



Graduate Theses, Dissertations, and Problem Reports

2003

Gene expression profiling in prepubertal and adult male mice using cDNA and oligonucleotide microarrays

Lisa Marie Tomascik-Cheeseman
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Tomascik-Cheeseman, Lisa Marie, "Gene expression profiling in prepubertal and adult male mice using cDNA and oligonucleotide microarrays" (2003). *Graduate Theses, Dissertations, and Problem Reports*. 1844.

<https://researchrepository.wvu.edu/etd/1844>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**GENE EXPRESSION PROFILING IN
PREPUBERTAL AND ADULT MALE MICE USING
CDNA AND OLIGONUCLEOTIDE MICROARRAYS**

Lisa Marie Tomascik-Cheeseman

Dissertation submitted to the
Genetics and Developmental Biology Program,
College of Agriculture, Forestry, and Consumer Sciences,
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Genetics

Joginder Nath, Ph.D., Co-Chair
Andrew J. Wyrobek, Ph.D., Co-Chair
Walter Kaczmarczyk, Ph.D.
Daniel Panaccione, Ph.D.
Sharon Wenger, Ph.D.

Morgantown, West Virginia
2003

Variations in gene expression are the basis of differences in cell and tissue function, response to DNA damaging agents, susceptibility to genetic disease, and cellular differentiation. The purpose of this dissertation research was to characterize variation in basal gene expression among adult mouse tissues for selected stress response, DNA repair and damage control genes and to utilize variation in temporal gene expression patterns to identify candidate genes associated with germ cell differentiation from mitosis through meiosis in the prepubertal mouse testis. To accomplish these goals, high throughput analyses of gene expression were performed using custom cDNA and random oligonucleotide microarrays. cDNA microarray technology was optimized by evaluating the effects of multiple hybridization and image analysis methodologies on the magnitude of background-subtracted hybridization signal intensities. The results showed that hybridizing lower probe quantities in a buffer developed at Lawrence Livermore National Laboratory to tryptone-blocked microarrays improved signal intensities. In addition, the error in expression ratio measurements was significantly reduced when microarray images were preprocessed. A custom cDNA microarray comprised of 417 genes and enriched for stress response, DNA repair, and damage control genes was used to investigate basal gene expression differences among adult mouse testis, brain, liver, spleen, and heart. Genes with functions related to stress response exhibited the most variation in expression among tissues whereas DNA repair-associated gene expression varied the least. Random oligonucleotide microarrays comprised of ~10,000 genes were

used to profile changes in gene expression during the first wave of spermatogenesis in the prepubertal mouse testis. Approximately 550 genes were differentially expressed as male germ cells differentiated from spermatogonia to primary spermatocytes. These findings suggest that the 313 unannotated sequences and 178 genes with known functions in other biological pathways have spermatogenesis-associated roles. This dissertation research showed that microarrays are a useful tool for quantitating the expression of large numbers of genes in parallel under normal physiological conditions and during differentiation. It has also provided candidate genes for future investigations of the molecular mechanisms underlying (1) tissue-specific DNA damage response and genetic disease susceptibility and (2) cellular differentiation during the onset and progression of spermatogenesis.

This dissertation is dedicated to my husband, Matthew,
for his advice, encouragement, and (most of all) love.

Thank you for giving me the courage to dream.

This work is also dedicated to my parents, Thomas and Carol

Tomascik,

and to my grandmother, Sophie Duzen

in honor of their tireless love and support.

In memory of my grandfather, Andrew Duzen.

Table of Contents

Abstract	ii
Acknowledgements	vii
List of Tables	viii
List of Figures	x
List of Abbreviations	xii
Chapter 1. Introduction	1
Dissertation Objectives	4
Chapter 2. Expression Microarray Technology	
Chapter Overview	10
Section 2.1 Fluorescent expression microarray technology review	
2.1.1 Introduction	11
2.1.2 Types of Expression Microarrays	13
2.1.3 Probe Generation, Hybridization, and Image Captu	14
2.1.4 Microarray Data Acquisition and Analysis	20
2.1.5 Summary	21
2.1.6 References	22
Section 2.2 Optimization of two-color fluorescence cDNA microarray hybridizations visualized with a white light imaging system	
2.2.1 Abstract	27
2.2.2 Introduction	27
2.2.3 Materials and Methods	31
2.2.4 Results	37
2.2.5 Discussion	46
2.2.6 References	50
Section 2.3 Accurate quantitation of fluorescence microarrays	
2.3.1 Abstract	51
2.3.2 Introduction	51
2.3.3 Materials and Methods	53
2.3.4 Image Processing and Analysis Methods	58
2.3.5 Results	69
2.3.6 Conclusions	73
2.3.7 Auspices	74
2.3.8 References	75

Chapter 3. Differential Basal Expression of Genes Associated with Stress Response, Damage Control, and DNA Repair Among Mouse Tissues	
3.1 Abstract	77
3.2 Introduction	78
3.3 Materials and Methods	80
3.4 Results	87
3.5 Discussion	100
3.6 Acknowledgements	106
3.7 References	107
Chapter 4. Mitotic and Meiotic Gene Expression Profiling of Male Germ Cells	
4.1 Abstract	114
4.2 Introduction	115
4.3 Materials and Methods	117
4.4 Results	121
4.5 Discussion	133
4.6 Acknowledgements	137
4.7 References	138
Chapter 5. Conclusions	152
Appendices	
A Symbols, names, biological pathways, I.M.A.G.E. clone IDs, and cDNA microarray data for 152 genes with differential expression among healthy adult mouse tissues	155
B All annotated genes showing significant differences in expression during the transition from mitosis to meiosis in the prepubertal mouse testis	160
C Unannotated sequences with significant expression differences during the first wave of spermatogenesis in the prepubertal mouse	167
D Functional genomics-related websites	173
E Curriculum Vitae	175
F Published research articles	180

Acknowledgements

Dr. Joginder Nath, Academic Advisor and graduate committee Co-chair: Words can't express how thankful I am to have been your student. Your guidance has enabled me to grow both personally and professionally.

Dr. Andrew J. Wyrobek, Research Advisor and graduate committee Co-chair: I am so grateful for the opportunity to be a member of your laboratory. Thank you for constantly challenging me and always finding time for my research.

Dr. Walter Kaczmarczyk, Dr. Daniel Panaccione, and Dr. Sharon Wenger, graduate committee members: Thank you for reviewing the dissertation and providing valuable comments.

Past and present members of the Wyrobek laboratory: Thank you for the spirited scientific conversations and camaraderie. I would especially like to recognize:

Dr. Francesco Marchetti: I could never thank you enough for everything you have taught me. I will always cherish our friendship.

Eddie Sloter: I have enjoyed every discussion (scientific and otherwise). You have been a wonderful friend, and I cannot imagine what this experience would have been like without you.

Dr. Matthew A. Coleman: Thanks for always keeping me on task by asking, "Is that dissertation finished yet???" I admire your scientific insight and appreciate your friendship.

Francesca Hill: Thanks for sharing your cytogenetic expertise and friendship.

Nancy Wrigley: Thank you for the administrative assistance and motherly advice.

Dr. David O. Nelson and Shea Gardner: I am truly thankful for all of your assistance with the microarray statistical analyses.

Laura Mascio Kegelmeyer: Thanks for patiently teaching me everything I ever needed to know about UNIX systems. Perhaps more importantly, thank you for the afternoon tea.

The Tomascik and Cheeseman families: Your love and unwavering support has made this dream possible. I promise to get a REAL job...someday.

Dr. Sara Frias, Ana Claudia Velazquez-Wong, Doug and Linda Hamilton, and the Jackson family: I will never forget the laughter, smiles, and many words of encouragement. Thank you for your friendship.

List of Tables

Chapter 2.	Page
Expression Microarray Technology	
Section 2.2 Optimization of cDNA microarray hybridizations visualized with a Xenon light source	
2.2.1 Primers used to amplify cDNA clones for the gene-specific hybridization experiments	33
2.2.2 Experimental design for the gene-specific and tissue sample hybridizations	35
2.2.3 Comparison of background-subtracted signal intensities obtained from gene-specific hybridizations	40
2.2.4 Characterization of the signal intensities obtained for each Alex Fluor following different tissue sample hybridization protocols	43
2.2.5 Relationship between tissue sample hybridization protocol and number of genes with signal intensity ratios ≥ 1.5	45
Section 2.3 Accurate quantitation of fluorescence microarrays	
2.3.1 Six genes were used to build a groundtruth series over ten microarray slides	56
2.3.2 Error measurements obtained for the groundtruth series following different image analysis methods	71
Chapter 3.	
Differential Basal Expression of Genes Associated with Stress Response, Damage Control, and DNA Repair Among Mouse Tissues	
3.1 Differences in the gene expression ratios from replicate independent hybridization:	91
3.2 Distribution of F ratios for differentially expressed genes with respect to biological pathway	96
3.3 Differentially expressed stress response, DNA repair, and damage control genes	97

Chapter 4.

Mitotic and Meiotic Gene Expression Profiling of Male Germ Cells

- | | | |
|-----|--|-----|
| 4.1 | Examples of genes with differential expression between mitosis and the onset of meiosis | 125 |
| 4.2 | Genes with significant differential expression between the onset and middle of meiosis I | 128 |
| 4.3 | Genes with complex patterns of differential expression from mitosis to mid-meiosis I | 130 |

List of Figures

Chapter 2.	Page
Expression Microarray Technology	
Section 2.1 Fluorescent expression microarray technology review	
2.1.1 Experimental outline for cDNA microarray hybridizations	15
2.1.2 Experimental outline for Affymetrix microarray hybridizations	16
Section 2.2 Optimization of cDNA microarray hybridizations visualized with a Xenon light source	
2.2.1 Effect of hybridization buffer and slide blocking protocol on background-subtracted microarray signal intensities	38
2.2.2 Distribution of background-subtracted median signal intensities obtained for tissue sample hybridizations	42
Section 2.3 Accurate quantitation of fluorescence microarrays	
2.3.1 A contrast-enhanced fluorescence microarray image with very high background	61
2.3.2 Removal of underlying autofluorescence and subsequent lowering of target spot intensities	63
2.3.3 Contours of hybridized target spots delineated by one of four segmentation methods	65
2.3.4 The groundtruth (ratio of known probe amounts) for <i>Rad52</i> was tracked among 10 slides	70
 Chapter 3.	
Differential Basal Expression of Genes Associated with Stress Response, Damage Control, and DNA Repair Among Mouse Tissues	
3.1 Intensity-based normalization of cDNA microarray data	88
3.2 Distribution of normalized expression ratios for testes vs. testes hybridizations	90
3.3 Distribution of gene expression results among tissues	93
3.4 Cluster analysis of genes using CLUSFAVOR	94
3.5 Comparison of northern blot and cDNA microarray results for selected genes	99

Chapter 4.

Mitotic and Meiotic Gene Expression Profiling of Male Germ Cells

4.1	Germ cell progression through mouse spermatogenesis	122
4.2	Genes that were differentially expressed between mitosis and the onset of meiosis	124
4.3	Genes that were differentially expressed between the onset of meiosis and mid-meiosis I	127
4.4	Genes with complex patterns of differential expression from mitosis to mid-meiosis I	129
4.5	Examples of gene sub-clusters obtained using CLUSFAVOR	132

List of Abbreviations

Abbreviation	Full name
AP	Apoptosis
BER	Base excision repair
bp	Base pair
CC	Cell cycle
cDNA	Complementary deoxyribonucleic acid
Cy3	Cyanine 3
Cy5	Cyanine 5
DAPI	4',6-diamidino-2-phenylindole
DDR	Direct damage reversal
dNTP	Deoxyribose nucleoside triphosphate
d.p.c.	Days post-coitum
DSBR	Double strand break repair
EST	Expressed sequence tag
FITC	Fluorescein isothiocyanate
GS	Generalized stress response
HS	Heat shock response
IVT	<i>In vitro</i> transcription
Kb	Kilobase pair
LLNL	Lawrence Livermore National Laboratory
MI	Meiosis I

MII	Meiosis II
NER	Nucleotide excision repair
p.n.	Postnatal
RT	Reverse transcription
OS	Oxidative stress response
PC	Personal computer
SAGE	Serial analysis of gene expression
SDS	Sodium dodecyl sulfate
SSC	Saline sodium citrate

1

Introduction

Every cell contains the complete genetic code for an organism, yet cell morphologies and functions are biologically diverse. The morphological, biochemical, and physiological diversity of cells within an organism is attributable to differences in their history of gene expression during differentiation (Strachan and Read, 1999). The profile of genes transcribed in a differentiated cell includes (a) genes whose expression is specific and required for a given cell type, (b) genes that are essential for general cell functions and are expressed in all cell types, and (c) genes that have tissue-specific functions but low levels of transcription are observed in all cell types (referred to as illegitimate expression; Strachan and Read, 1999). Understanding the variation in gene expression patterns among cells and tissues under normal physiological conditions and abnormal conditions (e.g., in response to genotoxic agents) is important for understanding cellular differentiation and function.

In 1961 Jacob and Monod reported that the expression of groups of genes in the bacterial operon is strictly coordinated. Expression studies up to the 1990s were typically limited to analyzing one or a small number of genes using techniques such as cell or tissue *in situ* hybridization, northern blot, RNA dot blot, etc. These low throughput investigations provided only limited insight on the molecular events underlying

individual biological processes (e.g., cellular differentiation, cell cycle, DNA repair, etc.). However, these techniques are not able to simultaneously analyze the expression of the many genes involved in any complex biological process. The advent of new molecular capabilities, such as automated PCR, together with the exponential increase of publicly accessible genome sequence information has facilitated the development of high throughput methods for gene expression analysis (i.e., serial analysis of gene expression or SAGE, expression microarrays, and mRNA differential display). Through the use of these genome-scale techniques, it is possible to analyze the expression of several thousand genes in parallel.

SAGE and mRNA differential display are used to determine which sequences are differentially represented between two samples. This information is then used to determine gene identities using other molecular techniques (i.e., DNA sequencing) and bioinformatics approaches (e.g., pairwise comparisons of each newly determined sequence with the non-redundant sequence database available through the National Center for Biotechnology Information). Microarrays, on the other hand, analyze the differential expression of a set of known genes and/or unannotated sequences that are selected *a priori*. Expression microarrays can be customized to include only specific genes of interest. It is feasible to build custom microarrays to simultaneously study the expression of all genes involved in a given biological process (e.g., cell cycle, apoptosis, DNA repair, etc.). Alternatively, expression microarrays may contain random sets of genes that represent a variety of biological pathways. These random microarrays are useful for discovering new genes associated with a biological process or tissue pathology.

A particularly promising application of both types of expression microarrays is the high throughput investigation of a broad range of biological questions in model organisms, such as the mouse. Microarray-based studies of tissues from model organisms promise to provide valuable insight on differential gene expression (a) under normal physiological conditions, (b) during development, (c) following exposure to genotoxic agents, and (d) during carcinogenesis. Although a genome-wide characterization of basal gene expression levels is necessary for a thorough understanding of the molecular mechanisms underlying tissue-specific differences in differentiation, response to DNA damaging conditions, and genetic disease susceptibility, the high throughput analysis of tissue-specific differences in basal gene expression has not been reported. Several laboratories, however, have started to examine global changes in gene expression in the mouse following exposure to genotoxic agents such as ionizing radiation (Amundson et al., 2001) and phencyclidine (Toyooka et al., 2002). Expression microarrays are also being utilized to profile transcriptional changes associated with various types of cancer (e.g., prostate cancer: Ho and Lau, 2002; breast cancer: Jiang et al., 2002; colon cancer: Zou et al., 2002; etc.).

Using expression microarrays, it is possible to obtain a more comprehensive evaluation of the sequential molecular changes that are correlated with cellular differentiation during development. The prepubertal mouse is an excellent model for studying the modulations in expression profiles that are associated with germ cell differentiation through the mitotic and meiotic phases of spermatogenesis. A characterization of gene expression across critical timepoints during spermatogenesis is essential for understanding the molecular mechanisms of meiosis (e.g., transition from

spermatogonial mitosis to meiosis: Wolfes et al., 1989; synaptonemal complex assembly: Allen et al., 1996; meiotic cell division: Koji et al., 1992; etc.). Furthermore, a detailed characterization of the normal baseline events may help us understand the genetic causes of male infertility (e.g., meiotic arrest of primary spermatocytes: Bailis et al., 2000; genetic causes for infertility previously characterized as idiopathic), and the induction and transmission of mutations to offspring following paternal exposure to genotoxic agents (e.g., ethylnitrosourea: Russell et al., 1979; cyclophosphamide; Schimenti et al., 1997; ionizing radiation: Matsuda et al., 1985; etc.).

Dissertation objectives

The objectives of this dissertation research were to characterize the differential basal expression of stress response, damage control, and DNA repair genes among healthy mouse tissues and investigate differential gene expression during the differentiation of spermatogonia into meiotic cells. Our approach utilized both custom cDNA microarrays and random oligonucleotide microarrays that contain known genes with diverse biological functions (including stress response, DNA repair, apoptosis, cell cycle, cellular differentiation, etc.) as well as unannotated sequences that may represent novel genes.

