DNA MICROARRAY TECHNOLOGIES: A NOVEL APPROACH TO GENOMIC RESEARCH

ROCHELLE HINMAN[A], BRIAN THRALL, PH.D.[B], AND KWONG-KWOK WONG, PH.D.[B]

ABSTRACT

A cDNA microarray allows biologists to examine the expression of thousands of genes simultaneously. Researchers may analyze the complete transcriptional program of an organism in response to specific physiological or developmental conditions. By design, a cDNA microarray is an experiment with many variables and few controls. One question that inevitably arises when working with a cDNA microarray is data reproducibility. How easy is it to confirm mRNA expression patterns? In this paper, a case study involving the treatment of a murine macrophage RAW 264.7 cell line with tumor necrosis factor alpha (TNF-α) was used to obtain a rough estimate of data reproducibility. Two trials were examined and a list of genes displaying either a > 2-fold or > 4-fold increase in gene expression was compiled. Variations in signal mean ratios between the two slides were observed. We can assume that erring in reproducibility may be compensated by greater inductive levels of similar genes. Steps taken to obtain results included serum starvation of cells before treatment, tests of mRNA for quality/consistency, and data normalization.

INTRODUCTION

"*Reconstructing the genome is just the beginning. Figuring out how the 30000 genes played like piano keys give rise to the rhythms and melodies of life is going to take even more calculating power.*"
- George Johnson (Johnson, 2001)

The media frenzy started with a simple press conference. United States President Bill Clinton announced in June of 2000 that scientists working under Francis S. Collins and J. Craig Venter had succeeded in completing the first map of the human genome (Cawley, 2001). The news spread rapidly. The Human Genome Project soon became the front headline in every facet of the media.

Reporters were correct in labeling the map of the human genome as well as those of other organisms as a major discovery. Yet, the work of biologists is far from complete. As George Johnson of the New York Times so aptly put it, reconstruction is just the beginning. The task of figuring out how sections of the code interact with one another and outside stimuli is still at hand. It is here where cDNA microarray technology is an important tool.

A cDNA microarray allows biologists to examine the expression of thousands of genes simultaneously. Researchers may analyze the complete transcriptional program of an organism in response to specific physiological or developmental conditions (Lodish et al., 2000). The construction of a cDNA microarray experiment is a lengthy process involving multiple steps. These steps are broadly divided into three main sections: array fabrication; probe preparation and hybridization; and data collection, data mining, and analysis (Hegde et al., 2000).

During array fabrication, cDNA clones representing the genes of interest, either in the form of known or expressed sequence tags (ESTs), are amplified by PCR techniques. Clones are then mechanically spotted/adhered to a glass microscope slide or other surfaces. Probe preparation begins with mRNA extraction from selected tissues or cell lines and their corresponding controls. The mRNA is reverse transcribed and undergoes fluorescent tagging. Hybridization against the clones on the slide takes place, and the slide is then scanned with two lasers. Each laser detects a different fluorescent label: one for the tissue/cell line probe; the other for the control probe. Images are overlapped and analysis software is used to compare the resulting signals at each spot and identify which genes have undergone a change in expression. These genes may then be examined in detail.

There are many possible uses for cDNA microarray technology. It may be used to compare the genomes of different subspecies (Lashkari et al., 1997). An important application is towards an understanding of the effects of chemicals (Schena et al., 1996) and/or diseases (Nelson, 1999). The reconstruction of complex gene control networks is a future goal (Herzel et al., 2001). Yet, in all cases the use of technology can present problems.

One question that inevitably arises when working with cDNA microarrays is data reproducibility. How easy is it to confirm mRNA expression patterns? A cDNA microarray is by design an experiment with many variables. Sample to sample fluctuations in mRNA preparation as well as varying success rates with reverse transcription and labeling can occur. It is also possible for target-spot sizes to fluctuate and unspecific hybridization to take place. Many factors must be taken into consideration when analyzing data.

