0.0021
hsa-mir-326	-14.679	0.923	1.13e-05	0.0024
hsa-mir-1249	-19.440	1.376	1.32e-05	0.0024
hsa-mir-374c	-17.047	0.884	0.000101	0.0148

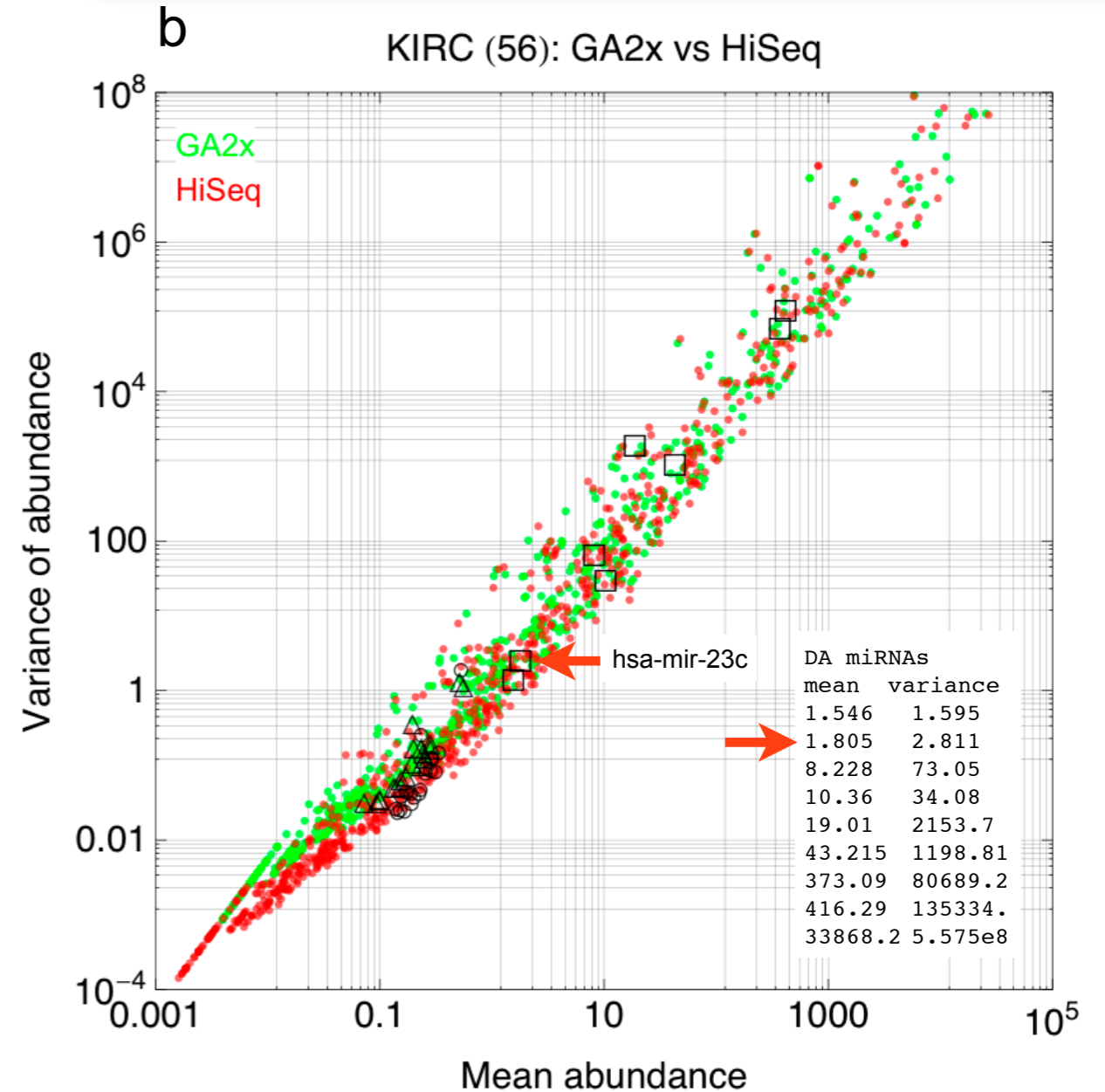
1. Adjusted data: GA2x vs. subsampled HiSeq.
2. Original data: GA2x vs. non-sampled HiSeq
3. GC content of miRNAs
4. PCA driver miRNAs and GC content