The custom cDNA microarray technology used in the basal gene expression studies (Chapter 3) was first optimized for the accurate quantitation of expression ratios by evaluating different hybridization and image analysis techniques. Various microarray hybridization strategies were investigated to increase signal intensity and reduce

background while maintaining hybridization specificity, including studies of the effects of (a) fluorescent probe quantity, (b) hybridization buffer composition, and (c) selection of slide blocking reagent. To provide accurate quantitation of hybridization signals, the microarray image capture system was calibrated by hybridizing a dilution series to the microarray and evaluating the following methods: (a) target spot segmentation, (b) quantitation, (c) background subtraction, (d) preprocessing, and (e) normalization. This work helped determine the optimal experimental conditions for accurately evaluating differential gene expression in subsequent biological studies.

The first biological objective of this dissertation research was to investigate the differential expression of stress response, damage control, and DNA repair genes among healthy adult mouse tissues using the optimized custom cDNA microarrays (Chapter 3). After assessing its reproducibility and precision, the cDNA microarray technology was used to (a) compare and contrast gene expression profiles among testis, brain, liver, spleen, and heart and (b) evaluate the differential tissue expression of genes with respect to their functions in different biological pathways. Microarray expression ratio measurements were compared with northern blot expression ratio measurements for selected genes in order to validate the microarray data. By characterizing *in vivo* differences in the baseline expression of stress response, damage control, and/or DNA repair-associated genes, this research provides insight on tissue-specific differences in the basal levels of cellular resources immediately available for responding to and processing DNA damage.

The second biological objective of this dissertation research was to investigate the temporal changes in gene expression that are associated with progressive stages of

cellular differentiation in the seminiferous tubules as germ cells proceed from mitotic spermatogonia to meiotic pachytene spermatocytes (Chapter 4). Random expression microarrays, which contain a diverse representation of known (annotated) genes and unannotated sequences, were used to establish gene expression profiles at specific times in this differentiation pathway (during spermatogonial mitosis, at the onset of meiosis, during mid-prophase of meiosis I (MI), and in the adult mouse to compare the first wave of spermatogenesis with full spermatogenesis). These profiles were compared in order to identify both annotated and novel genes whose expression was coordinately up- or down-regulated as germ cells differentiated through spermatogenesis. This specific research provides insight on the molecular mechanisms responsible for the onset and progression of meiosis and may also contribute to our understanding of the genetic causes of male infertility and the paternal transmission of mutations to offspring.

References

Allen, J.W., Dix, D.J., Collins, B.W., Merrick, B.A., He, C., Selkirk, J.K., Poorman-Allen, P., Dresser, M.E., and Eddy, E.M. (1996). HSP70-2 is part of the synaptonemal complex in mouse and hamster spermatocytes. *Chromosoma* *104*, 414-21.

Amundson, S.A., Bittner, M., Meltzer, P., Trent, J., and Fornace, A.J. (2001). Induction of gene expression as a monitor of exposure to ionizing radiation. *Radiat. Res.* *156*, 657-661.

Bailis, J.M., Smith, A.V., and Roeder, G.S. (2000). Bypass of a meiotic checkpoint by overproduction of meiotic chromosomal proteins. *Mol. Cell. Biol.* 20, 4838-48.

Ho, S.M. and Lau, K.M. (2002). DNA microarrays in prostate cancer. *Curr. Urol. Rep.* 3, 53-60.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318-56.

Jiang, Y., Harlocker, S.L., Molesh, D.A., Dillon, D.C., Stolk, J.A., Houghton, R.L., Repasky, E.A., Badaro, R., Reed, S.G., and Xu, J. (2002). Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays. *Oncogene* 21, 2270-82.

Koji, T., Jinno, A., Matsushime, H., Shibuya, M., and Nakane, P.K. (1992). In situ localization of male germ cell-associated kinase (mak) mRNA in adult mouse testis: specific expression in germ cells at stages around meiotic cell division. *Cell. Biochem. Funct.* 10, 273-9.

Matsuda, Y., Tobari, I., and Yamada, T. (1985). Studies on chromosome aberrations in the eggs of mice fertilized after irradiation. II. Chromosome aberrations induced in mature oocytes and fertilized eggs at the pronuclear stage following X-irradiation. *Mutat. Res.* 151, 275-80.

Niehrs, C. and Pollet, N. (1999). Synexpression groups in eukaryotes. *Nature* 204, 483-7.

Russell, W.L., Kelly, E.M., Hunsicker, P.R., Bangham, J.W., Maddux, S.C., and Phipps, E.L. (1979). Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci. USA* 76, 5818-9.

Schimenti, K.J., Hanneman, W.H., and Schimenti, J.C. (1997). Evidence for cyclophosphamide-induced gene conversion and mutation in mouse germ cells. *Toxicol. Appl. Pharmacol.* 147, 343-50.

Strachan, T. and Read, A.P. (1999). RNA transcription and gene expression. *Human Molecular Genetics*. 2nd Ed. New York: Wiley-Liss. pp. 9-14.

Toyooka, K., Usui, M., Washiyama, K., Kumanishi, T., and Takahashi, Y. (2002). Gene expression profiles in the brain from phencyclidine-treated mouse by using DNA microarray. *Ann. N.Y. Acad. Sci.* 965, 10-20.

Wolfes, H., Kogawa, K., Millette, C.F., and Cooper, G.M. (1989). Specific expression of nuclear proto-oncogenes before entry into meiotic prophase of spermatogenesis. *Science* 245, 740-3.

Zou, T.T., Selaru, F.M., Xu, Y., Shustova, V., Yin, J., Mori, Y., Shibata, D., Sato, F., Wang, S., Oлару, A., Deacu, E., Liu, T.C., Abraham, J.M., and Meltzer, S.J. (2002). Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene 21*, 4855-62.

Expression Microarray Technology

Chapter Overview

This chapter reviews the development and application of expression microarray technology (Section 2.1), discusses the optimization of cDNA microarray hybridizations for improved background-subtracted signal intensities (Section 2.2), and examines the effects of image acquisition and analysis on expression ratio measurements (Section 2.3). After determining the best methods for accurately quantitating expression ratios from fluorescent cDNA microarrays visualized using a white light image capture system, this research will be applied to the study of differential basal gene expression levels among healthy adult mouse tissues (Chapter 3).

2.1 Fluorescent expression microarray technology review

2.1.1 Introduction

Historically, gene expression studies have involved analyzing the expression of one gene, or a small number of genes, at a time. Over the past decade, the rapid incorporation of genome sequence information into publicly available databases, such as the National Center for Biotechnology Information's (NCBI) Entrez Nucleotide Query (<http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.html>) has facilitated the development of a new generation of high throughput methods for studying gene expression on a genome-wide scale, including serial analysis of gene expression (SAGE), PCR-based technologies (e.g., mRNA differential display), and expression microarrays. SAGE and the PCR-based technologies first identify transcripts that are differentially represented between experimental groups (i.e., cell or tissue samples) and then rely on downstream methods for gene identification and expression ratio quantitation. Microarrays quantitate the expression for up to tens of thousands of annotated genes or unannotated sequences that have been selected *a priori* for representation on the array.

To analyze gene expression using microarrays (transcript profiling), RNA is extracted from cells or tissue samples, fluorescently or radioactively labeled, and hybridized to a solid support that has been spotted with DNA (complementary DNA or oligonucleotide sequences) representing the genes of interest. Images of the hybridized array are captured, by either a white light system or a laser scanner, and the spot intensities are measured and then compared between samples in order to obtain gene expression ratios. Detailed descriptions of (a) the types of expression microarrays, (b)

probe generation, hybridization, and visualization, and (c) microarray data acquisition and analysis are contained in sections 2.1.2, 2.1.3, and 2.1.4, respectively.

The first microarray-based experiments were described by Schena et al. (1995) who applied complementary DNA (cDNA) microarray technology to the study of differential expression among 45 *Arabidopsis* genes. Using a dual color fluorescence *in situ* hybridization scheme, *Arabidopsis* mRNA was reverse transcribed in the presence of either fluorescein or lissamine and was hybridized to a microarray comprised of 96 targets (replicate spotting of 45 *Arabidopsis* cDNAs and 3 negative control cDNAs). This study showed that (1) fluorescently labeled probes could be hybridized to immobilized target cDNA with high specificity and (2) the differential expression of multiple genes could be quantitated in parallel, even for genes with low abundance transcripts.

Since their development in the mid-1990s, microarrays have become an increasingly popular tool for the analysis of differential gene expression. From January 2001 to January 2002 alone, over 900 microarray-related articles were indexed in PubMed, the database for biomedical literature sponsored by NCBI (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>). This technology has been used to address a broad range of research questions, including those related to (a) cancer (Khan et al., 1999) and other genetic diseases (e.g., Fragile X syndrome: Brown et al., 2001; Type 1 diabetes: Eaves et al., 2002); (b) host-pathogen interactions (e.g., comparative analysis of respiratory pathogens: Diehn and Relman, 2001; *Yersinia enterocolitica* infection: Sauvonnnet et al., 2002); (c) environmental and/or occupational exposures (Bartosiewicz et al., 2001); (d) development (Tanaka et al., 2000); (e) aging

(Weindruch et al., 2001); (f) behavior (Dent et al., 2001); (g) reproduction (e.g., Liu et al., 2001); and (h) pharmacology (e.g., therapeutic response: Chang et al., 2002; drug metabolism: Gerhold et al., 2001; drug discovery: Debouck and Goodfellow, 1999). Because this technology is easily adapted for studying any organism for which genome sequence information is available, it has been used to profile expression in diverse organisms including prokaryotes such as *Escherichia coli* (Oh and Liao, 2000), *Pseudomonas aeruginosa* (Ichikawa et al., 2000), and *Staphylococcus aureus* (Dunman et al., 2001) as well as lower eukaryotic organisms (e.g., *Saccharomyces cerevisiae*: Spellman et al., 1998). Expression microarrays have also been widely applied to the study of higher eukaryotes, including plants (e.g., *Arabidopsis*: Seki et al., 2001; loblolly pine: Whetten et al., 2001) and animals (e.g., mouse: Rockett et al., 2001; rat: Guo et al., 2000; and human: Schena et al., 1996).

2.1.2 Types of expression microarrays

Based on the type of “target” nucleotide sequence that is immobilized on the solid support (i.e., glass microscope slide or onto a nylon filter supported by a glass slide), there are two broad classifications of expression microarrays: cDNA or oligonucleotide. cDNA microarrays are generated by immobilizing expressed sequence tag (EST)-derived cDNA clones that are usually ~100 bp to 2 kb in size. Oligonucleotide microarrays are generated by immobilizing or synthesizing identified nucleotide sequences of approximately 25-100 bp in length that are unique to the transcript for the gene of interest.

cDNA and oligonucleotide microarrays are both generated by robotically arraying DNA sequences that represent the genes of interest directly onto the solid support. However, oligonucleotide microarrays have also been constructed by synthesizing DNA sequences directly on the solid support through processes such as photolithography (e.g., Affymetrix Expression Microarrays, Affymetrix, Inc., Santa Clara, CA).

2.1.3 Probe generation, hybridization, and image capture

Although specific protocols differ among laboratories, the overall experimental approaches for probe generation, hybridization, and visualization are generally very similar and include isolation of total RNA or mRNA, cDNA synthesis with fluorescent dye or radioactive label incorporation, hybridization (including pre-hybridization slide blocking, probe binding, and post-hybridization washes), and image capture. Figures 1 and 2 are schematic representations of the probe generation, hybridization, and visualization protocols used in our laboratory for the cDNA microarray study in Chapter 3 and the Affymetrix oligonucleotide microarray study in Chapter 4, respectively.

Probe generation

To generate probes for hybridization to the microarray, either total RNA or mRNA is isolated from cells or tissues and subsequently labeled with fluorophore-conjugated or radioactively-labeled dNTPs. Commonly, total RNA is isolated using guanidine isothiocyanate and phenol, and mRNA is isolated using an oligo-dT cellulose column [the polyA(+) tails of the mRNA bind to the oligo-dTs]. Microarray probes are

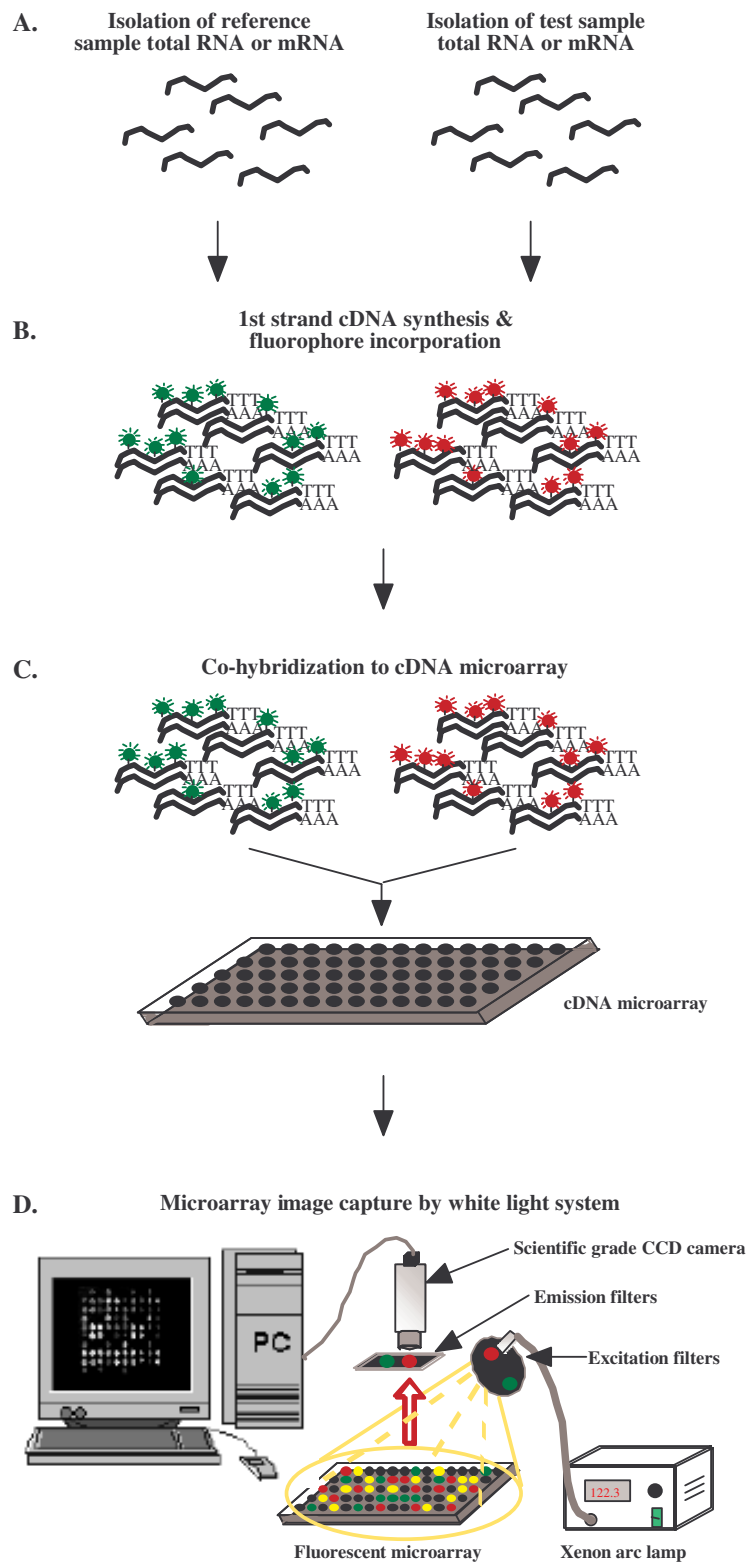


Figure 2.1.1. Experimental outline for cDNA microarray hybridizations. **(A)** Total RNA or mRNA is isolated from the cell or tissue source of interest. **(B)** Oligo-(dT) and an RNase H⁻ reverse transcriptase are used to incorporate fluorophore-conjugated dUTPs into 1st strand cDNA. **(C)** The two pools of labeled cDNA are co-hybridized onto the microarray. **(D)** Images of the hybridized array are captured using a white light imaging system.

