Making the Most of Small Samples: Developing a Benchmark to Compare Amplification Strategies

M. Sharma, Z. Liu, C. Virtanen, N. Winegarden. University Health Network Microarray Centre

Abstract

Gene expression profiling by microarrays has become a common practice for dissecting the aetiology of disease. However, as more understanding of the complexity of diseases such as various cancers has been revealed, it has become clear that further stratification of the tissues studied will be necessary. Laser capture microdissection allows for the isolation of pure populations of cells, but reduces the amount of material available for profiling. Researchers have typically solved this problem through the employment of RNA amplification methodologies. While a typical microarray experiment might require 5 to 20 µg of total RNA, amplification methodologies have allowed for a dramatic reduction in this requirement.

The Affymetrix GeneChip™ platform has long been the most popular gene expression profiling technology in use. The standard protocol uses a T7-polymerase-based amplification method in order to profile between 1 and 8 µg of total RNA, posing a challenge to researchers using small amounts of starting material. Affymetrix also offers a 2-round amplification procedure, which can be used down to 100 to 200 ng of Total RNA. While this represents a significant increase in sensitivity, it still is insufficient for many researchers. The two round protocol is also much more time and resource intensive (4 days to complete) posing a problem for service laboratories.

An alternative to the Affymetrix protocols is the Ovation™ Biotin RNA Amplification and labelling system from NuGen (San Carlos, CA). This kit uses a three-step process (2 days in total) called RiboSPIA, which provides amplified, biotin-labelled cDNA from as little as 5 ng of total RNA. One of the challenges when assessing the effectiveness of a new labelling system is that it is difficult to compare the data without a strong benchmark. The recent MAQC project (1) has provided a standard to which other data can be compared. The MAQC project included 60 Affymetrix Human U133 plus 2.0 microarrays hybridised with one of two commercially available RNAs. We have used these same RNAs and the MAQC data to determine how stable the Affymetrix platform is, as well as to develop the necessary benchmark to which we will be able to compare the data from our evaluation of several amplification methodologies.

Methods and Materials

The RNAs used in the MAQC project (Stratagene Human Universal Reference RNA, cat# 740000 and Ambion brain RNA ,cat# Am6051) were used for this project. The RNA was from the same lot as that used in the initial MAQC project. RNA quality was assessed by BioAnalyzer (Agilent) prior to amplification and labelling.

Three different labelling protocols were evaluated for this study: Affymetrix one-cycle, Affymetrix two-cycle and NuGen Ovation™ Biotin RNA amplification system. The two more sensitive methods (Affymetrix U133-cycle and NuGen) require far less starting material and as such 10 ng of total RNA input was used for these two methods compared to 8 µg of total RNA for the standard Affymetrix one-cycle method.

The Affymetrix protocols are based on in vitro transcription reactions and yield between 30 and 50 µg of amplified RNA (aRNA) (Figure 1A). The NuGen protocol is an isothermal amplification method which yields 4-7 µg of cDNA (Figure 1B). While it appears the Ovation kit produces far less material, only 2 µg are required for hybridisation to an Affymetrix GeneChip compared to 15 µg of aRNA. We used the Affymetrix Human 133 plus 2.0 Gene Chips for this study as they are compatible with all three amplification systems and were used in the MAQC project.