Differential abundance between HiSeq and GA2x

We identified differentially abundant miRNAs between HiSeq and GA2x (1/0) using edgeR v2.2.3 and an expression matrix of 767 miRNAs that had nonzero sums across 2x56 samples. Nine miRNAs had BH-corrected p-val < 0.05 (red in 'a', black squares in 'b', table). As expected, the FC plot ('smearPlot') identified no low-abundance miRNAs as differentially abundant. **hsa-mir-23c** was flagged as having a negative PC1 loading (slide 9), when all other 8 loadings had positive values.



gene	logConc	logFC	P.Value	adj.P.Val
→ hsa-mir-23c	-19.4337	-2.114	2.88E-11	2.21E-08
hsa-mir-326	-14.6473	1.072	1.52E-07	0.0000583
hsa-mir-874	-15.6407	1.031	1.01E-06	0.000257
hsa-mir-374c	-17.0304	1.000	6.31E-06	0.00121
hsa-mir-1249	-19.488	1.366	0.0000124	0.00190
hsa-mir-1976	-16.6154	0.8766	0.0000653	0.00833
hsa-mir-204	-11.4417	0.7045	0.000394	0.0431
hsa-mir-150	-11.2998	0.6859	0.000556	0.0481
hsa-mir-99b	-4.96516	0.6833	0.000565	0.0481

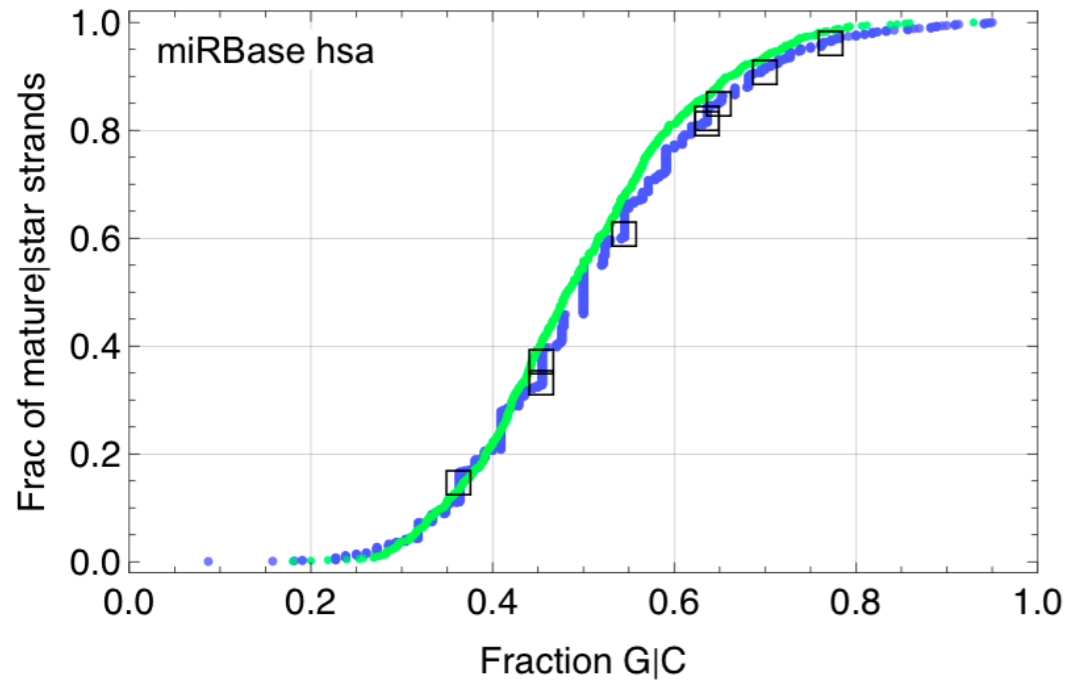


1. Open black circles/triangles show the 20 PCA 'driver' miRNAs. Squares show nine differentially abundant miRNAs for BH-pval<0.05.
2. Red and green point clouds appear to separate for mean abundances below ~2.