Yet, certain steps can be taken in order to obtain reliable re-

A: Whitworth College, Spokane, WA; B: Pacific Northwest National Laboratory, Richland, WA

sults. In this paper, a case study involving the treatment of a murine macrophage RAW 264.7 cell line with tumor necrosis factor alpha (TNF-α) will be used in order to outline several of these steps. Steps include serum starvation of cells before treatment, tests of mRNA for quality/consistency, data normalization, and trail repetition.

MATERIALS AND METHODS

Array Fabrication

The cDNA microarray slides used in this experiment were printed on site. The slides were printed using Cartesian Technologies PixSys 5000 workstation. They were printed with twenty-four ArrayIt™ ChipMaker Micro Spotting Pins on amino-silane coated slides. Clones were suspended in 50% DMSO and loaded into 384-well plates before printing. Approximately 6185 mouse cDNA clones, as well as a series of controls, were printed on each slide. Spots were then adhered to the slides by baking at 80 °C for two hours.

Extraction of mRNA

Four sets of cells from a murine macrophage RAW 264.7 cell line were prepared. Sets were 80% confluent and contained approximately five million cells. All cell sets were serum-starved overnight to minimize invariant cell-cycle differences that may hinder the impending results. Two cell sets were then treated with 25 ng/mL TNF-α. Two other cell sets were treated with an equal volume of DMSO as control. All cell sets were allowed to incubate for one hour before mRNA extraction. A Qiagen RNeasy® Midi Kit was used for mRNA extraction.

Extracted mRNA was resuspended in 200 µL RNAse free water. Optical densities on all samples were taken to determine concentration. A 1% agarose gel was run using 3 µg of each RNA sample in order to test for RNA purity. Samples were stored at -80 °C when not in use.

Probe Preparation and Hybridization

A MICROMAX™TSA™ Labeling and Detection kit was used to prepare probes and hybridize them to the printed slides. Probes of cDNA were prepared separately from each of the four mRNA samples using 3 µg of mRNA in each case. Probes originating from mRNA of TNF-α treated cells were labeled with fluorescein. Probes originating from mRNA of control cells were labeled with biotin.

Two pools of probes were prepared, each containing a TNF-α and control set. These pools were suspended in 20 µL of hybridization buffer. Each pool was hybridized against a separate slide (#1 and #2 respectively) overnight at 65 °C. At the end of hybridization, slides were washed. In the washing process, a Cy5 dye was bound to the biotin labeled control cDNA. A Cy3 dye was bound to the fluorescein labeled TNF-α cDNA. Slides were then spun dry in preparation for scanning.

Data Collection, Data Mining, and Analysis

Slides were scanned on ScanArray® 3000. This machine uses two lasers. One laser detects Cy3 and the other detects Cy5. Scanning was done sequentially, with Cy3 and Cy5 scans completed separately. Both slides were scanned two times with each laser.

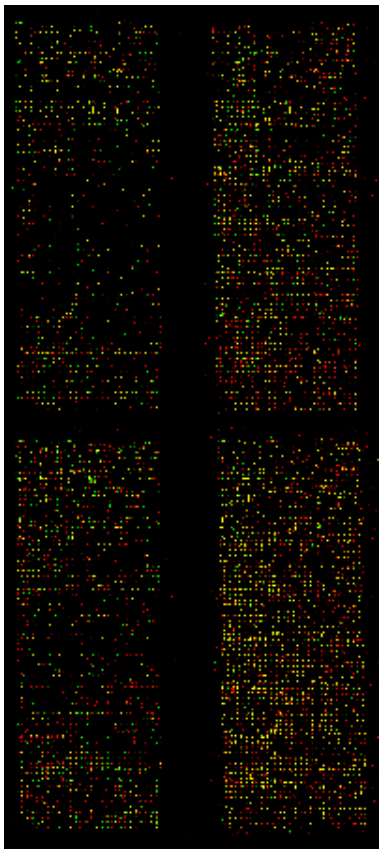

Figure 1.
Top Left Panel: Normal Scan of Slide #1
Top Right Panel: High Scan of Slide #1
Bottom Left Panel: Normal Scan of Slide #2
Bottom Right Panel: High Scan of Slide #2

One scanning was done at a high power; the other was done at a low power. The resulting images were loaded into ImaGene™ version 4.2 for analysis.