Replicate labellings (6 for NuGen and Affymetrix 1-cycle, 3 for Affymetrix 2-cycle) of each RNA type were performed. Nugen and Affymetrix 1-cycle labellings were performed on two separate days (3 replicates each day). Hybridisation and scanning of the NuGen and Affymetrix one-cycle labelling kit were performed contemporaneously, whereas the Affymetrix 2-cycle labelling was performed on a separate day. Placement of the chips into the scanner and fluidics station were randomised and were switched between different days to avoid confounding the statistical analysis.
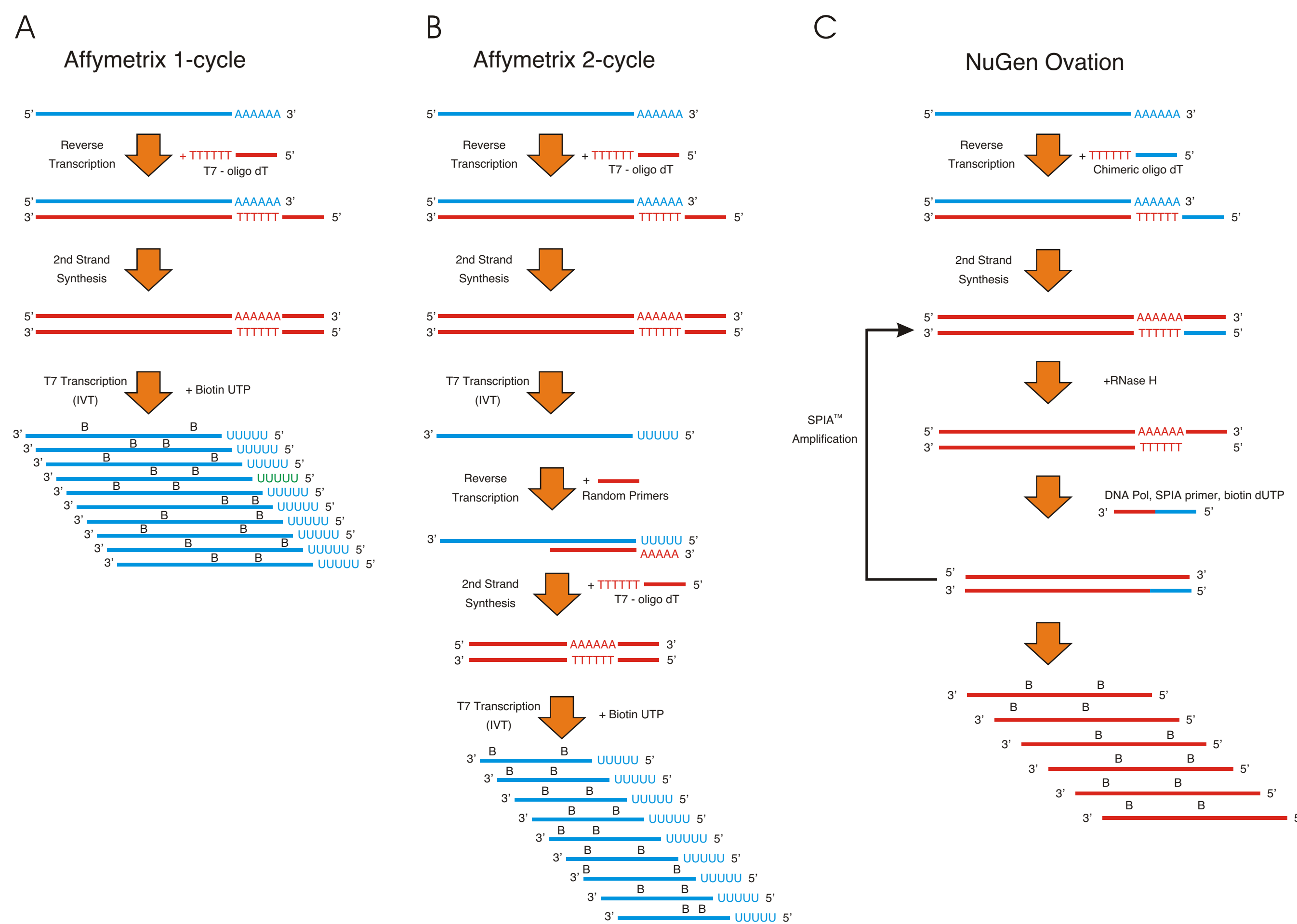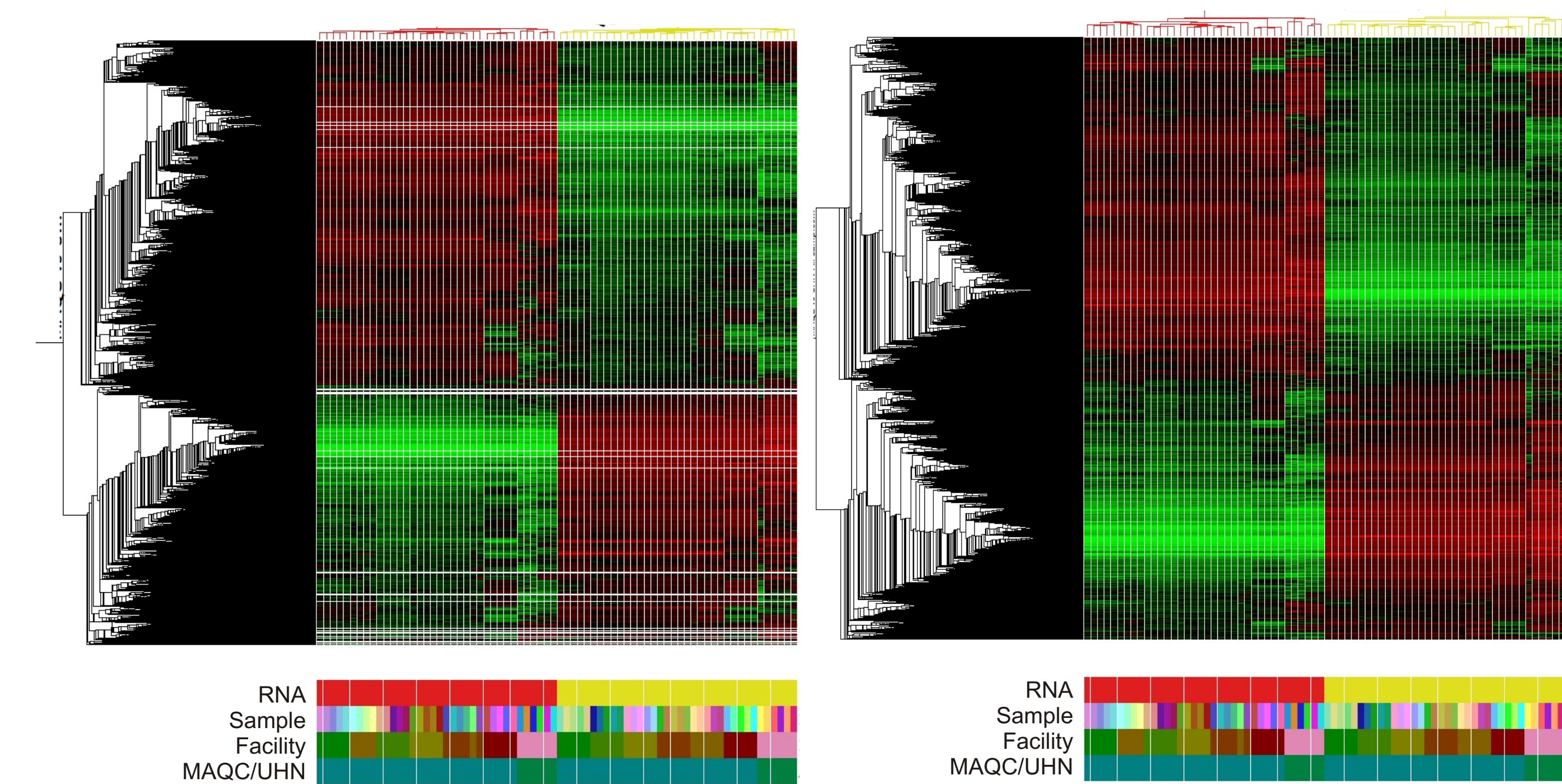