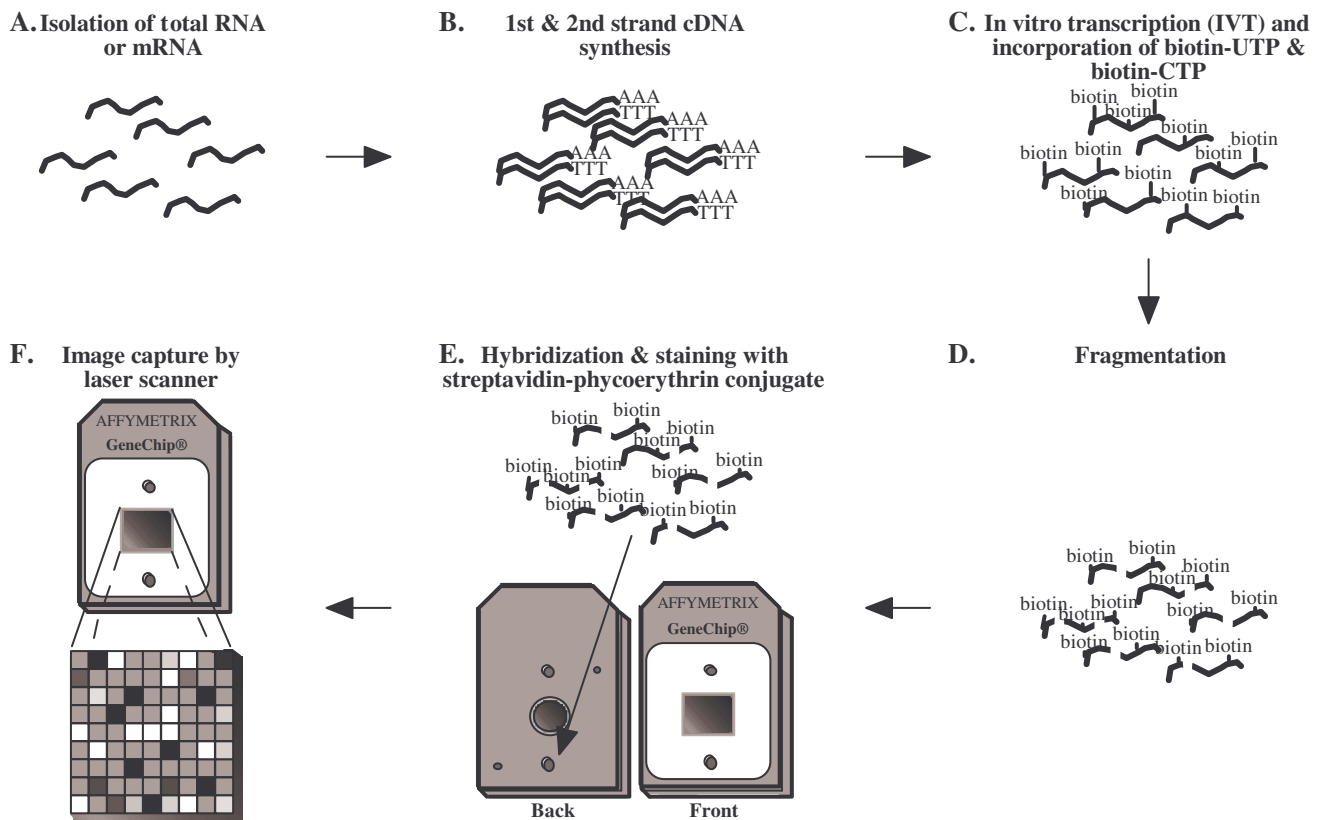


Figure 2.1.2. Experimental outline for Affymetrix oligonucleotide microarray hybridizations. **(A)** Total RNA or mRNA is isolated from the cell or tissue source of interest. **(B)** T7-(dT)₂₄ oligomer and an Rnase H⁻ reverse transcriptase are used to generate 1st strand cDNA. DNA polymerase I and DNA ligase are used in the presence of RNase H for 2nd strand cDNA synthesis. **(C)** T7 RNA polymerase is used to transcribe RNA *in vitro* while simultaneously incorporating biotin labeled UTPs and CTPs. **(D)** Transcribed RNA is fragmented in a magnesium acetate/ potassium acetate buffer. **(E)** Fragmented sequences are hybridized to the Affymetrix oligonucleotide expression arrays, and the hybridized sequences are stained with a streptavidin-phycoerythrin conjugate. **(F)** Images of the hybridized array are captured using a laser scanner.

labeled by incorporating fluorophore-conjugated dUTPs or dCTPs during either reverse transcription or nick translation. (Although radioactive labels have been used for nylon-based microarrays, most commercial and non-commercial microarray systems involve the hybridization of fluorescently labeled probes to glass slides. Therefore, this review focuses primarily on fluorescent microarray hybridizations.)

Probe generation techniques vary depending on the cell- or tissue-type or size of the tissue sample. Total RNA is usually isolated from smaller samples in order to prevent the loss of mRNA. In addition the isolated RNA may need to be linearly amplified through *in vitro* transcription in order to obtain enough RNA for hybridization to the microarray. The type of array being hybridized (cDNA or oligonucleotide) also contributes to variation in probe generation techniques between laboratories. For oligonucleotide microarrays, the probes must be fragmented prior to hybridization, but this step is not necessary for cDNA microarray hybridizations. Also, the characteristics of the microarray imaging system (e.g., wavelengths of the excitation and emission filters) must be matched to the type of fluorophore to be incorporated into the probe.

Hybridization

Typically, all microarray probe hybridization protocols include: a slide blocking procedure (which reduces non-specific binding of the fluorescently labeled probe mixture to areas outside of the arrayed spots); use of a special hybridization buffer; and post-hybridization washing techniques to further reduce non-specific probe binding.

Several reagents have been found to prevent the non-specific binding of the probes to the glass slide during *in situ* hybridizations (e.g., ammonium hydroxide, bovine

serum albumin, tryptone, acetic anhydride, etc.). In our experience, selection of the appropriate blocking reagent for microarray experiments depends upon the system used for image capture because the magnitude of the background intensity is directly affected by the combination of blocking reagent, excitation source (i.e., white light or laser), and excitation/emission filters.

Microarray hybridization stringency is determined by the combination of hybridization buffer, the length of time for hybridization, and hybridization temperature. More specifically, the composition of the hybridization buffer directly affects the length of time and temperature of the hybridization. The presence of large polymers (e.g., dextran sulfate) in the buffer increases the rate of reassociation and therefore decreases the length of time necessary for hybridization. The environmental temperature required for hybridization is also affected by whether or not the buffer contains formamide. The presence of formamide increases the effective hybridization temperature, (calculated as the environmental temperature + 0.4-0.6 times the % formamide), and, therefore, is usually determined prior to hybridization in order to maintain the correct level of hybridization stringency (i.e., allows the probe to hybridize, but prevents non-specific probe hybridization). Additional information on the optimization of fluorescent cDNA microarray hybridizations is contained in Section 2.2.

Post-hybridization washes usually involve using sodium phosphate buffers (i.e., PN), saline sodium citrate (SSC), and/or DNase-free water to remove excess probe and non-specific probe hybridization. In general, we have found it preferable to have high stringency hybridizations so that less rigorous washing procedures (which may affect the intensity of the fluorescently-labeled probe and, therefore, also lower the dynamic range

and ultimately affect expression ratio quantitation) are required. The concentrations of the wash solutions as well as the temperature of the washes are inversely dependent upon the stringency of the hybridization (i.e., the higher the hybridization stringency, the lower the wash stringency and vice versa).

Image capture

Following hybridization, the fluorescent microarray images are visualized using either a white light system or a laser scanner. White light systems are easily adapted for imaging a broad range of fluorophores, including those detected in the UV bandwidth. However, laser scanners are most commonly used for image capture because they avoid data analysis problems associated with non-uniform illumination. In addition, laser scanners are able to provide better resolution than the white light systems (~5 μm vs. ~15 μm).

Our white light image capture system (Figure 1D; Kegelmeyer et al., 2001) utilizes light (400 nm – 600 nm) from a source such as a Xenon arc lamp. The light is usually scrambled through a fiber optic cable and passed through excitation and emission filters appropriate for the fluorophore used to label the probe. The emitted light is captured by a scientific grade CCD camera connected to a personal computer (PC). The PC controls the camera and collects the microarray images for analysis. Integration time and gain can be adjusted by the user during image capture in order to maximize the signal without pixel saturation.

Laser image capture systems sequentially scan hybridized arrays with focused gas lasers (Application note from Packard Bioscience Company, Meriden, CT;

http://www.packardbioscience.com/reference_matl/827.asp). Briefly, the laser is reflected from a beamsplitter through an objective lens and onto the microarray. The resulting fluorescence is converted into parallel beams that travel back through the beamsplitter and through an emission filter to a detector lens. The detector lens focuses the beam onto the detector for image capture.

2.1.4 Microarray data acquisition and analysis

After capturing the microarray images, data acquisition and analysis software is used to extract signal intensity data from the images and generate expression ratios. A large number of microarray data acquisition and analysis software packages are currently available (e.g., ImaGene from Biodiscovery, Inc., Marina del Ray, CA and QuantArray from Packard Bioscience Company). Most acquisition and analysis software allows the user to select from a variety of parameters including methods for quantitation, color correction (for dual color hybridizations), background subtraction, normalization, and sample-to-sample or slide-to-slide comparisons. The selection of specific software varies depending upon the type of microarray (cDNA or oligonucleotide), imaging system, and computer platforms (i.e., PC, Macintosh, or UNIX) utilized. Consideration must also be given to the type of downstream analysis to be performed so that the formatted microarray data can be easily uploaded into bioinformatics databases, such as those for hierarchical clustering (e.g., CLUSFAVOR: Peterson, 2002), promoter analysis (e.g., PromoterInspector by Genomatix: Scherf et al., 2000), and biological pathway analysis (e.g., GenMAPP: Dahlquist et al., 2002). Furthermore, the design of microarray

experiments is critical (i.e., number of experimental groups; number of replicates per group; type of data analysis; selection of data to be reported in the literature and/or in public databases; and the type of downstream analysis). There are useful recommendations given by the Microarray Gene Expression Data Group (<http://www.mged.org/> or <http://industry.ebi.ac.uk/~alan/MicroArray/>), which continues to revise its guidelines based on input from the scientific community.

2.1.5 Summary

Although they have undeniably become a prominent tool in biological research since their first description in 1995, expression microarray technology varies greatly among laboratories (different: types of microarrays, probe generation and hybridization techniques, image capture systems, and data analysis methods) and the field continues to mature. Standards for conducting microarray experiments, as well as analyzing and reporting the voluminous data, are constantly being debated and revised. However, because this technology is maturing in concordance with a rapid increase in the amount of publicly accessible genome sequence information for both prokaryotic and eukaryotic organisms, it is expected that microarray (and related genome-scale technologies) usage will be even more prevalent in the future.

2.1.6 References

Bartosiewicz, M., Penn, S., and Buckpitt, A. (2001). Applications of gene arrays in environmental toxicology: fingerprints of gene regulation associated with cadmium chloride, benzo(a)pyrene, and trichloroethylene. *Environ. Health Perspect.* *109*, 71-4.

Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D., Darnell, R.B., and Warren, S.T. (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* *107*, 477-87.

Chang, B.D., Swift, M.E., Shen, M., Fang, J., Broude, E.V., and Roninson, I.B. (2002). Molecular determinants of terminal growth arrest induced in tumor cells by a chemotherapeutic agent. *Proc. Natl. Acad. Sci. USA* *99*, 389-94.

Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C., and Conklin, B.R. (2002). GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.* *31*, 19-20.

Debouck, C. and Goodfellow, P.N. (1999). DNA microarrays in drug discovery and development. *Nat. Genet.* *21(1 Suppl)*, 38-41.

Dent, G.W., O'Dell, D.M., and Eberwine, J.H. (2001). Gene expression profiling in the amygdala: an approach to examine the molecular substrates of mammalian behavior. *Physiol. Behav.* *73*, 841-7.

Diehn, M. and Relman, D.A. (2001). Comparing functional genomic datasets: lessons from DNA microarray analyses of host-pathogen interactions. *Curr. Opin. Microbiol.* *4*, 95-101.

Dunman, P.M., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., Brown, E.L., Zagursky, R.J., Shlaes, D., and Projan, S.J. (2001). Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the agr and/or sarA loci. *J. Bacteriol.* *183*, 7341-53.

Eaves, I.A., Wicker, L.S., Ghandour, G., Lyons, P.A., Peterson, L.B., Todd, J.A., and Glynne, R.J. (2002). Combining mouse congenic strains and microarray gene expression analyses to study a complex trait: the NOD model of type 1 diabetes. *Genome Res.* *12*, 232-43.

Gerhold, D., Lu, M., Xu, J., Austin, C., Caskey, C.T., and Rushmore, T. (2001). Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol. Genomics* *5*, 161-70.

Guo, Q.M., Malek, R.L., Kim, S., Chiao, C., He, M., Ruffy, M., Sanka, K., Lee, N.H., Dang, C.V., and Liu, E.T. (2000). Identification of c-myc responsive genes using rat cDNA microarray. *Cancer Res.* *60*, 5922-8.

Ichikawa, J.K., Norris, A., Banger, M.G., Geiss, G.K., van't Wout, A.B., Bumgarner, R.E., and Lory, S. (2000). Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc. Natl. Acad. Sci. USA* *97*, 9659-64.

Kegelmeyer, L.M., Tomascik-Cheeseman, L., Burnett, M.S., van Hummelen, P., and Wyrobek, A.J. (2001). A groundtruth approach to accurate quantitation of fluorescence microarrays. *SPIE Proceed.* *4266*, 35-45.

Khan, J., Saal, L.H., Bittner, M.L., Chen, Y., Trent, J.M., and Meltzer, P.S. (1999). Expression profiling in cancer using cDNA microarrays. *Electrophoresis* *20*, 223-9.

Liu, H.C., He, Z., and Rosenwaks, Z. (2001). Application of complementary DNA microarray (DNA chip) technology in the study of gene expression profiles during folliculogenesis. *Fertil. Steril.* *75*, 947-55.

Oh, M.K. and Liao, J.C. (2000). Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol. Prog.* *16*, 278-86.

Peterson, L.E. (2002). Factor analysis of cluster-specific gene expression levels from cDNA microarrays. *Comput. Methods Programs Biomed.* In press.

Rockett, J.C., Luft, C.J., Garges, B.J., Krawetz, S.A., Hughes, M.R., Kim, H.K., Oudes, A.J., and Dix, D.J. (2001). Development of a 950-gene DNA array for examining gene expression patterns in mouse testis. *Genome Biol.* 2, 0014.1-0014.9.

Sauvonnet, N., Pradet-Balade, B., Garcia-Sanz, J.A., and Cornelis, G.R. (2002). Regulation of mRNA expression in macrophages following *Yersinia enterocolitica* infection: role of different Yop effectors. *J. Biol. Chem.* In press.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-70.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93, 10614-9.

Scherf, M., Klingenhoff, A., and Werner, T. (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* 297, 599-606.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y., and Shinozaki, K. (2001). Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13, 61-72.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9, 3273-97.

Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W.H., Becker, K.G., and Ko, M.S. (2000). Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc. Natl. Acad. Sci. USA* 97, 9127-32.

Weindruch, R., Kayo, T., Lee, C.K., and Prolla, T.A. (2001). Microarray profiling of gene expression in aging and its alteration by caloric restriction in mice. *J. Nutr.* 131, 918S-23.

Whetten, R., Sun, Y.H., Zhang, Y., and Sederoff, R. (2001). Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Mol. Biol.* 47, 275-91.

2.2 Optimization of two-color fluorescence cDNA microarray hybridizations visualized with a white light imaging system

2.2.1 Abstract

The major parameters that affect the quantitation and interpretation of gene expression from cDNA microarrays are microarray quality, probe preparation, hybridization conditions, image capture, and methods of analysis. This research focused on the technological development of microarray hybridization strategies to increase signal intensity while decreasing background. Specifically, gene-specific and tissue sample probes were hybridized to cDNA microarrays to evaluate the effects of probe quantity (25 μg vs. 50 μg), hybridization buffer (commercial vs. LLNL formulas), and slide blocking procedure (ammonium hydroxide vs. tryptone) on hybridization signal intensities. The results show that improved signal intensities were obtained when lower probe amounts were hybridized in a buffer comprised of 42% formamide / 2 x SSC / 10% dextran sulfate to microarrays blocked with 0.25% tryptone in water. These conditions were applied to the study of differential basal expression levels among mouse tissues which is described in Chapter 3.

2.2.2 Introduction

Maximizing signal to background ratios is critical for obtaining accurate measurements of gene expression from cDNA microarrays. The magnitude of the signal

remaining after background subtraction is dependent upon several factors, including the quality of (a) the microarray (e.g., no inherent defects in the solid support and no degradation of target DNA spots), (b) the probe (e.g., efficient fluorophore incorporation and prevention of photobleaching), (c) the hybridization (e.g., correct hybridization stringency and blocking of non-specific probe binding), (d) the image capture system (e.g., use of fluorophore-appropriate, narrow bandpass excitation and emission filters and the ability to measure differences in exposure time between samples), and (e) the data acquisition software (e.g., ability to account for: the measured exposure differences, non-uniform illumination, spectral crosstalk, etc.).

Two major aspects of microarray fabrication, preparation of the solid support and preparation of the target spot cDNA, can affect background and must be controlled to obtain accurate gene expression measurements. To accomplish this in our experiments, the glass microscope slides used as the solid supports (a) were rigorously cleaned using concentrated acids such as HCL and H₂SO₄; (b) had a uniform distribution of the chemical used for cDNA attachment to the solid support (e.g., poly-L-lysine or aminopropyltrimethoxysilane); and (c) were inspected immediately prior to spotting to ensure that the area to be arrayed is free of chips or cracks. Additionally, the cDNA to be spotted onto the array was placed in buffers that facilitated cDNA attachment without compromising cDNA integrity (e.g., sodium carbonate/bicarbonate) and arrayed under environmental conditions (i.e., temperature and humidity) that allowed the spots to dry quickly without spreading into neighboring spots.