1. Adjusted data: GA2x vs. subsampled HiSeq.
2. Original data: GA2x vs. non-sampled HiSeq
- 3. GC content of miRNAs**
4. PCA driver miRNAs and GC content

GC content of differentially abundant miRNA mature|star strands

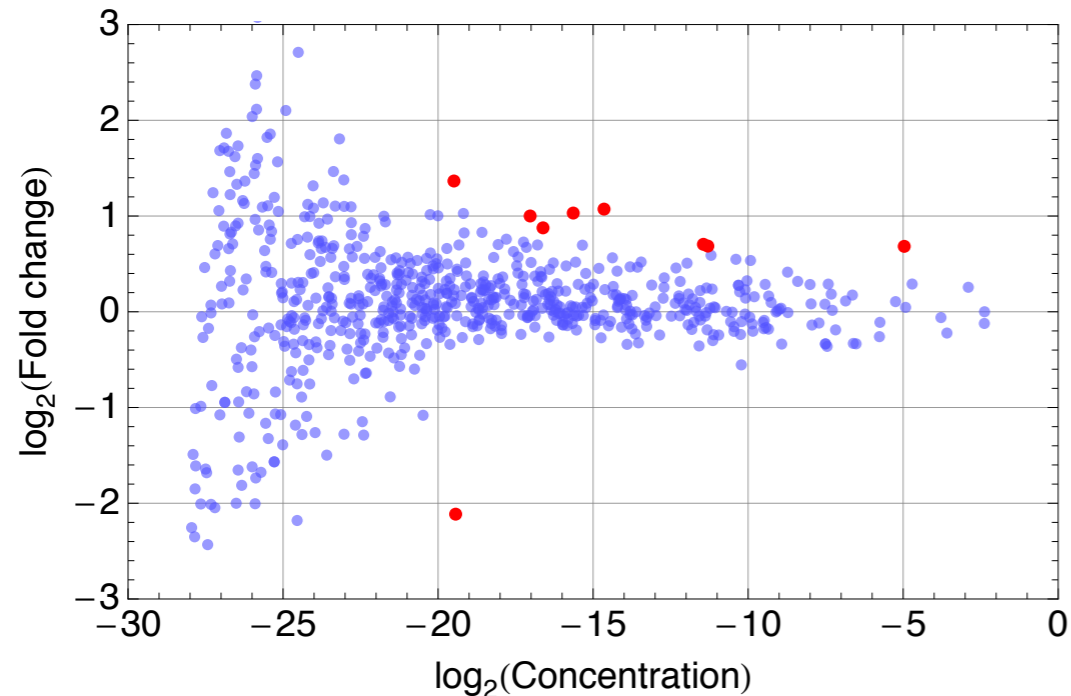
GC content of 1733 mature or star strands (v17, blue) and 939 primary transcripts (v15, green)



miRNA	MIMAT	sequence	f(GC)
hsa-miR-874	MIMAT0004911	CUGCCCUGGCCCGAGGGACCGA	0.773
hsa-miR-326	MIMAT0000756	CCUCUGGGCCCUUCCUCCAG	0.700
hsa-miR-1976	MIMAT0009451	CCUCCUGCCCUCCUUGCUGU	0.650
hsa-miR-1249	MIMAT0005901	ACGCCCUUCCCCCCCUCUUCA	0.636
hsa-miR-99b	MIMAT0000689	CACCCGUAGAACCGACCUUGCG	0.636
hsa-miR-150	MIMAT0000451	UCUCCCAACCCUUGUACCAGUG	0.545
hsa-miR-23c	MIMAT0018000	AUCACAUUGCCAGUGAUUACCC	0.454
hsa-miR-204	MIMAT0000265	UUCCCUUUGUCAUCCUAUGCCU	0.454
hsa-miR-374c	MIMAT0018443	AUAAUACAACCUGCUAAGUGCU	0.364