Figure 1. Labelling strategies for each of the methodologies used. (A) The standard Affymetrix 1-cycle IVT-based labelling method. (B) The Affymetrix 2-cycle IVT-based labelling method. (C) The NuGen Ovation Biotin labelling system. The one cycle reaction takes approximately 4.5 hours plus an overnight incubation. The two-cycle reaction takes approximately 9 hours plus two overnight incubations. The NuGen Ovation kit requires approximately 8 hours in total.

Results

Table 1. GCOS (Affymetrix) Quality Control report. Affymetrix recommends that the background should be below 100 counts. As can be seen some of the labelling reactions produced slightly higher background than this nominal cutoff. %Present calls tend to be in the range of 50% but are highly dependent on the RNA sample used. In general a higher %Present score is better. There is no recommended value provided by Affymetrix for the scaling factor. Noise typically should be in the range of 1-4. GAPDH and Actin scores should be in the range of 1-3. This value represents the ratio of 3' to 5' probe signals for these two control genes. A higher number indicates 3' bias. As can be seen, the NuGen kit and the two-cycle Affymetrix methods both produce values much higher than recommended. This is largely due to the fact that these methodologies produce shorter products.

Labelling Method	RNA	Average Background	% Present	Scaling Factor (500 Target)	Noise (RawQ)	GAPDH	Actin
Affymetrix 1-Cycle	UHRR	110.30	49.30	3.6	3.12	0.93	1.10
	UHRR	91.41	49.00	4.5	2.71	0.98	1.20
	UHRR	77.85	49.40	5.2	2.39	1.00	1.30
	UHRR	78.00	38.00	9.5	2.60	1.10	2.20
	UHRR	82.00	37.00	9.4	2.70	1.10	2.00
	UHRR	80.00	36.00	10.0	2.60	1.10	2.00
	Brain	77.38	46.20	5.7	2.34	1.50	2.10
	Brain	88.72	46.30	5.1	2.62	1.40	1.90
	Brain	76.08	45.40	6.0	2.30	1.50	2.00
	Brain	141.00	39.00	9.0	4.20	1.50	2.50
	Brain	146.00	40.00	4.6	4.20	1.20	1.90
	Brain	130.00	39.00	5.3	3.90	1.40	2.70
NuGen	UHRR	28.30	55.50	7.7	0.76	1.90	13.10
	UHRR	28.90	56.50	6.6	0.80	1.70	11.30
	UHRR	30.80	53.40	8.2	0.83	1.80	13.30
	UHRR	29.00	55.00	8.1	0.73	1.80	16.30
	UHRR	30.00	54.00	8.3	0.84	1.90	19.30
	UHRR	29.00	55.00	8.1	0.75	1.90	14.40
	Brain	32.20	56.00	6.7	0.88	2.20	18.60
	Brain	29.60	54.10	7.9	0.78	2.50	24.40
	Brain	30.70	56.80	6.6	0.80	2.40	19.10
	Brain	28.00	51.00	10.0	0.73	3.30	46.60
	Brain	30.00	50.00	10.9	0.81	3.10	35.20
	Brain	51.00	42.00	8.6	1.60	3.50	38.05
Affymetrix 2-Cycle	UHRR	43.70	18.30	81.0	1.30	26.40	68.30
	UHRR	44.20	19.20	74.0	1.30	26.70	52.80
	UHRR	43.00	17.20	84.0	1.30	37.40	97.49
	Brain	50.00	36.50	19.0	1.50	24.36	60.52
	Brain	52.00	34.80	23.0	1.50	30.12	84.00
	Brain	50.00	34.30	25.0	1.50	26.35	94.70

Run 1
Run 2
Run 3

Figure 2. Quality control data generated by BioConductor. (A) Normalised Unscaled Standard Errors (NUSE). Data consistency is represented by similar sized boxes clustered around the 1-line. Longer boxes indicate more data scatter. (B) Pseudo coloured representations of the array. This view allows for the identification of artifacts that may affect overall data quality. (C) Intensity distributions. Consistency will be represented by overlapping traces. Traces shifted to the right show a distribution that is tending toward higher signal intensities. (D) The RNA Degradation Plot indicates the 3' bias of each sample. Lower slopes indicate less bias.