Probe quality has a large impact on signal to background ratios and, therefore, also on the accurate quantitation of expression. In our experiments, fluorescently labeled

probes were generated by incorporating both unlabeled dNTPs and fluorophore-conjugated dUTPs or dCTPs during a reverse transcription reaction. The appropriate balance of labeled and unlabeled dNTPs had to be determined, because if too many fluorophore-conjugated dNTPs were incorporated in succession, the reverse transcriptase could stall and fall off of the transcript. Irrespective of the dNTP that was selected, the fluorophores have a substantial effect on the signal to background ratios. The most common fluorescent dyes currently used for microarray experiments are the Cyanine dyes (Cy3 and Cy5) and the Alexa Fluor dyes (Alexa Fluors 488 and 546). Initially, the Cy dyes were used for microarray experiments due to the limited wavelengths available for laser scanners. However, white light imaging systems (and a few recently developed laser scanners) are capable of detecting a broad range of fluorophores, including the Alexa Fluor dyes. We found that the Alexa Fluor 488 and 546 dyes (which have excitation and emission spectra similar to FITC and Cy3, respectively) were preferable to the Cy dyes because they are brighter, more photostable, and less sensitive to alterations in pH. Also, when used for dual color hybridizations, these Alexa Fluor dyes have less spectral overlap with each other compared to Cy3 and Cy5. Regardless of the fluorophore selected, the use of antifades typically provided better signal to background ratios because they preserve signal intensity by preventing photobleaching during the image capture procedure. Unfortunately, antifades can only be used when images are captured by white light imaging system because laser scanners require dry slides.

Hybridization strategies (e.g., increasing probe concentration and the rate of DNA reassociation by the addition of large polymers such as dextran sulfate and polyethylene glycol; Wetmur, 1971) also have a significant impact on signal to background ratios.

Stringent hybridization conditions are critical for accurate expression ratio measurements because they increase probe binding while reducing non-specific probe hybridization. In our experiments, stringency was optimized by matching the probe hybridization buffer with hybridization temperature and time conditions. This required the use of the proper combinations of formamide (which increases the effective hybridization temperature; Miesfeld, 1999) and salt concentration (which stabilizes the hybridized probe; Strachan and Read, 1999) for a given hybridization temperature and time.

To obtain accurate gene expression information from microarray images using a white light image capture system, we ensured that the excitation and emission filters had narrow, non-overlapping bandpasses (to circumvent/reduce spectral cross-talk) with wavelengths that were suitable for the excitation and emission spectra of the fluorophore used in probe labeling. For additional information, see Section 2.3 “Accurate quantitation of fluorescence microarrays”. Improvements to background-subtracted signal intensities were also made in our experiments by correcting for camera dark noise, non-uniform illumination, spectral cross-talk, and integration time differences. In addition, the selection of segmentation and quantitation methods was dependent upon the target spot shapes and the array grid layout (i.e., horizontal x vertical number of spots and their position within each row and column) as well as the area hybridized on individual spots (e.g., accurate measurements for irregularly shaped spots or spots that only hybridize along the edges). The selection of data acquisition software that included flexible background subtraction methods for hybridizations with either high or low background was also important for accurately measuring expression ratios. (See Section 2.3.).

As described above, there are many parameters (and combinations of these parameters) that may affect the magnitude of background-subtracted microarray signal intensities. The purpose of this research was to develop cDNA microarray hybridization protocols that increased signal to background ratios by concurrently improving signal intensity and reducing background intensity. Probe quantity, hybridization buffer, and microarray slide blocking procedures were evaluated for hybridizations of gene-specific (simple) and/or tissue sample (complex) probes.

2.2.3 Materials and Methods

cDNA microarray preparation

Two custom cDNA microarrays were spotted: the microarray used for gene-specific hybridizations was comprised of 72 genes (322 target spots) and the microarray used for the tissue sample hybridizations was comprised of 53 genes (252 target spots). For each gene, one to four cDNA clones were obtained from the I.M.A.G.E. Consortium at LLNL. Up to five lambda phage and bacterial sequences served as controls. Clones were PCR-amplified with 5'-C6 amino-modified vector-specific primers, purified on Qiagen purification columns (Qiagen Inc., Valencia, CA), ethanol precipitated, and resuspended in 0.1M sodium carbonate/bicarbonate (pH = 10.2) for spotting onto the microarray slides.

The glass slides used for arraying were cleaned for 30 min. in 1:1 concentrated hydrochloric acid:methanol, soaked overnight in concentrated sulfuric acid, and washed 10 x 10 min. in room temperature water and 1 x 10 min. in boiling water. Slides were

then submerged in 1% 3-aminopropyltrimethoxysilane (Sigma, St. Louis, MO) / 95% acetone/ water for 2 min., washed 10 x 5 min. in acetone, and heated at 110°C for 45 min. (Guo et al., 1994). Silane-coated slides were submerged in 0.2% 1,4-phenylene diisothiocyanate (Sigma), 10% pyridine (Sigma) and dimethylformamide (Aldrich, Milwaukee, WI) for 2 hours, washed in methanol (2 x 10 min.) and acetone (2 x 10 min.), and air-dried. Target spots for both the gene-specific (simple) and tissue sample (complex) hybridizations were robotically arrayed (Norgren Systems, Palo Alto, CA) in duplicate with a 250 µm center-to-center distance.

Probe generation

Gene-specific (simple) probes were generated for each of the following genes: *Cdc2*, *Eif-4c*, *Lig1*, *Tp53*, and *Xrcc1*. A cDNA clone representing each gene was obtained from the I.M.A.G.E. Consortium at LLNL and was PCR-amplified with gene-specific primers designed using the Oligo v4.0 primer design software (National Biosciences, Inc., Plymouth, MN; Table 2.2.1). The expected amplicon sizes (range: 150-470 bp) were verified by electrophoresis on a 2% agarose gel. The quantity of each PCR-product was determined using the GeneQuant spectrophotometer (Amersham Pharmacia Biotech, Piscataway, NJ). To incorporate the fluorescent label (Alexa Fluor 488-dUTP for *Cdc2* and *Tp53*; Alexa Fluor 546-dUTP for *Eif-4c*, *Lig1*, and *Xrcc1*; Molecular Probes Inc., Eugene, OR), 0.5 µg of each product was nick translated using the Nick Translation System from Life Technologies (Rockville, MD). A master mixture containing all of the fluorescently labeled genes was prepared and then divided equally,

Table 2.2.1. Primers used to amplify cDNA clones for the gene-specific hybridization experiments.

Gene symbol	Gene name	I.M.A.G.E. clone	Primers (5' 3') →	Estimated amplicon size (bp)	
				Expected	Actual
<i>Cdc2</i>	Cell division cycle 2	763260	TTTGG AATACCGATACGAGT CGACCAGCAGACAGGGACAT	470	450
<i>Eif-4c</i>	Elongation initiation factor 4C	959700	AAGAAGTCTGAAGGCCTATG CAGAGAACTTGGAATGTAGC	150	150
<i>Lig1</i>	Ligase I	605700	ATGCAAGCTGGGAACTGGATT TGAACCGAGGAAAACGAAGAG	250	230
<i>Tp53</i>	Tumor suppressor protein 53	464741	AAGTGAAGCCCTCCGAGTGT CCATAGTTGCCCTGGTAAGT	150	150
<i>Xrcc1</i>	X-ray repair complementing defective repair	1022963	GACTGTCACCACATGCGGCG GGCTGCCTTTGTTCCCTCTG	350	300

such that each slide was hybridized with ~10 ng of labeled DNA per gene. The labeled probes were co-purified on Qiagen columns (Qiagen, Inc.) and ethanol precipitated.

Tissue sample (complex) probes were generated by synthesizing first strand cDNA from 25-50 µg of adult mouse testis or brain total RNA (Clontech, Inc., Palo Alto, CA). Specifically, testis and brain total RNA were reverse transcribed at 42°C using an oligo-dT primer in the presence of Superscript II reverse transcriptase (Life Technologies). Testis and brain cDNA were labeled during the reverse transcription reaction by incorporation of Alexa Fluor 488-dUTPs and Alexa Fluor 546-dUTPs (Molecular Probes, Inc.), respectively. The labeled probes were co-purified on Qiagen columns (Qiagen, Inc.) and ethanol precipitated.

Hybridization

To prevent non-specific probe binding (and thereby reduce background intensity), slides were blocked prior to hybridization with ammonium hydroxide (Sigma, St. Louis, MO) or tryptone (Difco Laboratories, Detroit, MI). Slides blocked with ammonium hydroxide were incubated at room temperature in 1% NH₄OH in water for 10 min., washed 3 x 10 min. in double distilled water, and air-dried. To block with 0.25% or 0.50% tryptone in water, the slides were submerged in the solution and agitated at 100 rpm for 1 hour and then immediately denatured. Regardless of the blocking procedure used, slides were denatured for 6 minutes in 70% formamide / 2xSSC / water at 78°C, passed through a 70% / 85% / 100% ethanol series, and air-dried. An *in situ* frame (Eppendorf, Westbury, NY) was affixed to the slide in order to confine the probe mixture over the arrayed area.

Table 2.2.2. Experimental design for the gene-specific and tissue sample hybridizations.

<u>Hybridization probe</u>	<u>Probe quantity^a</u>	<u>Hybridization buffer</u>	<u>Slide blocking reagent</u>
<u>Gene-specific</u>	50 ng	Commercial ^b	1% NH ₄ OH ^b
	50 ng	LLNL	1% NH ₄ OH
	50 ng	LLNL	0.50% Tryptone
	50 ng	LLNL	0.25% Tryptone
<u>Tissue sample</u>	50 μg ^b	Commercial ^b	1% NH ₄ OH ^b
	25 μg	Commercial ^b	1% NH ₄ OH
	50 μg	LLNL	0.50% Tryptone
	25 μg	LLNL	0.25% Tryptone

^aGene-specific hybridizations: total quantity of DNA probe; Tissue sample hybridizations: quantity of total RNA used for probe generation

^bProbe quantity, hybridization buffer, and slide blocking reagent traditionally used in our laboratory for microarray hybridizations

The specific combination of probe quantity, hybridization buffer, and slide blocking procedure used for each hybridization is shown in Table 2.2.2. For the gene-specific hybridizations, each slide was hybridized with 15 μ l of probe mixture (10 μ l hybridization buffer + 1 μ l herring sperm DNA + 4 μ l labeled probe resuspended in double distilled water). For the tissue-specific hybridizations, each slide was also hybridized with 15 μ l of probe mixture; however, the detergent SDS was added to promote probe movement across the array, and poly (A)+ (Amersham Pharmacia Biotech) as well as cot-1 mouse DNA (Life Technologies) were added to increase hybridization specificity. Both hybridization buffers contained formamide and were either commercially obtained (Vysis, Inc., Downers Grove, IL) or made at LLNL. The LLNL hybridization buffer was prepared by incubating 4.2 ml of formamide, 1.05 ml of 20 x SSC, and 1 g of dextran sulfate overnight at 65°C. Following incubation, water was added to bring the hybridization buffer volume to 7 ml. Probes prepared with the LLNL hybridization buffer had a pH = 7.0 and a final concentration of 42% formamide, 2 x SSC, and 10% dextran sulfate.

All hybridization mixtures were denatured at 78°C for 6 min. and snap-cooled on ice. Slides were hybridized for 12 hours in a gently rocking moist chamber at 37°C and washed 2 x 2 min. in 42°C water. Prior to image capture, Vectashield mounting media (Vector Laboratories, Inc., Burlingame, CA) was placed over the arrayed area to prevent photobleaching.

Image capture and analysis

Image capture and processing was performed as described by Kegelmeyer et al. (2001). Briefly, images were acquired with a full-field white light imaging system. Arrays were exposed to bandpass-filtered excitation light from a Xenon arc lamp. The resulting emitted light was bandpass-filtered and collected by a scientific-grade CCD camera. All images used for the gene-specific experiments were captured using comparable integration time and gain settings. This was also true for the set of tissue sample experiments. Custom algorithms, built within SCIL-Image (Delft, The Netherlands), corrected for CCD dark noise, spectral cross-talk, misaligned images and integration time variation. Additional processing algorithms determined the “segmentation mask” for each cDNA spot. The red and green intensities for all spots were computed by taking the median of all pixels within the segmentation mask that were greater than zero after background subtraction. Spots covered by debris were eliminated from all subsequent processing.

2.2.4 Results

Effects of hybridization buffer and slide blocking procedure on signal intensity for gene-specific (simple probe) hybridizations

Five genes, labeled with either Alexa Fluor 488 or Alexa Fluor 546, were co-hybridized to the 72 gene microarray to evaluate the effects of hybridization buffer and slide blocking on the background-subtracted median signal intensities. Figure 2.2.1 shows the median background-subtracted signal intensities (\pm standard error for all target

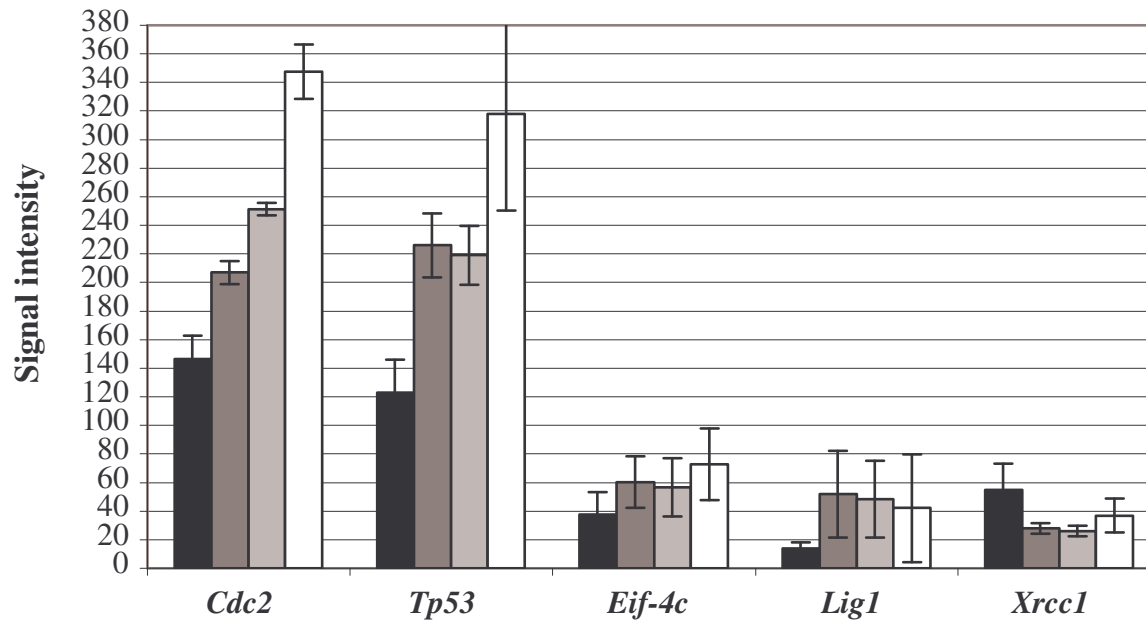


Figure 2.2.1. Effect of hybridization buffer and slide blocking protocol on background-subtracted microarray signal intensities obtained for the gene-specific probes. The x-axis represents the genes selected for co-hybridization to the cDNA microarray. *Cdc2* and *Tp53* were labeled with Alexa Fluor 488. *Eif-4c*, *Lig1*, and *Xrcc1* were labeled with Alexa Fluor 546. The y-axis represents the median background subtracted signal intensity. Each column represents a different combination of hybridization buffer and slide blocking protocol: black – commercial fluorescent probe hybridization buffer and 1% ammonium hydroxide blocking reagent; dark gray – LLNL hybridization buffer and 1% ammonium hydroxide blocking reagent; light gray – LLNL hybridization buffer and 0.50% tryptone blocking reagent; and white - LLNL hybridization buffer and 0.25% tryptone blocking reagent. Bars represent the standard error among the replicate spots used to represent each gene.

spots representing the gene of interest) obtained from each hybridization buffer and slide blocking combination listed in Table 2.2.2. The sum of the median intensities (calculated as all Alexa Fluor 488 probe intensities + all Alexa Fluor 546 probe intensities per hybridization condition) for the commercial buffer / 1% NH₄OH, LLNL buffer / 1% NH₄OH, LLNL buffer / 0.50% tryptone, and LLNL buffer / 0.25% tryptone hybridization conditions were approximately 375, 572, 601, and 817, respectively. Based on the average intensity values (average Alexa Fluor 488 vs. average Alexa Fluor 546) within each set of hybridization conditions, probes labeled with Alex Fluor 488 had 3.8 to 6.6-fold higher intensities than those labeled with Alexa Fluor 546.