In ImaGene™ 4.2 corresponding Cy3 and Cy5 images for each scan type were overlaid. Cy3/TNF-α images were shown in red; Cy5/Control images were shown in green. For the resulting compilation, a red spot indicated a gene induced by TNF-α, a green spot indicated a gene repressed by TNF-α, and a yellow spot indicated a gene unchanged by the treatment. Some spots did not light up at all.

ImaGene™ 4.2 was used to calculate the Cy3 and Cy5 signals at each spot. Calculations were completed both with and without provided normalization methods. Provided normalization methods subtracted local background hybridization and took spot quality into account. Scatterplots and data tables were then produced. The resulting signal mean ratios (SMR) between Cy3/ TNF-α and Cy5/Control were compiled for further analysis. A signal mean ratio greater than one indicated that the gene was induced by TNF-α. A signal mean ratio less than one indicated a suppression of expression. In normalized data, all negative signal mean ratios were assigned a value of one. Thus, for purposes of this analysis, only genes induced 2-fold or greater were considered.

Table 1. Comparison of differences between Signal Mean Ratios (SMR) in Slide #1 and Slide #2.

	High Scan Non-normalized	High Scan Normalized	Normal Scan Non-normalized	Normal Scan Normalized
Difference Between SMR > 0.25	5293	4212	5669	4287
Difference Between SMR > 0.5	4169	2601	4899	2673
Difference Between SMR > 0.75	3225	1667	4210	1693
Difference Between SMR > 1.0	2498	1150	3578	1129
Difference Between SMR > 2.0	959	480	1858	452
Difference Between SMR > 3.0	402	301	1006	275

Table 2. Comparison of gene induction levels in Slide #1 and Slide #2.

	#1 High Scan Normalized	#2 High Scan Normalized
Genes Induced > 2 Fold	683	734
Genes Induced > 4 Fold	157	192

RESULTS

Compiled images of both slides (Figure 1) were visually compared. While some differences within the slides were evident, the slides displayed many of the same general trends in expression patterns. This was especially evident in those images produced via high scanning.

Comparisons of signal mean ratios between the two slides were made (Table 1). The differences between signal mean ratios in non-normalized data were higher than those in normalized data. High scanning produced signal mean ratios in closer correlation than normal scanning.

For high scanning, the number of genes induced by factors greater than two, three, four, six, eight, and ten-fold were determined (Table 2). For each slide, the number of these genes was slightly different. For example, slide #1 was shown to have 683 genes induced by a factor greater than two-fold. Slide #2 had 734 genes induced by the same factor. Although both treated slides shared just 147 common genes (20% reproducibility), differences in induction levels between the two were low (Table 3). Interestingly, when we examined the common genes that were induced by greater than four-fold, 20 genes (11% reproducibility) were markedly induced in both data sets (Table 4).

Several genes of interest were identified based on the compiled gene expression profiles. These genes included, but were not limited to, tumor necrosis factor, zinc finger protein 36, immunoresponsive gene 1, and leptin receptor.

DISCUSSION AND CONCLUSIONS

In a sense, a cDNA microarray can be thought of as a high throughput Northern Blot. Thousands of genes undergo a generalized procedure to obtain expression data (Chin, 2000). It is important to remember that conditions may not be optimal for all the genes involved. Certain genes may never register on the slide although mRNA for these genes is present in the cells.

In working with many expression profiles generated simultaneously, an exact duplication of results is extremely difficult. However, certain steps may help to ensure that results from multiple trials do have some correlation. In this study, all cells were serum-starved overnight. This process put all cells at the same stage of their life cycle before the experiment started. Differences in mRNA expression based on cell cycle stage were therefore minimized.