/Users/grobertson/GENEREG/Data/miRNA/mirBase/v16/mature.hsa.fraction-GC.txt

Nine differentially abundant miRNAs (BH-pval<0.05, red) for 767 miRNAs (blue) in GA2x-HiSeq KIRC/KIRP test



gene	logConc	logFC	P.Value	adj.P.Val
hsa-mir-23c	-19.43	-2.114	2.88E-11	2.21E-08
hsa-mir-326	-14.64	1.072	1.52E-07	0.0000583
hsa-mir-874	-15.640	1.031	1.01E-06	0.000257
hsa-mir-374c	-17.030	1.000	6.31E-06	0.00121
hsa-mir-1249	-19.488	1.366	0.0000124	0.00190
hsa-mir-1976	-16.615	0.877	0.0000653	0.00833
hsa-mir-204	-11.442	0.705	0.000394	0.0431
hsa-mir-150	-11.30	0.686	0.000556	0.0481
hsa-mir-99b	-4.965	0.683	0.000565	0.0481

1. Adjusted data: GA2x vs. subsampled HiSeq.
2. Original data: GA2x vs. non-subsampled HiSeq
3. GC content of miRNAs
4. **PCA driver miRNAs and GC content**

Hi Gordon.

Using all of the tests on the subsampled data we found just one result that was significant. Anecdotally, most tests showed a difference between the two platforms, but only this one was able to differentiate them statistically. In this figure, the GA is green and the High seq is red. The median of the mean normalized expression for this set of microRNAs is ~5000, so we're not talking about noise here.

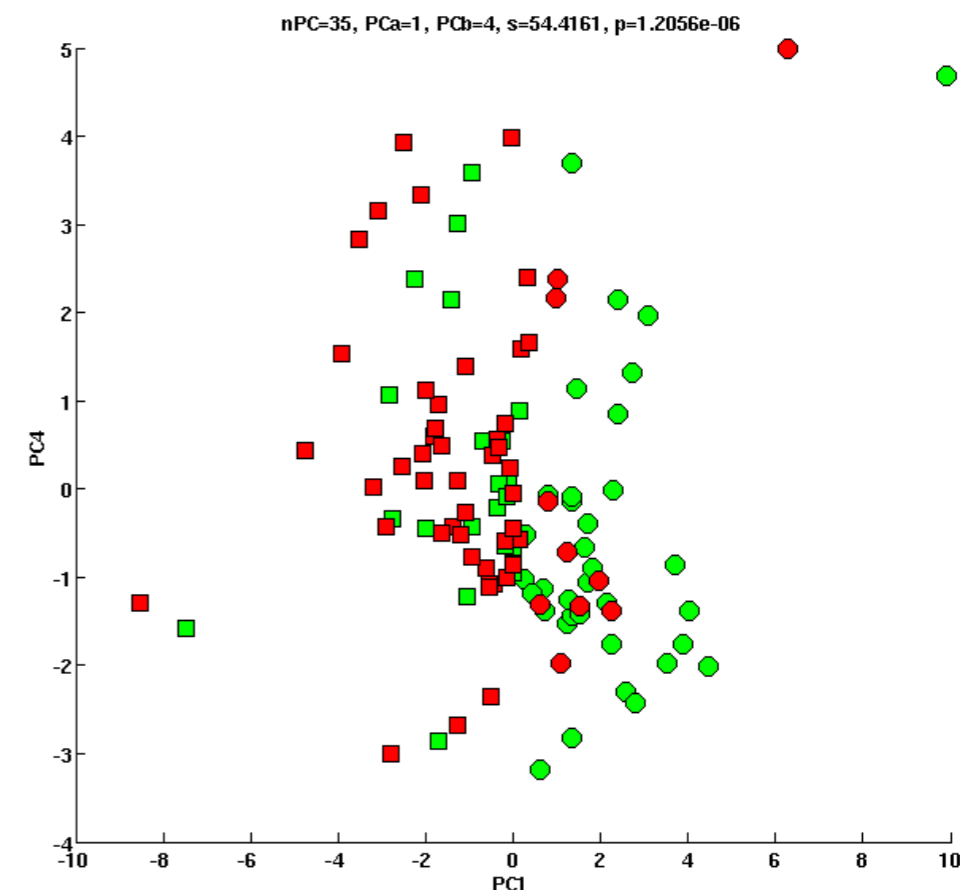
Richard

Here are the top 20 driving microRNAs:

hsa-let-7a-3
 hsa-let-7a-2
 hsa-let-7a-1
 hsa-let-7f-2
 hsa-mir-101-1
 hsa-mir-148a
 hsa-mir-144
 hsa-mir-30d
 hsa-mir-199b
 hsa-mir-451
 hsa-mir-199a-2
 hsa-mir-30a
 hsa-mir-99b
 hsa-let-7b
 hsa-mir-100
 hsa-mir-21
 hsa-mir-29a
 hsa-mir-9-1
 hsa-mir-9-2
 hsa-mir-25

GC content

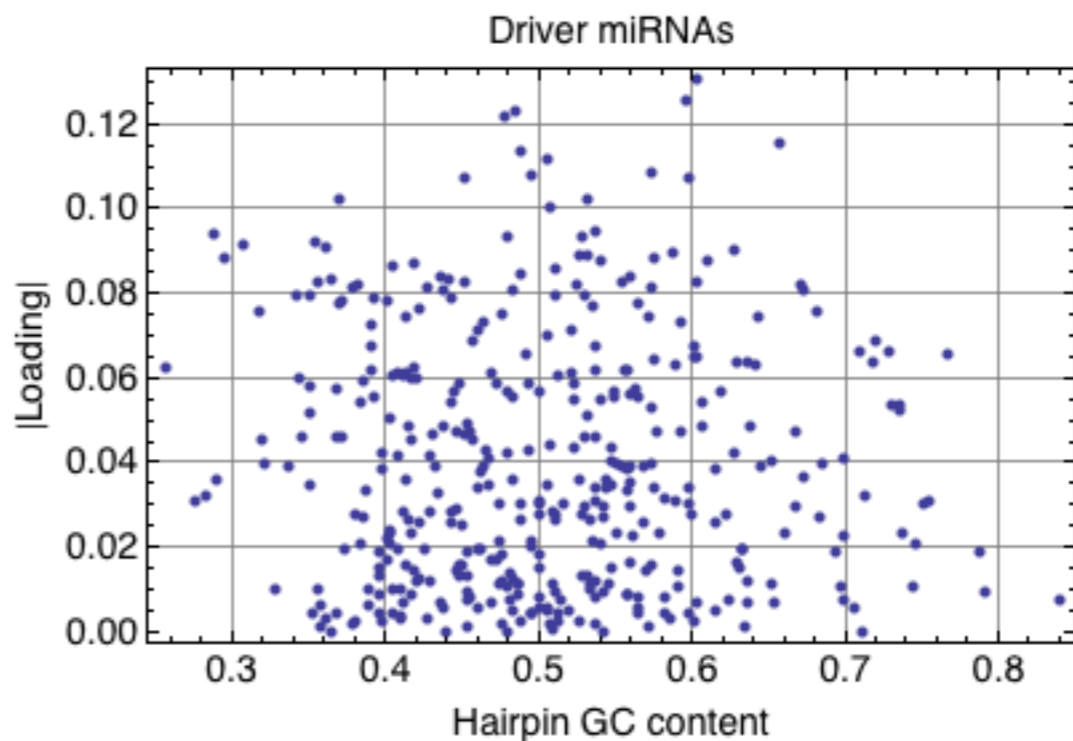
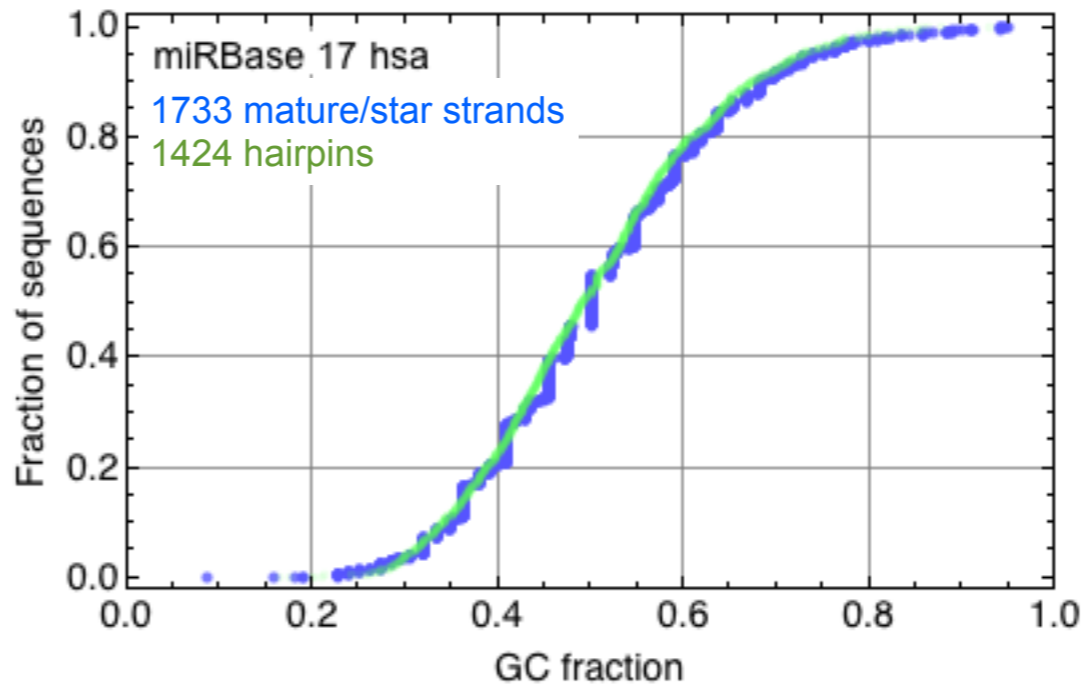
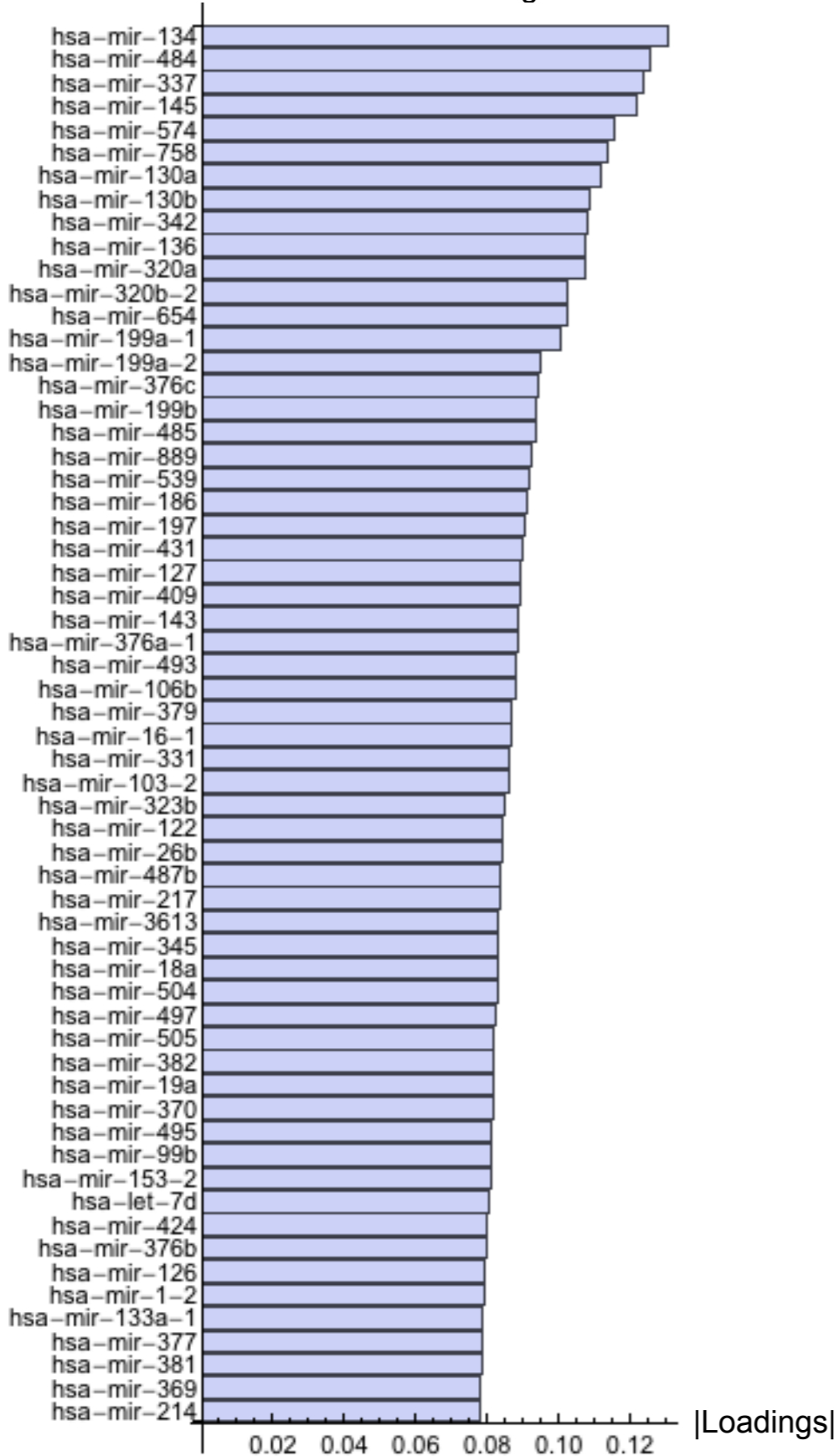
hsa-let-7a	MIMAT0000062	UGAGGUAGUAGGUUGUAUAGUU	0.363636