To further compare the effects of hybridization buffer and slide blocking protocol on signal intensity, the percent increases in signal intensities, obtained after using different hybridization methods, were calculated (Table 2.2.3). The results showed that, irrespective of the blocking reagent utilized, the LLNL hybridization buffer yielded signal intensities that were ~41 to 275% higher than those obtained using the commercial hybridization buffer. Next, the effect of slide blocking reagent on signal intensity was determined for all slides hybridized using the LLNL buffer. From these comparisons, it was determined that the 0.50% tryptone and 1% NH₄OH reagents gave similar results and that use of the 0.25% tryptone blocking reagent resulted in signal intensities that were up to ~45% higher than those from 0.50% tryptone-blocked slides and up to ~68% higher than those from 1% NH₄OH-blocked slides. Furthermore, signal intensities for both of the LLNL buffer/tryptone combinations ranged from ~51 to 250% higher than those obtained using the commercial buffer / 1% NH₄OH hybridization procedure, with the exception of *Xrcc1*.

Table 2.2.3. Comparison of background-subtracted signal intensities obtained from gene-specific hybridizations using different combinations of hybridization buffer and slide blocking reagent.

Genes	Relative percent increase in signal intensity ^a					
	Effect of hybridization buffer ^b	Effect of slide blocking reagent ^c			Combined effect of hybridization buffer and slide blocking reagent	
	LLNL vs. Commercial	0.50% tryptone vs. 1% NH ₄ OH	0.25% tryptone vs. 1% NH ₄ OH	0.25% tryptone vs. 0.50% tryptone	(LLNL & 0.50% tryptone) vs. (Commercial & 1% NH ₄ OH)	(LLNL & 0.25% tryptone) vs. (Commercial & 1% NH ₄ OH)
<i>Cdc2</i>	41.3	21.4	67.9	38.3	71.5	137.2
<i>Eif-4c</i>	60.8	NI ^d	20.7	28.8	50.7	94.1
<i>Lig1</i>	275.4	NI ^d	NI ^d	NI ^d	250.0	204.3
<i>Tp53</i>	84.0	NI ^d	40.7	45.2	78.3	159.0
<i>Xrcc1</i>	NI ^d	NI ^d	32.4	41.5	NI ^d	NI ^d

^aPercent increase in the background subtracted signal intensity for condition 1 when compared to condition 2. For example, the signal intensity for *Cdc2* increased by 41.3% when the LLNL hybridization buffer was used instead of the commercial hybridization buffer.

^bSlide blocking reagent: 1% NH₄OH

^cLLNL hybridization buffer. See Materials and Methods for details.

^dNI = no increase in signal intensity was observed

Effects of probe quantity, hybridization buffer, and slide blocking procedure on signal intensity for tissue sample (complex probe) hybridizations

Mouse tissue samples were labeled with either Alexa Fluor 488 (testis) or Alexa Fluor 546 (brain) and co-hybridized to the 53 gene microarray in order to evaluate the combined effects of probe quantity, hybridization buffer, and slide blocking protocol on the background-subtracted median signal intensities for complex probe hybridizations. As shown in Panel A of Figure 2.2.2, probe quantities of 25 μg tended to yield higher Alexa Fluor 488 intensities, regardless of the hybridization buffer and slide blocking reagent utilized. However, the same trend was not observed for the Alexa Fluor 546 labeled probes (Panel B) which generally had the highest intensities when 25 μg of probe was hybridized using the LLNL buffer to microarray slides blocked with 0.25% tryptone. The remaining three probe quantity / hybridization buffer / slide blocking reagent combinations had similar (lower) signal intensities.

As shown in Table 2.2.4, the maximum signal intensities obtained for Alexa Fluor 488 probes ranged from 50.3 for the 50 μg / commercial buffer / 1% NH_4OH hybridization to 237.0 for the 25 μg / LLNL buffer / 0.25% tryptone hybridization. Minimum signal intensities, however, were similar across all procedures (range: 9.0 to 15.5). In general, hybridizations using 25 μg of probe had higher intensities than the 50 μg probe hybridizations. The highest and lowest average signal intensities were observed for the 25 μg / commercial buffer / 1% NH_4OH hybridization (40.9) and the 50 μg / LLNL buffer / 0.50% tryptone hybridization (14.2), respectively. In fact, the 25 μg / commercial buffer / 1% NH_4OH protocol had even the highest signal intensity at the 75th percentile. The maximum signal intensities for probes labeled with Alexa Fluor 546

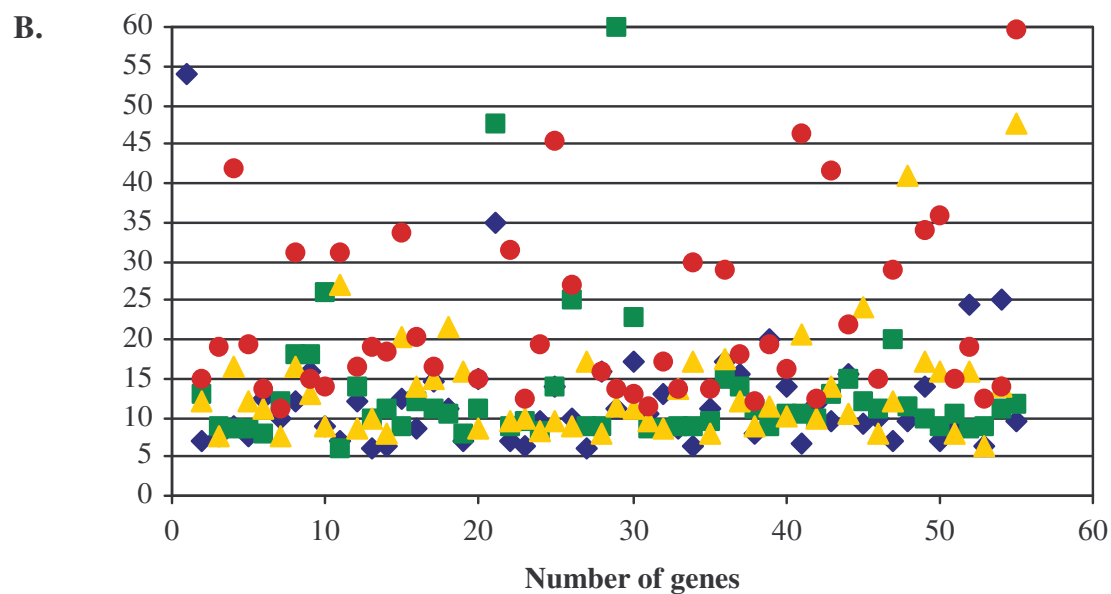
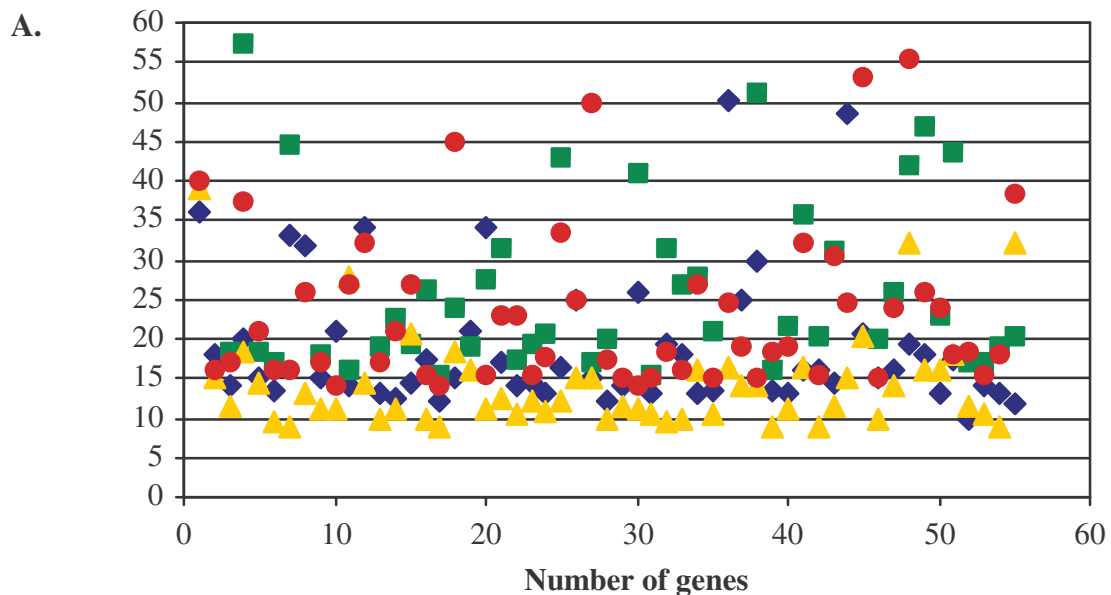


Figure 2.2.2. Distribution of background-subtracted median signal intensities obtained for tissue sample hybridizations performed under different conditions. The x-axis represents the number of genes evaluated. The y-axis represents the median background subtracted signal intensity. The combinations of probe quantity, hybridization buffer, and slide blocking protocol are represented as follows: blue diamond – 50 µg probe / commercial buffer / 1% NH₄OH blocking reagent; green square – 25 µg probe / commercial buffer / 1% NH₄OH blocking reagent; yellow triangle – 50 µg probe / LLNL buffer / 0.25% tryptone blocking reagent; and red circle – 25 µg probe / LLNL buffer / 0.25% tryptone blocking reagent. A.) Distribution of Alexa Fluor 488 intensities. B.) Distribution of Alexa Fluor 546 intensities.

Table 2.2.4. Characterization of the signal intensities obtained for each Alexa Fluor following different tissue sample hybridization protocols.

	Signal intensity distributions			
	50 µg / commercial / NH₄OH^a	25 µg / commercial / NH₄OH^a	50 µg / LLNL / 0.50% tryptone^a	25 µg / LLNL / 0.25% tryptone^a
Alexa Fluor 488				
Percentile				
Minimum	10.0	15.5	9.0	14.0
25th	13.5	19.0	10.5	16.0
50th	15.0	26.0	12.0	19.0
75th	20.4	45.8	16.0	27.0
Maximum	50.3	162.5	39.0	237.0
Average	19.0	40.9	14.2	27.3
Alexa Fluor 546				
Percentile				
Minimum	6.0	6.0	6.5	11.0
25th	7.5	9.0	9.0	15.0
50th	10.0	10.5	12.0	19.0
75th	14.0	13.5	16.5	33.9
Maximum	54.0	72.0	136.0	418.0
Average	12.3	14.3	17.6	36.9

^aProbe quantity / hybridization buffer / slide blocking procedure

ranged from 54.0 for the 50 μ g / commercial buffer / 1% NH₄OH hybridization to 418.0 for the 25 μ g / LLNL buffer / 0.25% tryptone hybridization. The minimum signal intensity for the 25 μ g / LLNL buffer / 0.25% tryptone hybridization was slightly higher (11.0) than the other procedures which resulted in minimum intensities of ~6.0. The highest and lowest average signal intensities were observed for the 25 μ g / LLNL buffer / 0.25% tryptone hybridization (36.9) and the 50 μ g / commercial buffer / 1% NH₄OH hybridization (12.3), respectively. At the 75th percentile, the 25 μ g / LLNL buffer / 0.25% tryptone hybridization had 2.1 to 2.5-fold higher signal intensities compared to the other three hybridization protocols.

Median background-subtracted signal intensities obtained from the different hybridization protocols were also evaluated by determining the number of genes with signal intensities ≥ 1.5 times the average negative control intensity (Table 2.2.5). The negative control intensities obtained for the Alexa Fluor 546 probes ranged from a low of 9.5 to a high of 14.1 (data not shown). The negative control intensities obtained for the Alexa Fluor 488 probes were slightly higher and ranged from 10.1 to 18.5 (data not shown). Among the probes labeled with Alexa Fluor 488, the hybridization of 25 μ g of probe with the commercial buffer to 1% NH₄OH-blocked slides resulted in the largest number of genes with intensities at least 1.5 times the average negative control intensity (25 genes ≥ 1.5 ; 12 genes ≥ 3.0). Increasing the probe quantity for this procedure to 50 μ g, however, resulted in the lowest number of genes with signal intensity ratios that were at least 1.5 (14 genes ≥ 1.5 ; 2 genes ≥ 3.0). Among the Alexa Fluor 546 labeled probes, microarrays blocked with 0.25% tryptone and hybridized using 25 μ g of probe in the LLNL buffer had the highest number of genes with intensities ≥ 1.5

Table 2.2.5. Relationship between the tissue sample hybridization protocol utilized and the number of genes with signal intensity ratios ≥ 1.5 .^a

Signal Intensity Ratio	Number of genes			
	50 μ g / Commercial / NH ₄ OH ^b	25 μ g / Commercial / NH ₄ OH	50 μ g / LLNL / Tryptone	25 μ g / LLNL / Tryptone
Alexa Fluor 488				
≥ 1.5	14	25	16	16
≥ 2.0	9	20	6	9
≥ 2.5	3	14	4	5
≥ 3.0	2	12	3	3
Alexa Fluor 546				
≥ 1.5	13	11	19	23
≥ 2.0	5	7	9	21
≥ 2.5	4	5	5	13
≥ 3.0	2	3	4	11
Alexa Fluor 488 & Alexa Fluor 546				
≥ 1.5	21	27	21	25
≥ 2.0	13	22	10	22
≥ 2.5	6	15	5	13
≥ 3.0	3	13	4	11

^aCalculated for each gene as: (average gene intensity)/(average negative control intensity)

^bProbe quantity / hybridization buffer / slide blocking procedure

times the average negative control intensity (23 genes ≥ 1.5 ; 11 genes ≥ 3.0), while probes hybridized using the commercial buffer and 1% NH_4OH blocking reagent had the lowest number of genes (11-13 genes ≥ 1.5 and 2-3 genes ≥ 3.0). When both Alexa Fluors were considered together, the number of genes with signal intensity ratios ≥ 1.5 but less than 2.0 was similar across all procedures (21-27 genes). However, slides hybridized with 25 μg of probe, irrespective of the hybridization buffer and blocking procedure, had a 2.8 to 4.3-fold increase (over slides hybridized with 50 μg of probe) in the number of genes with signal intensity ratios ≥ 3.0 .

2.2.5 Discussion

The results of the gene-specific and tissue sample hybridizations show that probe quantity, hybridization buffer, and slide blocking procedures all affect background-subtracted signal intensities, and therefore may also affect accurate microarray expression ratio quantitation, especially for genes with low hybridization signal intensities.

Gene-specific probe hybridizations

Gene-specific hybridizations were performed to determine the effects of hybridization buffer and microarray slide blocking procedure on background-subtracted signal intensities. These simple probes were ideal for comparing the intensities obtained for specific genes following different hybridization procedures because the amplicons (which do not degrade as readily as the tissue RNA used as the starting material for tissue

sample hybridizations) can be simultaneously labeled, combined, and then aliquoted from a single master mix for each hybridization.

Irrespective of the set of hybridization conditions utilized, Alexa Fluor 546 intensities were always lower than the Alexa Fluor 488 intensities. Therefore, these differences are most likely the result of intrinsic differences between the fluorophores and/or differences between the filter sets used for image capture. When considering total intensities, however, the LLNL buffer / 0.25% tryptone procedure yielded ~2.2-fold higher intensity than the method that had been traditionally used in our laboratory (commercial buffer/ammonium hydroxide). In fact, microarray slides hybridized using the LLNL hybridization buffer and blocked with 0.25% tryptone generally had the highest signal intensities, followed by LLNL buffer / 0.50% tryptone and LLNL buffer / 1% NH₄OH which behaved similarly. With the exception of the signal detected for *Xrcc1*, the commercial buffer / 1% NH₄OH combination gave the lowest background-subtracted signal intensities. Overall, the LLNL buffer / 0.25% tryptone protocol increased signal intensity ~94% to 204% compared to the signal intensities obtained for the commercial buffer / 1% NH₄OH protocol previously used in our laboratory.

Tissue sample probe hybridizations

Tissue sample hybridizations were performed to determine the combined effects of probe quantity, hybridization buffer, and microarray slide blocking procedure on background-subtracted signal intensities. The hybridization conditions traditionally used in our laboratory (50 µg probe / commercial buffer / 1% NH₄OH) were compared to those which used 1/2 the probe quantity and/or a different buffer and blocking reagent.

The results showed that decreasing probe quantities only increased the background-subtracted signal intensities for particular fluorophores. Unlike the trend toward increased signal intensity which was observed for lower quantities of the Alexa Fluor 488 probe, there were no notable differences in intensities based on the Alexa Fluor 546 probe quantity alone. This may suggest that the lower probe amounts resulted in reduced background for the Alexa Fluor 488 images (which would increase the magnitude of the background-subtracted signal intensity). A similar effect was probably not observed for the Alexa Fluor 546 probes because the background for these images was substantially lower than that for the Alexa Fluor 488 images. Between the two protocols which used 25 μg of probe, the commercial buffer / ammonium hydroxide combination resulted in the highest number of genes with Alexa Fluor 488 signal intensities that were at least 1.5 times greater than negative control intensities, whereas the LLNL buffer / 0.25% tryptone combination resulted in the highest number of genes with Alexa Fluor 546 signal intensities that were at least 1.5 times greater than negative control intensities. However, when the results for both Alexa Fluors were considered together, the total number of genes with signal intensity ratios of at least 1.5 was similar between the hybridization procedures. This suggests that, like probe quantity, the combination of hybridization buffer and slide blocking procedure also affects signal intensity in a fluorophore-dependent manner. With respect to the results for Alexa Fluor 488, it is possible that ammonium hydroxide is superior to tryptone in masking the autofluorescence of the silane-coated microarray. For Alexa Fluor 546, which tends to have much lower background, the LLNL buffer / 0.25% tryptone combination may produce larger signal

intensities due to a decrease in fluorophore quenching during hybridization to the tryptone-blocked slide.