All mRNA extracted from the cells underwent a gel electrophoresis test for purity. The mRNA was shown to be pure and without degradation. This is important. Degraded mRNA can still undergo reverse transcription to create cDNA probes. However, these probes will anneal to the microarray slide in random fashion. They will anneal wherever their shorter cDNA sequences can be found, not necessarily on the spot corresponding to their genetic origin. Consequently, collected expression profiles will be difficult, if not impossible, to duplicate. For this reason, it is always a good idea to test mRNA quality before probe preparation, especially if some time has passed since the mRNA was extracted.

Data normalization in a cDNA microarray experiment can aid in obtaining reproducible results. Normalization processes can be used to correct for variations in background fluorescence on the slide. The degree of background on a slide can vary greatly from experiment to experiment. Normalization also compensates when one dye naturally fluoresces more than another dye. Spot size is taken into consideration as well as signals from surrounding spots. As indicated in this experiment, normalization can result in greater correlation between signal mean ratios on different slides.

In this experiment, two trials were run. Each trial used different sources of mRNA. These two trials were enough to obtain a rough estimate of data-reproducibility and identify probable genes of interest. However, further trials are necessary. Switching dyes would verify that the data obtained was not the result of dye location. Testing the two controls against one another would locate genes where differences in expression were a result of baseline mRNA variations and not TNF-α exposure. It would also be helpful to hybridize each of the TNF-α probes against the opposite control.

Once enough genes of interest have been identified, a smaller cDNA microarray may be printed for further experimentation. In a smaller microarray, additional steps may be taken in order to verify results. Clones may be spotted multiple times on the slide. These spots may be centered in one location on the slide or spread out evenly across the surface. In either case, the result is multiple signal mean ratios that can be compared and averaged for each experiment.

The technology available through a cDNA microarray has many applications. The vast majority of these applications center

Table 3. Genes induced by a factor greater than two-fold on Slide #1 and Slide #2 (High Scan Normalized Data).

Gene ID	*SM R #1*	*SM R #2*
tumor necrosis factor	25.936	24.575
"AI385562""CytochromeP450,2b9,phenobarbitolinducible,typea"""	3.9947	4.8613
"AI894223""MusmusculusP38deltaMAPkinasemRNA,completecds"""	3.4195	2.3631
acetylcholine receptor epsilon	2.0807	3.6037
serum-inducible kinase	4.2471	2.4695
protein kinase C and casein kinase substrate in neurons 1	12.359	3.79
AI323310Fos-likeantigen2	2.2174	2.3988
AI323330SRY-boxcontaininggene17	4.2432	2.094
426318ESTs	2.24	81.142
AI426345ESTs	4.4362	3.3971
AI429085ESTs	2.0703	2.6932
AI449073ESTs	8.6087	2.8756
AI528691B-cellleukemia/lymphoma3	2.1262	2.0028
AI528708MYELOIDDIFFERENTIATIONPRIMARYRESPONSEPROTEINMYD116	2.8851	3.8973
AI430953ESTs	4.6128	7.5741
AI447215EST	6.1161	4.0212
AI447814ESTs	2.8165	2.2979
AI464404ESTs	3.1168	2.2491
procollagen, type III, alpha 1	2.9925	2.1645
AI528691B-cellleukemia/lymphoma3	2.1262	2.0028
AI893411Zincfingerprotein36	10.958	14.238
AI428386ESTs	3.4921	2.1343
AI448307ESTs	3.7874	2.8138
AI452258ESTs	3.2407	9.0342
AI447233ESTs	2.7646	2.1711
"AI447335""EST,HighlysimilartosimilartoAL031532[H.sapiens]"""	2.6391	2.8585
AI451116EST	5.8156	5.3604
small inducible cytokine subfamily D, 1 Scyd1	16.972	5.4329
"AI323555""Solutecarrierfamily2(facilitatedglucosetransporter),member2"""	2.0114	2.3185
AI323680IMMEDIATEEARLYPROTEINGLY96	6.2085	4.0352
"AI414038""ESTs,WeaklysimilartoapolipoproteinA-IVprecursor[M.musculus]"""	2.8466	2.1769
AI447522ESTs	2.0874	2.1013
"AI451916""Ubiquitouslytranscribedtetratricopeptiderepeatgene,XChromosome"""	2.7528	3.8049
5'-CCACCACTTCAGTGTGGTTTGGAAAAGGGACAGATGAGCCCCTGAAGACGAGGTGAAAAGTCAATTTTAC-3'	4.3773	6.58
Homo sapiens TNF-alpha stimulated ABC protein (ABC50) mRNA,	10.905	3.2333
"AI660999""ESTs,WeaklysimilartoNG28[M.musculus]"""	2.338	2.524
AI414026ESTs	5.1666	6.2979
AI414480ESTs	3.9582	2.477
AI425989ESTs	2.3381	2.6646
"AI426007""ESTs,HighlysimilartoNEURONALPROTEINNP25[Rattusnorvegicus]"""	2.05	7.5056
AI450130ESTs	4.9952	28.109
Mus musculus WNT-2 gene, partial cds; putative ankyrin-related protein	3.1488	4.2211
AI385610SemaphorinF	4.4217	2.1641
AI323667Immunoresponsivegene1	9.5027	5.6812
AI413231ESTs	2.9306	2.9918
AI413235ESTs	6.246	2.891
AI447645ESTs	2.1168	2.8712
"AI449667""ESTs,ModeratelysimilartoCTM[M.musculus]"""	26.67	2.4479
5'-TCCAGTTCCTGTCCCAGCAGACTGGATGAACCGTGTGGAGATGAACGAGACCCAGTACAGTGAAATGTTT-3'	3.9799	14.039
Mus musculus ferritin light chain 1 (Ftl1), mRNA	12.583	2.5916
Mus musculus serine protease inhibitor 6 (Spi6), mRNA	11.247	2.8004