hsa-let-7f	MIMAT0000067	UGAGGUAGUAGAUUGUAUAGUU	0.318182
hsa-miR-101	MIMAT0000099	UACAGUACUGUGUAACUGAA	0.333333
hsa-miR-9	MIMAT0000441	UCUUUGGUUAUCUAGCUGUAUGA	0.347826
hsa-miR-148a	MIMAT0000243	UCAGUGCACUACAGAACUUUGU	0.409091
hsa-miR-144	MIMAT0000436	UACAGUAUAGAUGAUGUACU	0.3
hsa-miR-30d	MIMAT0000245	UGUAAACAUCCCGACUGGAAG	0.5
199b			
hsa-miR-451	MIMAT0001631	AAACCGUUACCAUACUGAGUU	0.363636
199a			
hsa-miR-30a	MIMAT0000087	UGUAAACAUCCUCGACUGGAAG	0.454545
hsa-miR-99b	MIMAT0000689	CACCCGUAGAACCGACCUUGCG	0.636364
hsa-let-7b	MIMAT0000063	UGAGGUAGUAGGUUGUGUGGUU	0.454545
hsa-miR-100	MIMAT0000098	AACCCGUAGAUCCGAACUUGUG	0.5
hsa-miR-21	MIMAT0000076	UAGCUUAUCAGACUGAUGUUGA	0.363636
hsa-miR-29a	MIMAT0000086	UAGCACCAUCUGAAAUCGGUUA	0.409091
-9			
hsa-miR-25	MIMAT0000081	CAUUGCACUUGUCUCGGUCUGA	0.5



27 May 2011, 18h00

400 PCA-driver miRNAs (from Richard Corbett, later on Friday)

50/400 driver loadings



397/400 drivers