RNA
Sample
Facility
MAQC/UHN

Figure 3. Hierarchical clustering of the MAQC array data and the UHN Affymetrix one-cycle data. (A) Data was filtered based on raw intensities, removing all probes with signal less than 150 in half of the arrays. A T-test was then performed to look for genes that were significantly differently expressed between UHRR and Brain using only the MAQC data for the statistic. Hierarchical clustering of all samples was then performed using this gene list. (B) Intensity filtering was performed as in (A) but the t-test was performed with the MAQC and UHN arrays present. The hierarchical cluster was then performed using the resultant gene list.

Discussion and Conclusions

Microarray research has frequently been hampered by the inability to generate enough starting material. Biological samples obtained via laser capture microdissection, flow cytometry, and fine needle aspirations tend to contain too little RNA to perform a standard microarray analysis. While a lot of work has been done to develop amplified labelling strategies, critical analysis of these labelling strategies has been difficult. With the Affymetrix platform, this comparison is further complicated by the fact that the standard labelling procedure itself is amplification based. When comparing labelling modalities, invariably certain genes will show bias to one method or the other. Without brute force validation, determining which of the two modalities is producing the more accurate data can be very difficult.

The recent publication of the Microarray Quality Control (MAQC) project's findings has helped provide a benchmark to which other platforms, strategies, and reagents can be compared. The data from the MAQC was cross validated among several platforms (array and qPCR based) to provide a relatively stable reference. Fortunately the RNAs used in this study are commercially available which allowed us to perform a direct comparison.

The initial results from the amplification tests based on the QC report generated by the Affymetrix GCOS software was encouraging. Percent present calls, background and noise levels from the NuGen kit were generally equivalent to or in many cases superior to the Affymetrix one-cycle labelling strategy (Table 1). When compared to the two cycle labelling strategy the NuGen showed much better results based on these QC metrics (admitedly however the Affy 2-cycle experiment was being conducted with the minimal recommended RNA input). While this was a strong indication that the NuGen system was the superior methodology, more in depth investigation was warranted. Further quality control tests using the Bioconductor package also indicated that the NuGen kit was performing fairly well showing fairly strong intra- and inter-experiment reproducibility and generally even background signals (much better than the standard Affymetrix protocol in this regard). However, the NuGen protocol interestingly also showed lower signal intensity distribution. This suggests that the improved % present calls are largely a result of the more even and lower background signals generated by this kit.