Summary

Hybridization conditions for the cDNA microarray experiments performed in our laboratory previously involved using commercially prepared hybridization buffers to hybridize large probe quantities ($\geq 50 \mu\text{g}$) to slides blocked with ammonium hydroxide. The results for both the gene-specific hybridizations (irrespective of fluorophore) and the Alexa Fluor 546-labeled tissue sample hybridizations indicated that the largest signal intensities were obtained when the LLNL buffer was used in combination with the 0.25% tryptone blocking reagent. However, experience with multiple microarray hybridizations has shown that for the white light capture system used in our laboratory, images for the Alexa Fluor 546 probes are normally captured using longer integration times and higher camera gain settings than those used for the Alexa Fluor 488 probes. This indicates that the Alexa Fluor 546 signal intensities are inherently lower than the Alexa Fluor 488 intensities. Because these research findings suggest that microarray experiments visualized with this white light image capture system may show improved signal intensities by hybridizing lower probe quantities ($\sim 25 \mu\text{g}$) with the in-house hybridization buffer (LLNL; 42% formamide / 2 x SSC / 10% dextran sulfate), to 0.25% tryptone-blocked slides, this protocol was utilized in our subsequent cDNA microarray experiments (Chapter 3).

2.2.6 References

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., and Smith, L.M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* 22, 5456-5465.

Kegelmeyer, L.M., Tomascik-Cheeseman, L., Burnett, M.S., van Hummelen, P., and Wyrobek, A.J. (2001). A groundtruth approach to accurate quantitation of fluorescence microarrays. *SPIE Proceed.* 4266, 35-45.

Miesfeld, R.L. (1999). *Biochemical basis of applied molecular genetics. Applied Molecular Genetics.* New York: Wiley-Liss. pp. 6-10.

Strachan, T. and Read, A.P. (1999). *Principles of nucleic acid hybridization. Human Molecular Genetics.* 2nd Ed. New York: Wiley-Liss. pp. 100-5.

Wetmur, J.G. (1971). Excluded volume effects on the rate of renaturation of DNA. *Biopolymers* 10, 601-13.

2.3 Accurate quantitation of fluorescence microarrays

2.3.1 Abstract

To more accurately measure fluorescent signals from microarrays, our image acquisition and analysis systems were calibrated using groundtruth samples comprised of known quantities of gene-specific DNA probes that were labeled with Cy3 and/or FITC and hybridized to cDNA targets. Slides were imaged with a full-field, white light CCD imager and analyzed using custom analysis software. The results obtained with and without preprocessing (alignment, color crosstalk compensation, dark field subtraction, and integration time) were compared for multiple genes. The accuracy of various image processing and analysis techniques (background subtraction, segmentation, quantitation and normalization) was also evaluated. This methodology was used to calibrate and validate our system for accurate, quantitative measurement of microarrays. The results show that preprocessing the images resulted in measurements that were substantially closer to the known groundtruth for these samples.

2.3.2 Introduction

Expression microarrays provide a means for monitoring the expression of many genes in parallel. Therefore, this technology can provide in-depth understanding of biological processes such as DNA repair, cellular differentiation, and development. Careful target selection is essential for utilizing custom-built cDNA microarrays to address pathway-specific topics. These targets must not only represent the potentially

interesting and relevant genes but also minimize sequence homology (and thus cross-hybridization) among spots. Since subtle differences in gene expression measurements can have a large impact on the interpretation of biological data, specificity and measurement accuracy are especially important.

There are many ways to perform the image processing and analysis steps needed to derive quantitative information from a microarray image (Brown et al. 2000; Chen et al. 1997; Pie'tu et al. 1996). We evaluated a number of different methods for processing and analyzing images and showed the effects of preprocessing on quantitation. Segmentation, quantitation, background subtraction, preprocessing, and normalization were defined as follows:

- Segmentation delineated the extent of each spot, and thus distinguished spots from surrounding background.
- Quantitation involved measuring intensities within the spot boundaries determined by segmentation.
- Background subtraction was used to remove the effects of autofluorescence and other effects that are not due to specific fluorescent hybridization.
- Preprocessing involved characterizing and accounting for the acquisition system parameters such as camera dark field, spectral crosstalk, image alignment, integration time and camera gain.

- Normalization made the measurements from the 2 color channels (e.g. red and green) commensurable so they could be sensibly compared or arithmetically manipulated relative to one another. It also enabled slide-to-slide comparisons.

Each of these processes has optional techniques and implementations. Not only are there a number of different possible algorithms for each step, but the optimal combinations may also differ from system to system. We used groundtruth to determine which techniques and combination of the above options were optimal for correlating computed intensities with known probe amounts for our system. However, this methodology can be used to calibrate any acquisition and analysis system in order to improve measurement accuracy.

2.3.3 Materials and Methods

Generation of amino modified clones

Expressed sequence tag (EST) clones were obtained from LLNL's I.M.A.G.E. Consortium and used as the target cDNA for the microarrays. PCR with sequence-specific primers verified that each clone represented the correct gene. Two ESTs with an insert size ranging from 500 to 1500 base pairs were selected to represent each gene. Prior to spotting, inserts from the selected clones were PCR amplified from plasmid preparations using 5' C6 amino-modified, vector-specific primers and purified using

Qiagen PCR purification columns (Qiagen Inc., Chatsworth, CA). The amplicons were precipitated and resuspended for spotting in 0.1M Sodium Carbonate/Bicarbonate (pH=10.2) to a final concentration of 2 μ g/ μ l.

Slide preparation

Slides were derivatized according to Guo et al. (1994). Briefly, glass slides were coated with 1% 3-aminopropyltrimethoxysilane [Sigma, St. Louis, MO] in 95% acetone/water for 2 min., washed 10 times in acetone, and baked at 110°C for 45 min. Prior to spotting, the silane was activated by incubating the slides for 2 hours in 0.2% 1,4-phenylenediisithiocyanate [Sigma, St. Louis, MO], 10% pyridine [Sigma, St. Louis, MO] and dimethylformamide [Aldrich, Milwaukee, WI]. Slides were then washed in methanol (2x 10 min.) and acetone (2x 10 min.). Slides were air dried and used immediately for spotting.

Robotic arrayer

A custom robotic high speed arrayer was used to grid cDNA onto glass microscope slides. The gridding system had a 3-axis DC servo driven gantry (GM2340R, Glentek, El Segundo, CA). The full travel of the Z-axis was 0.25m with 5 μ m resolution and of the X-and Y-axis (powered by Newport-Klinger MD4 servo motor driver), 2m and 1m, respectively, with 20 μ m resolution each. The system controller was a Newport-KlingerMM2000 card with 3DC modules in an Intel 80486 PC. The spotting tool had 2 Beryllium-Copper plated pins with a spacing of 4.5mm. The grid density was 4.5 x 4.5

mm for each of 2 pins with a center to center spacing of 375 μm . After spotting, the slides were humidified in a 37°C incubator for 5 minutes, air dried and stored under vacuum at room temperature until hybridization.

Probe labeling

To create groundtruth over a series of 10 slides, 10 probe mixtures were made using clones for the following genes: *Globin*, *Dna-pkcs*, *Tp53*, *Rad50*, *Rad52*, and *Ku80*. Each mixture contained both Cy3 and FITC labeled probes for each of the above genes. The ratio of the FITC and Cy3 probe amounts shown were expected to correlate with the subsequent ratio of intensity measurements. *Globin* and *Dna-pkcs* probes had equal proportions across all slides to serve as controls (each was expected to yield a Cy3/FITC ratio of 1). *Globin* was used for normalization. For each of the remaining four genes, the probe amount for one dye was held constant while the probe amount for the other dye varied, forming a dilution series as shown in Table 2.3.1. Each mixture was hybridized to one slide.

Probes were generated by PCR amplification of the same I.M.A.G.E. clones that were spotted onto the array using gene-specific primers. A spectrophotometer was used to measure the quantity of each probe. The probe labeling was performed by mixing serial dilutions for each gene together per color and afterwards replacing dCTP by Cy3-dCTP (Amersham Pharmacia Biotech, Piscataway, NJ) or dUTP by FITC-dUTP (dUTP (Roche, Indianapolis, IN) using a Nick Translation kit (Life Technologies, Rockville, MD) which incorporated labeled nucleotides by using the enzyme DNase I to “nick” the DNA and DNA polymerase to replace the excised nucleotides with a mixture of

Table 2.3.1. Six genes were used to build a groundtruth series over ten microarray slides.^a

Globin	FITC	2.3 ng	→	2.3 ng	Rad52	FITC	18.2 ng	→	0.01 ng
	Cy3	2.3 ng		2.3 ng		Cy3	3.0 ng		3.0 ng
Dna-pkcs	FITC	2.4 ng	→	2.4 ng	Ku80	FITC	3.2 ng	→	3.2 ng
	Cy3	2.4 ng		2.4 ng		Cy3	18.6 ng		0.01 ng
Tp53	FITC	20.4 ng	→	0.01 ng	Rad50	FITC	3.0 ng	→	3.0 ng
	Cy3	3.3 ng		3.3 ng		Cy3	17.3 ng		0.01 ng

^aArrows represent the serial change in the probe amount across the 10 slides.

unconjugated and fluorescently-conjugated nucleotides. The Cy3 and FITC labeled probe mixtures were added together and purified with Qiaquick spin columns (Qiagen Inc., Chatsworth, CA).

Hybridization

To block non-specific probe binding, slides were incubated for 10 min. in 1% NH₄OH (Sigma, St. Louis, MO) and washed 3 x 10 minutes in double distilled water prior to hybridization. Slides were then denatured for 6 minutes in 70% Formamide/2xSSC at 78°C, dehydrated through a 70- 85-100% ethanol series, and air dried at room temperature. The probe mixtures hybridized to each slide had similar total concentrations. Labeled probes were concentrated using speed vacuum centrifugation and resuspended in 5-10 µl hybridization mix containing 70% formamide, 2x SSC, 10% dextran sulfate and 1 µl herring DNA. The hybridization mixtures were placed on the microarray under a 22x22 mm coverslip, sealed with rubber cement, and denatured at 72°C for 3 minutes. Hybridization was performed for 24 hours at 37°C in a moist chamber, followed by 2 x 10 minute washes at 37°C in 2xSSC. Vectashield mounting media (Vector Laboratories Inc., Burlingame, CA) was used to prevent photobleaching of the fluorescent dyes.

Image acquisition

Fluorescent microarray images were acquired with a full-field (15 mm square; resolution = 0.015mm/pixel), white light imaging system (Norgren Associates, Palo Alto, CA). Light from a 500 watt Xenon light source (400 nm - 600 nm) was scrambled

through a fiber optic and passed through a bandpass excitation/emission filter pair before reaching a scientific grade CCD camera. The camera was controlled via a PC with a MATROX PULSAR digital acquisition and display board. Images were collected onto the PC and then transferred to a Unix workstation for analysis. This system was modified by adding a second light source (a mercury arc lamp) for UV excitation. The UV wavelength allowed the excitation of the nucleic acid stain DAPI (4',6-Diamidino-2-phenylindole) which assisted in segmentation by delineating all spots, including those with weak or absent hybridization signals.

Image analysis

Custom software was written within the SCIL-Image development platform (TNO/TPD, Delft, The Netherlands) for preprocessing, automated background subtraction, grid detection, spot detection and quantitation. Custom Perl scripts were used to ratio, normalize, summarize and plot data resulting from the image analysis.

2.3.4 Image Processing and Analysis Methods

Preprocessing

To calibrate the acquisition system, inherent system response characteristics were measured and corrected or accounted for in the analysis. For example, all CCD cameras have thermal noise even when no light is incident on the CCD. Dark noise, which is noise from all sources except photons, was measured and then subtracted from the images (Mullikin et al. 1994). Additionally, because the filters in the system do not have perfect

bandpass cutoffs and light through the filters is not always parallel, spectral crosstalk (FITC signal leaking into the Cy3 channel and vice versa) occurs. All systems should also measure and correct for this (Castleman et al. 1996). Image misalignment was also corrected, via correlation methods, during preprocessing. Misalignment occurs whenever there is a physical shift of the microarray slide between capturing images with different filters. Lastly, we corrected for integration time differences between the FITC and Cy3 images. Although it is feasible to leave this correction until normalization, it is prudent to correct for known disparities in a straightforward manner. Then, the factors remaining to be corrected by normalization are those for which reliable characterization is not available.

Problems associated with two of the system response characteristics, camera gain and non-uniform illumination, were ameliorated even though they are difficult to correct. The camera gain is non-linear at higher settings, so gain was held constant during acquisition for all images in the same experiment. Non-uniform illumination was difficult to correct because of wavelength dependencies (chromatic aberration) and variation of the light source over time; therefore, the illumination pattern was optimized for uniformity as much as possible during acquisition and was not corrected computationally.

Background subtraction

Quantitative information was extracted from hybridization signals following preprocessing. First, background subtraction was performed in order to account for the various sources of fluorescent background, such as autofluorescence inherent to the glass slide or other substrate, autofluorescence of the antifade, and the non-specific binding of

fluorescent probe to material in the area surrounding the target spots of the array. This procedure is different from camera dark field subtraction which accounts for parameters inherent to the camera.

Informal investigations have shown that different substrates have varying levels of autofluorescence. For some systems, high quality, polished quartz, such as Corning brand slides had a lower contribution to background fluorescence than standard glass microscope slides. If background fluorescence was attributable to autofluorescence of the substrate, then there was a slowly varying background intensity under the array which served as a "DC offset" to the spot signals. (All spot intensities should be above this background.) In such a case, background subtraction was a logical and necessary step for accurate quantitation of spot signals.

If background fluorescence was attributable to non-specific binding of the labeled probe, then the surrounding signal was only around the spots, because the target DNA spotted on the slide prevents the non-specific binding at the spots. In this case, the background could have higher intensity than the spots (Figure 2.3.1). When background intensity is higher than target intensity, background subtraction is unwarranted because it would result in negatively valued spot intensities. Therefore, images with this type of background were not quantitated and another array was hybridized after applying an appropriate blocking procedure to prevent non-specific probe binding. Alternatively, spots could be compared to the negative controls (spots that had no probe hybridization) to get a relative sense of their brightness, but accurate ratios may be hard to obtain, since there may still be autofluorescence from the substrate which can't be characterized or subtracted.

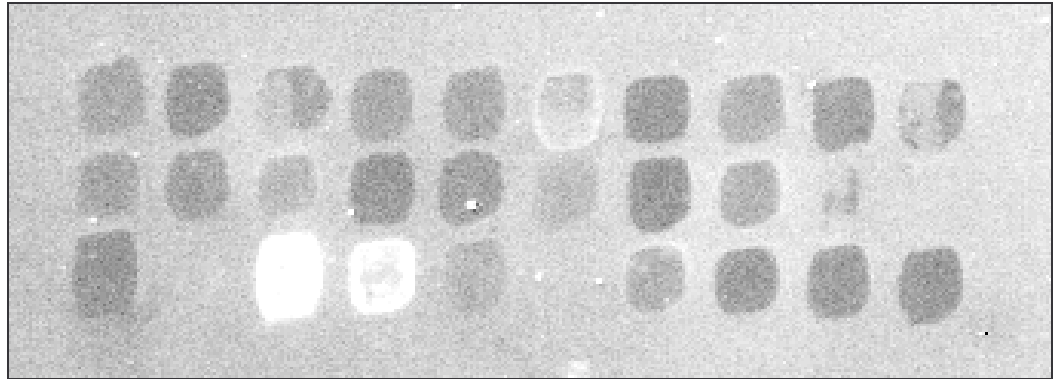


Figure 2.3.1. A contrast-enhanced example of a fluorescence microarray image with very high background. The bright white spots indicate hybridized cDNA targets. Many target cDNA spots were not hybridized and appeared darker than the surrounding background (as indicated by the black spots within the white background above). Therefore, the high background was not additive to the spot intensities, and the background intensities should not be subtracted from the target spots.

After determining which manner of background subtraction was most appropriate for a given image, we selected from a number of methods to perform the task (sampling pixel intensities near the extremes of the image, sampling just outside of spots, etc.). The background subtraction method we employed was an adaptive, non-linear technique which continuously sampled the background and created a 2-dimensional, slowly changing "sheet" or blanket that followed the low frequency trend of intensity outside of the spots. This method, called the lower envelope subtraction (Verbeek, et al. 1988), lowered the signal intensities relative to a base value near zero (Figure 2.3.2).

Semi-automated grid placement

To detect spots with shape and location anomalies, a semi-automated grid placement algorithm was utilized which requires an operator to indicate the extent of the array and the number of spots present. With this information, the system divided the area into self-adjusted, not necessarily uniformly sized, grid squares and each square contained one spot.