Table 3. (Continued)

Gene ID	*SMR #1*	*SMR #2*
adrenergic receptor, beta 3	3.1604	4.5372
AI415019ESTs	3.3453	2.2811
AI430799ESTs	4.1444	3.6952
AI449378ESTs	3.053	3.8083
small inducible cytokine subfamily B, member 15	3.0449	2.7636
"AI426665""ESTs,WeaklysimilartoCALCIUM-BINDINGPROTEINP22[R.norvegicus]"""	2.6975	2.5254
AI452320ESTs	2.6611	2.5237
cytokine inducible SH2-containing protein 2 Cish	2.8703	2.7936
ESTs, Highly similar to ATRIAL NATRIURETIC PEPTIDE CLEARANCE RECEPTOR PRECURSOR [Rattus norvegicus]	3.0732	3.2623
AI666741ESTs	2.8351	6.5842
AI528709Mousegeneformuchainassociationprotein(8hs20)	7.1497	6.4497
AI430805EST	3.9898	5.8722
"AI431039""ESTs,WeaklysimilartoKIAA0584protein[H.sapiens]"""	2.6259	6.8584
AI426162ESTs	2.6697	2.0884
"AI426674""EST,WeaklysimilartosimilartoTRNA[H.sapiens]"""	3.4883	2.0884
AI428422ESTs	2.4028	2.3146
AI464549ESTs	3.1995	3.0491
"AI323701""MusmusculusSKD3mRNA,completecds"""	9.2152	4.1922
AI414514EST	16.772	2.7323
AI414525EST	9.7347	2.0268
ESTs, Highly similar to PROBABLE G PROTEIN-COUPLED RECEPTOR GPR18 [H.sapiens]	2.7856	32.813
Mm.29844 ESTs	2.7781	3.0611
AXL receptor tyrosine kinase	2.033	2.1216
AI385752Distal-lesshomeobox5	3.1178	2.3148
"AI326823""ESTs,HighlysimilartoLENSFIBERMEMBRANEINTRINSICPROTEIN[M.musculus]"""	2.4196	6.5354
"AI414810""ESTs,ModeratelysimilartoS-ACYLFATTYACIDSYNTHASETHIOESTERASE,MEDIUMCHAIN[R.norvegicus]"""	2.1662	3.55
AI414821ESTs	2.6085	2.5915
AI429810ESTs	3.347	3.3936
"AI426455""MusmusculusSin3-associatedprotein(sap30)mRNA,completecds"""	2.5355	2.0139
AI451295ESTs	2.8999	3.5372
small inducible cytokine A2	2.5952	6.392
AI449520ESTs	2.3186	2.3155
AI447908ESTs	2.5892	2.0078
gap junction membrane channel protein beta 5 Gjb5	2.2108	4.4055
nuclear receptor subfamily 0, group B, member 2	2.2327	2.2174
AI449420ESTs	2.2606	2.1599
AI447436ESTs	3.2093	2.9957
homeodomain interacting protein kinase 2	2.8656	2.2825
AI528743Myxovirus(influenzavirus)resistance2	3.7441	3.3611
AI326155MusmusculusmRNAforvery-long-chainacyl-CoAsynthetase(VLACS)	3.4356	2.0425
AI428447EST	2.3858	2.1103
interleukin 15 receptor, alpha chain	4.8704	2.0444
cytochrome P450, 4a10	23.493	2.3457
AI451958ESTs	2.2101	3.4445
"AI448672""MusmusculuscloneL5uniformgroupof2-cell-stagegenefamilymRNA,completecds"""	3.5762	2.5035
AI447753EST	2.6478	2.9979
AI447183ESTs	3.0624	2.8211
AI447862ESTs	2.7553	3.0587
AI449628EST	2.925	2.1751
AI450841ESTs	2.4811	2.9297
AI451315EST	2.2694	2.2659
AI465471ESTs	4.035	4.7052
small inducible cytokine B subfamily (Cys-X-Cys), member 10	2.691	2.8823
procollagen, type VII, alpha 1	5.735	2.2219
endothelial-specific receptor tyrosine kinase	2.6609	2.3717
AI323614Myelocytomatosisoncogene	2.6817	2.2608
AI528697Zincfingerprotein62	2.2608	3.5774
AI426040ESTs	4.2577	8.0409