Comparison of the NuGen data to the Affymetrix data however began to show differences when statistical tests and hierarchical clustering was performed (data not shown). The reason for this discrepancy was unknown suggesting further calibration of our system may be required. In order to test our results, the Affymetrix one-cycle data was compared directly to the 60 arrays that are part of the MAQC project. When our data is compared to the MAQC data there is very strong agreement. A hierarchical cluster of genes found to be significantly different between the two RNA samples using only the MAQC data shows that our data is in strong agreement with the MAQC data. It is worth noting that our experiments were conducted more than one year after the original MAQC data using different lots of arrays and reagents, which indicates that the platform itself is relatively robust and consistent. This larger data set will now be used as a benchmark to compare both the NuGen and the 2-cycle Affymetrix data. We suspect that part of the reason for some discrepancy is due to the smaller products produced by these two more sensitive methodologies, which may lead to several probes not reporting information properly. We intend to conduct an analysis that takes probe position into account in order to fully evaluate the NuGen system.
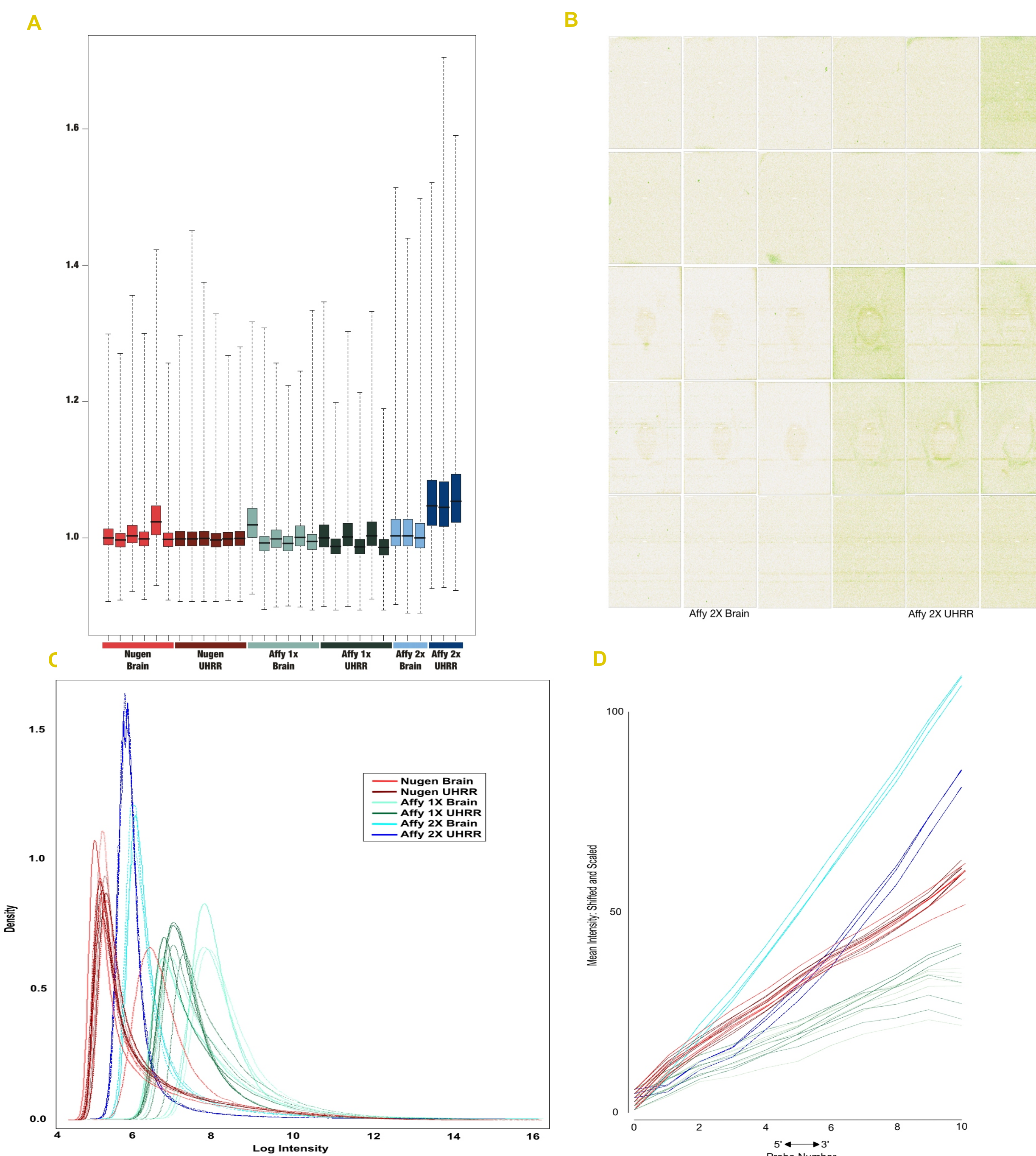
The hierarchical clustering data also provides some other interesting insights. While the MAQC experiments were all performed in a relatively narrow window of time using the same lots of chips, and reagents, the clustering algorithm is still able to distinguish each of the 6 different sites that were involves in the study. The UHNMAC data forms a distinct (but similar) 7-th cluster. This suggests that in order to fully control experiments, it is critical that researchers choose a single site to perform the experiments, and that a core facility that will dedicate a single technician to the entire experiment should be sought to prevent technicians or site based bias.

References

1. Shi et al. (2006) Nature Biotechnology **24**:1151-1161

Acknowledgements

The UHN Microarray Centre
University Health Network
Toronto Medical Discovery Tower
101 College Street
9th floor
1-877-294-4410 (Toll Free)
www.microarrays.ca

µA
UHN Microarray Centre