Three-color segmentation

Once a grid was overlaid on the array, segmentation methods were used to determine which pixels inside a grid square belonged to the spot under investigation and which belonged to the background. As expected, segmentation was more difficult for dim or non-expressing spots than for bright spots. To address this problem, a third image of the array was acquired after ratio data was obtained for the first two images (Cy3 and FITC). This third image was captured after applying a DAPI DNA counterstain to the

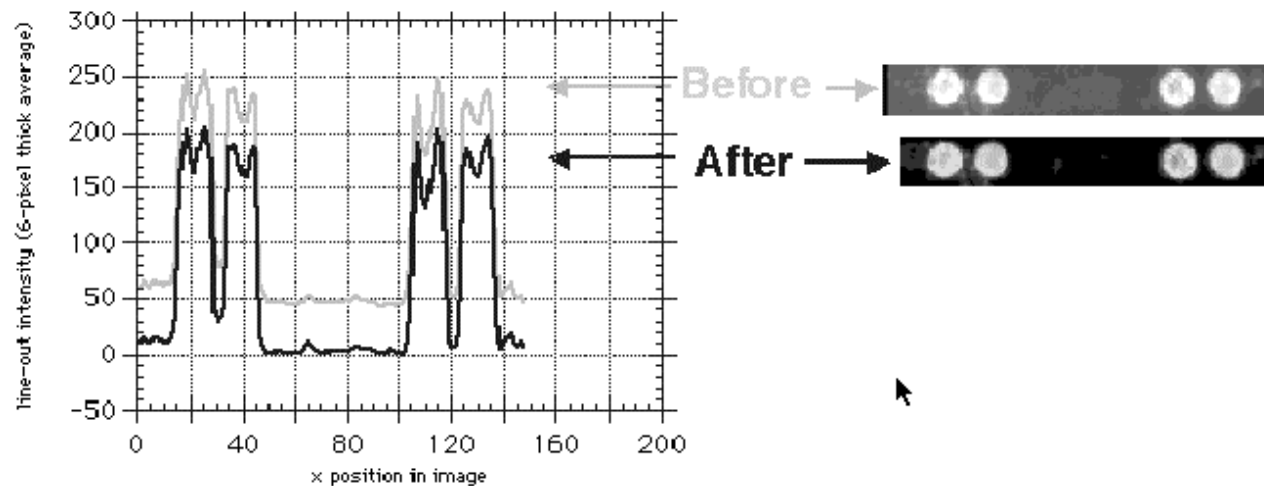


Figure 2.3.2. Removal of underlying autofluorescence and subsequent lowering of target spot intensities through background subtraction. The gray-scale microarray images of target spots (right) represent a row of hybridized spots before (top image) and after (bottom image) background subtraction. The graph (left) plots position within the image (x-axis) vs intensity for the corresponding position in the image (y-axis). Intensity measurements were made using a six pixel wide horizontal line through the image.

slide to mark the location of all DNA target spots. The presence of DAPI stain removed any ambiguity about the presence of a valid target DNA when no expression signal was observed. It also allowed better determination of the size and shape of the DNA target for segmentation. However, DAPI was not always present in spots with a strong hybridization signal. (It is possible that the hybridization of the labeled probes changed the properties of the DNA molecule and prevented intercalation of the counterstain.) Therefore, images from all 3 dyes (Cy3, FITC, and DAPI) were superimposed and aligned in order to obtain the best signal from each spot. This resulting composite image was subsequently used for segmentation.

Methods for spot segmentation

When microarray target spots appear to be very dim, "bagel-shaped", or have other irregularities in the absence of counterstain such as DAPI, different segmentation methods define different spot areas which in turn affects the intensity measurement. As shown in Figure 2.3.3, the following segmentation methods were studied:

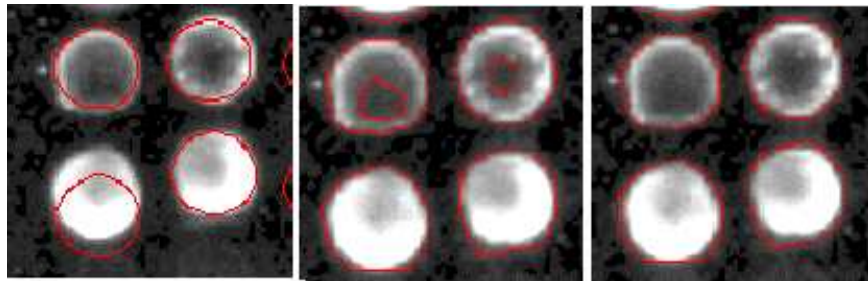


Figure 2.3.3. Contours of hybridized target spots delineated by different segmentation methods. From left to right these segmentation methods are: Circle Hough transform, Trian threshold, and Hulled Trian. Each of these segmentation methods were evaluated to determine which method resulted in ratios of computed intensities closest to the groundtruth (i.e., yielded the smallest average error for ratios of known probe quantities).

- Circle Hough Transform: Best fit circle was found using an edge detector to determine the edges of the spot and then determine their magnitude and direction. This information was used to determine the center of the circle that best fits the edge information (Verbeek et al. 1998)
- Trian Threshold: An intensity-based histogram was drawn in a grid square containing one spot. The low-intensity peak of the histogram (which represents background values) was found and a line was fit to the falling slope on the right side, forming a triangle (hence the name trian) with the y-axis. The clipping level (intensity threshold) was set where this line crosses the x-axis of the histogram (Ballard et al. 1981).
- Circles + Trian: A logical OR statement was used to combine the results from the two methods above.
- Hulled Trian: The morphological convex hull operation was applied to the result of the Trian Threshold.

Methods for quantitation of signal intensity

A number of methods could be used to compute the relative intensities of the red and green probe signal for a spot. The differences between the methods lie primarily in how well they accurately quantify the signal in the presence of noise or other artifacts. We implemented and evaluated a number of methods:

- Total Intensity. Summed the intensities of all of the pixels assigned to a spot, for each color, and then divided them to get the ratio:

$$\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}$$

where Y was the FITC signal, X was the Cy3 signal, i was a pixel in the spot and n was the total number of pixels in the segmented region.

- Median. Found the median pixel intensity in each color and took the ratio:

$$(\text{Median of FITC pixel values})/(\text{Median of Cy3 pixel values})$$

This method was superior to the total (or averaging) method when there were noisy outliers (e.g. a few really bright or dim pixels). It also worked especially well when there was a similar number of low and high outliers.

- Pixel-to-pixel mean and average ratios. Divided the intensities of a spot on a pixel:pixel basis. Then, calculated the mean or the median of the pixel:pixel ratios for the entire spot. This method assumed that the two wavelength images were perfectly aligned and that the spot shape was consistent in both colors.
- Fit line. Since the pixel:pixel ratios within a spot can vary a lot, especially for bagel-shaped spots, another approach was to plot the pixel intensities in one wavelength versus another for each pixel and then fit a line to the resulting scatterplot. Each pixel was represented by a point on the scatterplot, with the x location its intensity in one wavelength and the y location its intensity in

another. The slope of the line that passed through the 0 y-intercept represented the final ratio.

- Log geometric mean. The log geometric mean of the intensities in a spot is 10^L , such that

$$L = \sum_{i=1}^n \log Y_i / X_i$$

where Y represented the FITC signal, X represented the Cy3 signal, i represented one pixel in a spot and n was the number of pixels in the spot.

Normalization

Once an intensity ratio was determined for a particular spot, it had to be normalized to a standard in order to account for variations due to differences in exposure time, amount of target, amount of probe, dye incorporation, rate of photobleaching, hybridization conditions, imaging conditions, etc. Normalization was achieved by dividing all spot intensities on a slide by the average ratio of positive controls, which were spots designed to have the same intensity and ratio on all slides. (*Globin* was hybridized using equal amounts of each color and served as the positive control for this groundtruth series.) When positive controls didn't have the expected ratio, we assumed the differences were due to the above variations, and we used the factor by which they differed as a correction factor for all other spots.

2.3.5 Results

Preprocessing Results

In order to measure the effects of various analysis steps, we compared the computed intensity ratios against the known ratio of known probe quantities. As shown in Figure 2.3.4, the two were plotted such that the distance between the curves represented the error between measured and known quantities, with and without preprocessing. The curves were much closer together over a greater range of probe quantities with preprocessing. It was also possible to gauge the probe amounts at which the two curves diverge and to note that the computed intensity ratio was more accurate down to lower probe amounts following preprocessing.

Error evaluation for various analysis methods and combinations

In order to quantify the disparity among the curves for one gene, we calculated an error by summing the distance from measurement to the truth for all 10 dilutions in the series:

$$\text{Error} = \sqrt{\sum_{n=1}^{10} [\log(\text{observed intensity ratio}) - \log(\text{quantity ratio})]^2}$$

where intensity ratio was computed using results from one of the quantitation methods described earlier and the quantity ratio was computed using the known probe amounts. For the graphs in Figure 4, the error for quantitation type was 1.28 without preprocessing and 0.73 with preprocessing. These values are also shown in Table 2.3.2.

Based on Table 2.3.2 error chart, preprocessing was most valuable when signal strength was low and when there was an intensity disparity between the two probe amounts. Probes that were plentiful and equal in the two colors, such as *Globin* and

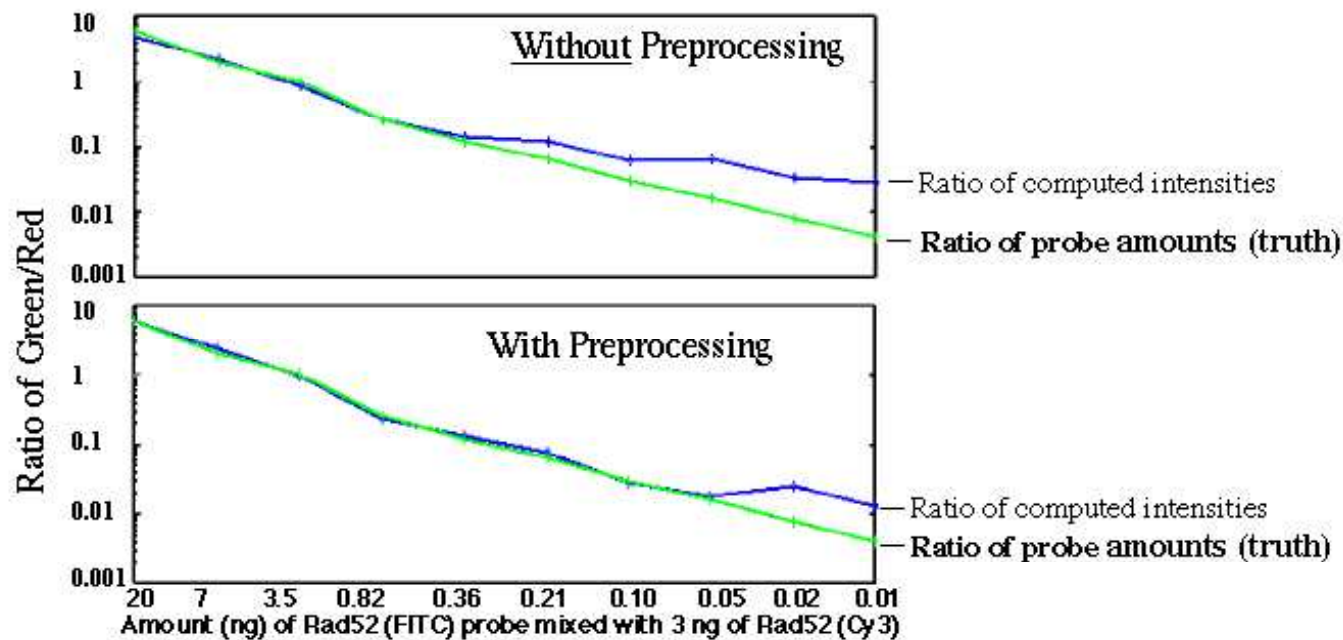


Figure 2.3.4. The groundtruth (ratio of known probe amounts) for *Rad52* was tracked among the 10 slides as shown by the (lower) green line. The ratio of measured intensities is represented as the (upper) blue line. The cumulative distance between these two lines provided an error measurement that enabled different methods to be applied and compared in a quantitative manner.

Table 2.3.2. Error measurements obtained for the groundtruth series following different image analysis methods.^a

	Full preprocessing			No preprocessing, except alignment		
	Total intensity	Median	Log geometric mean	Total intensity	Median	Log geometric mean
<i>Globin</i>	0.05	0.07	0.07	4.34	0.07	0.07
<i>Dna-pkcs</i>	0.33	0.36	0.35	0.27	0.29	0.29
<i>Tp53</i>	1.00	1.06	0.98	2.88	1.60	1.57
<i>Rad50</i>	1.25	1.29	1.33	1.89	1.74	2.07
<i>Rad52</i>	0.65	0.73	0.73	1.25	1.28	1.35
<i>Ku80</i>	1.32	1.39	1.42	1.91	1.81	2.10
<i>Tp53, Rad50, Rad52, and Ku80</i>						
Average error	1.06	1.12	1.12	1.98	1.61	1.77
Standard deviation	0.30	0.29	0.32	0.67	0.24	0.37

^aError was determined by measuring the cumulative distance between the truth (quantity ratio) and measured (observed intensity ratio) curves (e.g., Figure 2.3.4).

Dna-pkcs, did not benefit substantially from preprocessing. The other four genes, however, showed significantly larger improvements with preprocessing. This was primarily due to the color correction which accounted for spectral crosstalk by removing intensities from one color channel and adding them back to the appropriate color channel. For dim spot signals, this incremental change could be substantial, whereas bright spot signals were minimally affected.

Also shown in Table 2.3.2 are the effects of utilizing different quantitation methods. As seen in the average error along the last row of the chart, the total intensity quantitation method could result in the lowest average error (1.06 with full preprocessing), but it could also give the highest average error (1.98 with no preprocessing, except alignment) when data was corrupted by debris or other noise. Therefore it was not a robust quantitation method. Instead, the median and the log geometric mean error values were more consistent with lower standard deviations and overall lower error than the total intensity method and all other methods evaluated (data not shown).

Since all of the analyses used to generate the data for the chart were normalized via the *Globin* positive control spots, it stands to reason that the largest error occurred when the *Globin* control spot contained undetected debris, and the quantitative method of "total" included those values in the intensity measure for *Globin* (Table 2.3.2). Since that measurement was flawed, and since that value was subsequently used to normalize all of the gene intensity ratios, all computed ratios using that method were flawed.

Just as these errors were generated for various combinations of preprocessing and quantitation parameters, similar values were generated for the other analysis steps. For

evaluation of background subtraction, we held constant the settings for preprocessing, segmentation, quantitation and normalization and only changed whether or not background was subtracted. For the genes of the dilution series, the average error improved from 1.41 to 1.12 when background was subtracted and the median or log geometric mean was used for quantitation. Further, almost all of the improvement was realized for the genes which had low probe amounts. As expected, the ratios for spots with strong intensity were least affected by background subtraction.

2.3.6 Conclusions

While there are numerous approaches to microarray analysis, accurate quantitation required system calibration. The methodology outlined in this paper can be used to calibrate any acquisition and analysis system and can be tailored and optimized for specific data and specific experiments for improved measurement accuracy. In order to calibrate and validate our system, we generated a dilution series with known quantities for the red and green hybridization probes. We used that groundtruth sample to evaluate the results of a number of computational techniques for preprocessing and analysis. Computed ratios were compared to the known groundtruth, and an error was calculated for each method. In order to obtain the least overall error for our groundtruth samples and our system, we found that:

- preprocessing was superior to not preprocessing
- subtracting background was superior to not doing so

- spots were best segmented by combining a best circle fit with an intensity threshold
- quantitation was most accurate when using the median or the log geometric mean
- normalizing by a positive control was superior to not doing so.

Preprocessing had the greatest impact on bringing the computed ratios closest to the groundtruth. For the gene with the lowest cumulative error across all analysis methods, *Rad52*, preprocessing enabled accurate quantitation of DNA probe to 0.05 ng (improved from a limit of 0.36 ng without preprocessing). Future efforts will establish whether this translates to the detectable amount for samples using cDNA or RNA probes and will also evaluate metrics for signal brightness and uniformity.

2.3.7 Auspices

This work was performed under the auspices of the U.S. DOE by the University of California, LLNL under contract W-7405-Eng-48 with funding from NIH ES09117-02, DOE KP110202, and additional support from West Virginia University and the University of California Campus Laboratory Collaboration.

2.3.8 References

Ballard, D.H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recog.* 13, 111-122.

Brown, M.P.S., Grundy W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *PNAS* 97, 262-267.

Castleman, K.R., Riopka, T.P., and Wu, Q. (1996). FISH image analysis. *IEEE Engineer. Med. Biol.* 15, 67-75.

Chen, Y., Dougherty, E.R., and Bittner, M.L. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics* 2, 364-374.

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., and Smith, L.M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* 22, 5456-5465.