Table 3. (Continued)

Gene ID	*SMR #1*	*SMR #2*
AI430784ESTs	2.2549	7.1683
AI447507EST	3.2541	3.2948
AI449234CD48antigen	2.2482	2.1122
Mus musculus SWI/SNF related, matrix associated, actin dependent	2.4076	2.3079
grow th accentuating protein 43	2.0039	4.1651
trefoil factor 1	2.2367	2.123
T-cell receptor gamma, variable 4	2.9209	2.4972
CD3 antigen, zeta polypeptide	2.2674	2.3616
"AI528680""MusmusculustranscriptionfactorGlFmRNA,completecds"""	4.3126	2.0259
lung carcinoma myc related oncogene 1	2.1714	4.6654
"AI323965""ATPase,Na+/K+transporting,beta1polypeptide"""	2.1048	2.3011
AI448121ESTs	4.018	3.7943
AI428470ESTs	2.7799	2.0604
AI449547ESTs	5.394	3.2503
AI447293EST	2.3183	2.0301
AI447603ESTs	6.2375	2.5641
AI448810ESTs	3.2941	2.8877
AI662536ESTs	2.3631	2.2141
"AI464440""ESTs,WeaklysimilartoHYPOTHETICAL139.1KDPROTEINC08B11.3INCHROMOSOMEII[Caenorhabditiselegans]"""	2.2688	2.0232
AI464450ESTs	3.5499	2.4294
Cytochrome P450, 2a4	2.8682	2.9134
phospholipase A2, group IB, pancreas, receptor	5.9435	2.0658
"AI661491""Guaninenucleotidebindingprotein,alphao"""	2.5411	2.0717
"AI323836""Colonystimulatingfactor,macrophage"""	2.2982	2.4251
AI449557EST	2.2322	4.8917
AI449453EST	2.0107	2.0181
AI449461ESTs	3.825	6.1049
AI447443ESTs	2.3871	2.047
AI464308ESTs	5.5826	2.2535
AI465162ESTs	6.699	2.6199
toll-like receptor 2	3.7917	3.2385
cAMP response element binding protein-related protein	3.0204	6.0496
uracil-DNA glycosylase	2.3573	2.4672
AI327008M.musculusmRNAforcytochromeP450IIIA25	17.853	3.1653
AI415104ESTs	3.3717	2.7119
AI447289ESTs	3.7061	2.1763
AI447940ESTs	2.5051	2.1495
Mus musculus cytochrome P450 CYP2D22 mRNA, complete cds	27.356	2.2019

Table 4. Genes Induced by a Factor Greater than 4-fold on Slide #1 and Slide #2 (High Scan Normalized Data)

Gene ID	*SM R #1*	*SM R #2*
"AI385598""MurineGlvr-1mRNA,completecds"""	5.16	4.78
AI414709ESTs	5.17	4.05
tumor necrosis factor	25.94	24.58
AI661215ESTs	6.67	7.14
AI325160MARCKS-RELATEDPROTEIN	5.26	4.16
AI430953ESTs	4.61	7.57
AI447215EST	6.12	4.02
AI893411Zincfingerprotein36	10.95	14.24
AI451116EST	5.82	5.36
small inducible cytokine subfamily D, 1 Scyd1	16.97	5.43
AI323680IMMEDIATEEARLYPROTEINGLY96	6.21	4.04
5'-CCACCACTTCAGTGTGGTTTGGAAAAGGGACAGATGAGCCCCTGAAGACGAGGTGAAAAGTCAATTTTAC-3'	4.38	6.58
AI414026ESTs	5.17	6.3
AI450130ESTs	5	28.11
AI323667Immunoresponsivegene1	9.5	4.68
AI528709Mousegeneformuchainassociationprotein(8hs20)	7.15	6.45
"AI385688""MousemRNAforZfp-57,completecds"""	5.19	13.71
"AI323701""MusmusculusSKD3mRNA,completecds"""	9.22	4.19
AI323343Leptinreceptor	6.82	10.09
AI465471ESTs	4.03	4.71

on deciphering the genetic code of life. Although many factors must be considered, the technology still generates a list of genes of interest faster than would be possible if working with traditional methods on a gene-by-gene basis. Based on the differences in generation of signal mean ratios between the two similarly treated slides, we can assume that greater inductive levels of similar genes may alleviate variations in reproducibility.

ACKNOWLEDGEMENTS

The research described in this paper was preformed at the Life Sciences Laboratory, a part of Pacific Northwest National Laboratories in Richland, Washington. I would like to take this opportunity to thank the Department of Energy, Office of Science for their creation, organization, and funding of the ERULF program and to the National Science Foundation for their sponsorship. My thanks also go to my mentors Dr. Brian Thrall and Dr. Kwong-Kwok Wong, as well as Dr. B.Y. Chin and Meng Markillie for their assistance and patience over the course of this project.

REFERENCES

Cawley, Janet (2001, January). Biography of the year. Biography, 50-59.

Chin, B.Y. "Bioformatics." John Hopkins University, Baltimore. 12 Febuary 2000.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughs, J.E., Snesrud, E., Lee, N., & Quackenbush, J. (2000, September). A concise guide to cDNA Microarray Analysis. BioTechniques, 29, 548-562.

Herzel, H., Beule, D., Kielbasa, S., Korbel, J., Sers, C., Malik, A., Eickhoff, H., Lehrach, H., & Schuchhardt, J. (2001, March). Extracting information from cDNA arrays. CHAOS, 11 (1), 98-106.

Johnson, George (2001, March 25). All science is computer science. The New York Times, pp. 4.1.

Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., & Davis, R.W. (1996, October). Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. USA, 94, 13057-13062.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., & Darrell, J. (2000). Molecular Cell Biology (4th ed.) New York: W.H. Freeman and Company.

Nelson, N.J. (1999, December). Genetic profiling for cancer surfaces slowly in the clinic. J. Natl. Cancer Inst., 91, 1990-1992.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., & Davis, R.W. (1996, October). Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. USA, 93, 10614-10619.