Mascio, L.N., Verbeek, P.W., Sudar, D., Kuo, W-L., and Gray, J.W. (1995). Semiautomated DNA probe mapping using digital imaging microscopy: I. System development. *Cytometry* 19, 51-59.

Mullikin, J.C., van Vliet, L.J., Netten, H., Boddeke, F.R., van der Feltz, G., and Young,

I.T. (1994). Methods for CCD camera characterization. *Image Acquisition and Scientific Imaging Systems*, Helen C. Titus and Amir Waks Eds. 2173, 73-84.

Pie'tu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariage-Sampson, R., Houlgatte, R., Soularue, P., and Auffray, C. (1996). Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Gen. Res.* 6, 492-503.

Verbeek, P.W., Vrooman, H.A., and van Vliet, L.J. (1988). Low Level image processing by max-min filters. *Signal Process.* 17, 249-258.

Zack, G.W., Rogers, W.E., and Latt, S.A. (1977). Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* 25, 741-753.

Differential Basal Expression of Genes Associated with Stress Response, Damage Control, and DNA Repair Among Mouse Tissues

3.1 *Abstract*

Tissues must be capable of efficiently recognizing and repairing various types of DNA damage in order to maintain genomic integrity and the overall health of an organism. However, differential DNA damage susceptibilities and cancer incidences have been observed among tissues, and the molecular mechanisms underlying tissue-specific differences are not well understood. The purpose of this research was to compare and contrast transcription profiles among healthy adult mouse tissues (testis, brain, liver, spleen and heart) using a 417 gene cDNA microarray enriched for genes involved in DNA damage recognition and repair processes. Several tissue-specific patterns of expression were identified through cluster analysis. With respect to specific biological pathways, we found that ~41% of the stress response genes, ~23% of the damage control genes and ~10% of DNA repair-associated genes were significantly differentially expressed among the tissues examined. In general, stress response genes exhibited the highest expression in liver and heart while DNA repair genes exhibited the highest

expression in testis. Damage control genes associated with cell cycle regulation often had the highest expression in testis. The finding that tissues differ in their basal expression of stress response, damage control and DNA repair-associated genes raises important questions regarding tissue-specific responses to endogenous and exogenous genotoxic agents and differential genetic susceptibility to various diseases, including cancer.

3.2 Introduction

Health maintenance requires that all mammalian tissues be capable of recognizing and repairing a variety of insults to genomic DNA. Yet, tissue-specific differences have been observed in the response to endogenous and exogenous genotoxic agents and cancer incidence.

Several studies have reported differential tissue responses (ranging from modulations in gene expression to differences in the amount of damage induced) to a variety of DNA damaging agents. Recently, it has been suggested that transcriptional responses to ionizing radiation vary among cell lines derived from different tissue-types (Amundson et al., 1999). Valverde et al. showed that the extent of carcinogen-induced genotoxic damage also differs among tissues exposed to cadmium chloride (2000). Evaluation of single strand breakage and alkali-labile sites in various mouse tissues (e.g., lung, liver, kidney, brain, testis) suggested that a single exposure to cadmium induced relatively high amounts of DNA damage in brain and bone marrow, while the liver, testis, and kidney exhibited lower damage.

In addition to the observed differences in DNA damage produced following various environmental exposures, cancer incidence rates are also known to vary among tissues (American Cancer Society; http://www.cancer.org/eprise/main/docroot/stt/stt_0). The distribution of the approximately 1.3 million new cancer cases expected in the United States during 2002 varies significantly, ranging from a high of 1 in 6 new cases attributable to breast cancer to a low of 1 in 1070 new cases attributable to male genital cancers, excluding testicular and prostate cancers. Cancers of the brain/nervous system and liver/intrahepatic bile duct are expected to account for 1 in ~75 and 1 in ~77 new cases, respectively. However, soft tissue cancers (including heart) and testicular cancer only account for 1 in ~150 and 1 in ~170 cases, respectively. The molecular processes contributing to these tissue-specific differences in cancer rates are not well understood.

Tissues may respond to DNA damaging conditions by activating signaling cascades involved in stress response (i.e., heat shock or oxidative stress depending on the type of exposure), regulating cell cycle progression, repairing DNA damage, and/or inducing apoptosis. Although many of the major genes in each of these biological pathways have been identified, information regarding their basal mRNA expression levels is limited. A better understanding of the *in vivo* baseline expression levels for these genes would provide insight on the cellular resources that are immediately available for responding to and processing DNA damage.

Several studies have examined tissue-specific differences in the baseline expression of single genes or small numbers of genes associated with damage response and repair (Burns et al., 2001; Leasure et al., 2001; Pittman et al., 1998). In contrast, genome-scale technologies such as cDNA microarrays can provide a comprehensive

parallel evaluation of expression differences across large numbers of genes, biological pathways, and tissues. Although cDNA microarrays have been utilized to profile gene expression in tumor tissues (Alizadeh et al., 2001; DeRisi et al., 1996) as well as in tissues that have been exposed to DNA damaging agents, such as ionizing radiation (Amundson et al., 1999; Fornace et al., 1999), a microarray-based assessment of the differential basal expression of damage response genes among healthy mouse tissues has not yet been reported.

The purpose of this research was to characterize and contrast the relative basal levels of gene expression among five tissues from healthy adult male mice (testes, brain, liver, spleen, and heart) using a cDNA microarray enriched for genes associated with DNA damage recognition and repair. After assessing the reproducibility and precision of our cDNA microarray technology, we (a) compared basal gene expression profiles among tissues, (b) characterized expression differences by biological pathway, and (c) validated cDNA microarray accuracy against northern blot results for selected genes.

3.3 *Materials and Methods*

Custom cDNA microarray generation

Custom cDNA microarrays were generated to represent 417 genes involved in nucleotide excision repair, base excision repair, mismatch repair, homologous and non-homologous recombination, stress response, apoptosis, cell cycle, transcription, translation, growth regulation, meiosis, spermatogenesis, and chromatin-remodeling. For each gene, one to four cDNA clones were obtained from either the I.M.A.G.E.

Consortium at LLNL or Research Genetics (Huntsville, AL). Several bacterial and lambda phage sequences served as controls. Clones were PCR-amplified with 5'-C6 amino-modified vector-specific primers, purified on Qiagen purification columns (Qiagen Inc., Valencia, CA), ethanol precipitated, and resuspended in 0.1M sodium carbonate/bicarbonate (pH = 10.2) for spotting onto the array slides.

Glass slides were cleaned for 30 min in 1:1 concentrated hydrochloric acid:methanol, soaked overnight in concentrated sulfuric acid, and washed 10 x 10 min in room temperature water and 10 min in boiling water. Slides were then submerged in 1% 3-aminopropyltrimethoxysilane (Sigma, St. Louis, MO) / 95% acetone/ water for 2 min, washed 10 x 5 min in acetone, and heated at 110°C for 45 min (Guo et al., 1994). Silane-coated slides were submerged in 0.2% 1,4 phenylene diisothiocyanate (Sigma), 10% pyridine (Sigma) and dimethylformamide (Aldrich, Milwaukee, WI) for 2 hours, washed in methanol (2 x 10 min) and acetone (2 x 10 min), and air-dried. 608 clones were robotically arrayed (Norgren Systems, Palo Alto, CA) in duplicate with a 220 µm center-to-center distance.

RNA isolation and labeling

Eight male B6C3F1 mice (Harlan Sprague Dawley, Indianapolis, IN) were euthanized by CO₂(g) at 2 months of age. Mice were allowed food and water *ad libitum* and were not exposed to genotoxic agents. Testes from six mice were extracted and pooled for use as the reference tissue in each hybridization. Heart, spleen, liver, testes and whole brain were extracted from each of the two remaining mice. Tissues were stored at -80°C until RNA isolation.

Total RNA was isolated by homogenization (Omni Tissue Homogenizer, Warrenton, VA) in TRIzol Reagent (Life Technologies, Rockville, MD). RNA was ethanol precipitated twice, resuspended in RNase-free water (Sigma), and quantified using the GeneQuant spectrophotometer (Amersham Pharmacia Biotech, Piscataway, NJ). 25 µg of RNA was primed with 4 µg of 20mer oligo-dT (Life Technologies) and reverse transcribed at 42°C using Superscript II reverse transcriptase (Life Technologies) in the presence of unlabeled dUTPs (Amersham Pharmacia Biotech) and either Alexa Fluor 488-dUTPs (pooled testes reference; green fluorescence) or Alexa Fluor 546-dUTPs (test tissue; red fluorescence) from Molecular Probes, Inc. (Eugene, OR). For each hybridization, pooled testes and test tissue probes were co-purified on Qiagen columns (Qiagen, Inc.) and ethanol precipitated.

Microarray hybridization

Prepared slides were placed in 0.25% tryptone (Difco Laboratories, Detroit, MI) in water, agitated at 100 rpm for 1 hour, denatured 2 min. in 94-98°C water, dehydrated using an ice cold 70-85-100% ethanol series, and air dried. Probes were resuspended in 2.75 µl water and added to 0.25 µL 10%SDS, 10 µg mouse cot-1 DNA (Life Technologies), 10 µg poly A+ (Amersham Pharmacia Biotech) and 10 µL of 60% formamide / 3xSSC / 10% dextran sulfate. Hybridization mixtures were denatured at 78°C for 6 min. and snap-cooled on ice. An *in situ* frame (Eppendorf, Westbury, NY) was used to confine the probe mixture over the arrayed area. Slides were hybridized 12 hours at 37°C and washed 2 x 2 min. in 42°C water.

Image capture and processing

Image capture and processing was performed as described by Kegelmeyer et al. (2001). Briefly, images were acquired with a full-field white light imaging system. Arrays were exposed to bandpass-filtered excitation light from a Xenon source. The resulting emitted light was bandpass-filtered and collected by a scientific-grade CCD camera (resolution = 0.015 mm/pixel). Custom algorithms, built within SCIL-Image (Delft, The Netherlands), corrected for CCD dark noise, spectral cross-talk, misaligned images and integration time variation. Additional processing algorithms determined the “segmentation mask” for each cDNA spot. The raw red and green intensities for all spots were computed by taking the geometric mean of all pixels within the segmentation mask that were greater than zero after background subtraction. Spots covered by debris were eliminated from all subsequent processing.

Normalization and expression ratio calculations

Logarithms were used to transform the raw red to green intensity ratio for each spot into red minus green intensity differences. This logarithmic transformation helped stabilize the variance. Logarithms to the base 2 were used for convenience in interpretation: a log-ratio of plus or minus one corresponds to a red to green ratio of 2:1 or 1:2, respectively.

Log-ratios were normalized using an intensity-dependent normalization similar to Yang et al. (2001). Two normalizations were performed in succession: (1) a normalization that adjusted the log-ratio based on the average of the log-transformed red and green intensities (called “A”) and (2) a subsequent normalization that adjusted the

log-ratio based on the Euclidean distance of the spot from the center of the array. This second normalization was performed to counter illumination differences between the center and edges of the array. Both normalizations used the "lowess" procedure in S-PLUS (Venables and Ripley, 1999) to draw a smooth curve through the data while down-weighting outliers that might affect the fit. The log-ratio for each spot was normalized by subtracting the value of the smooth curve at the "A" value (or Euclidean distance) associated with that spot.

An expression ratio was then computed for each gene on a slide by averaging the normalized log-ratios for all cDNA target spots representing that gene. The resulting ratios for each gene were then averaged across replicate hybridizations. Log-ratios were then converted into fold-differences.

Statistical analysis of expression differences among tissues

For each gene, ten log-ratio measurements (two mice x five tissues) were combined into an F statistic to determine whether expression ratio measurements varied among tissues. The F statistic consisted of a ratio of a numerator (the variability of the average log-ratio measurements among tissues) to a denominator (the variability of measurements for a given tissue between mice). Higher variability *among* tissues than *within* a tissue indicated that the average log-ratios differed between tissues. The numerator was computed by summing twice the squared differences between individual tissue average log-ratios (averaged across the two observations for a tissue) and the overall average log-ratio, and dividing by four (the number of tissues minus one). The denominator was computed by summing the squared difference between each

measurement and the average of the two measurements for that tissue, and dividing by five (the number of measurements minus the number of tissues). Some F statistics were inflated by artificially low denominators (data not shown); therefore, the denominator was replaced by the maximum of the computed denominator for that gene and the median of the computed denominators across all genes.

The significance of each F-ratio was assessed by re-sampling methods (Good, 1994; Dudoit et al., in press). Briefly, if there were no expression differences among tissues, then the tissue label associated with each measurement would be considered irrelevant and could be permuted without affecting the random properties of the F statistic. An F statistic was computed for each possible assignment of the labels to measurements. The p value was then calculated as the proportion of the resulting F statistics that are at least as large as the statistic actually obtained. For a balanced experiment of five tissues and two mice per tissue, the number of distinct assignments is 113,400. However, these assignments can be broken down into 945 distinct groups, each of size 120, in which the statistic has the same value. Hence, in this experimental design only 945 F statistics need to be evaluated for each gene, and the p value is one of $1/945$, $2/945$, ..., $945/945$.

Adjusted p values were calculated to account for the fact that 412 p values, one for each gene with a full complement of ten measurements, were calculated. The step-down algorithm described by Dudoit et al., and based on Algorithm 4.1 of Westfall and Young (1993), was applied to adjust p values upward.

Expression ratios for a gene were considered indistinguishable whenever the unadjusted p value exceeded 0.05 or the log-ratios differed by less than 2.6 times an

estimate of the residual standard deviation for that gene. The number 2.6 corresponds to the threshold for Fisher's LSD method of multiple comparisons, which simplifies to the value of the upper 0.25 tail of the t distribution with five degrees of freedom.

Cluster analysis of gene expression

Expression trends were evaluated using CLUSFAVOR (CLUSter and Factor Analysis Using Varimax Orthogonal Rotation; Peterson, 2002). Briefly, \log_{10} gene expression ratios were clustered using the centroid average Euclidean distance between joining nodes. The resulting dendograms identified groups of differentially expressed genes among tissues.

Northern blot hybridization

Expression ratios were verified for *Araf*, *Catalase*, *Cdk2*, *Gadd153*, *Gpx1*, *Gstp2*, *Rad52*, *Stat4* and *Tdg* using Multiple Tissue Northern (MTN) Blots (CLONTECH Laboratories, Inc., Palo Alto, CA). Probes were generated by PCR-amplification of mouse QUICK-Clone cDNA (CLONTECH Laboratories, Inc.) using the following primer sequences: *Araf* fwd: 5'-ATCCGCTCCACATCTACTCC-3' and rev: 5'-CTATCCCCAAACCCAAGAGG-3'; *Cas-1* fwd: 5'-CCGTTTCGGTTCTCCACAGTC-3 and rev: 5'-TGC GTTCTTAGGCTTCTCAG-3'; *Cdk2* fwd: 5'-TGCAGAGGGGTCCATCAAGC-3' and rev: 5'-AGGCCAGGGTCAAGTCAGA-3'; *Gadd153* fwd: 5'-GTCCCTGCCTTTCACCTTGG-3' and rev: 5'-GGCGCTCGATTCNTGCTTG-3'; *Gpx1* fwd: 5'-CCACCGTGTATGCCTTCTCC-3' and rev: 5'-AGGCTATCCAAAAGGTGACA-3'; *Gstp2* fwd: 5'-

TGGAGACCTCACCTTTACC-3' and rev: 5'-CCACTACTGTTTGCCATTGC-3';
Rad52 fwd: 5'-GTTACAATGGCTGGGCACAC-3' and rev: 5'-
CCCACATTTCAAGGTTCTCT-3'; *Stat4* fwd: 5'-ACTGGGAGTAAAGGAAACGAG-
3' and rev: 5'-GCACCAAGTGAGAAAGAGAGC-3'; and *Tdg* fwd: 5'-
GACCCGAGAGCAGGAAGAAG-3' and rev: 5'-CCCCGGACTCGTTACTCACC-3'.
Primer specificity was confirmed using NCBI's BLAST. Amplicon sizes were confirmed
on a 2% agarose gel. PCR products were purified, labeled, hybridized, and visualized as
described by Coleman et al. (2000). Briefly, 100 ng of each probe was [α -³²P]dCTP-
labeled using the Megaprime DNA Labeling System (Amersham, Arlington Heights, IL).
Hybridizations were performed according to manufacturer's instructions. Probes were
visualized with the Storm 860 PhosphorImager (Molecular Dynamics, Sunnyvale, CA).
Quantitation was performed using a gel documentation system (NucleoTech, San Mateo,
CA). All blots were normalized to actin- β . Normalized gene expression ratios were
calculated by dividing each tissue measurement by the testis measurement. Ratio
differences \leq 2.0-fold indicated similar expression among tissues.

3.4 Results

Normalization and expression ratio measurement variation

The plot of the non-normalized average log intensity vs. log expression ratio for each cDNA target in the pooled testes vs. pooled testes hybridization (i.e., two separate aliquots from pooled testes reference labeled with either Alexa Fluor 488-dUTPs or Alexa Fluor 546-dUTPs) showed a striking dependence on intensity (Figure 3.1A). After applying a lowess normalization (see Materials and Methods), the data scattered around a

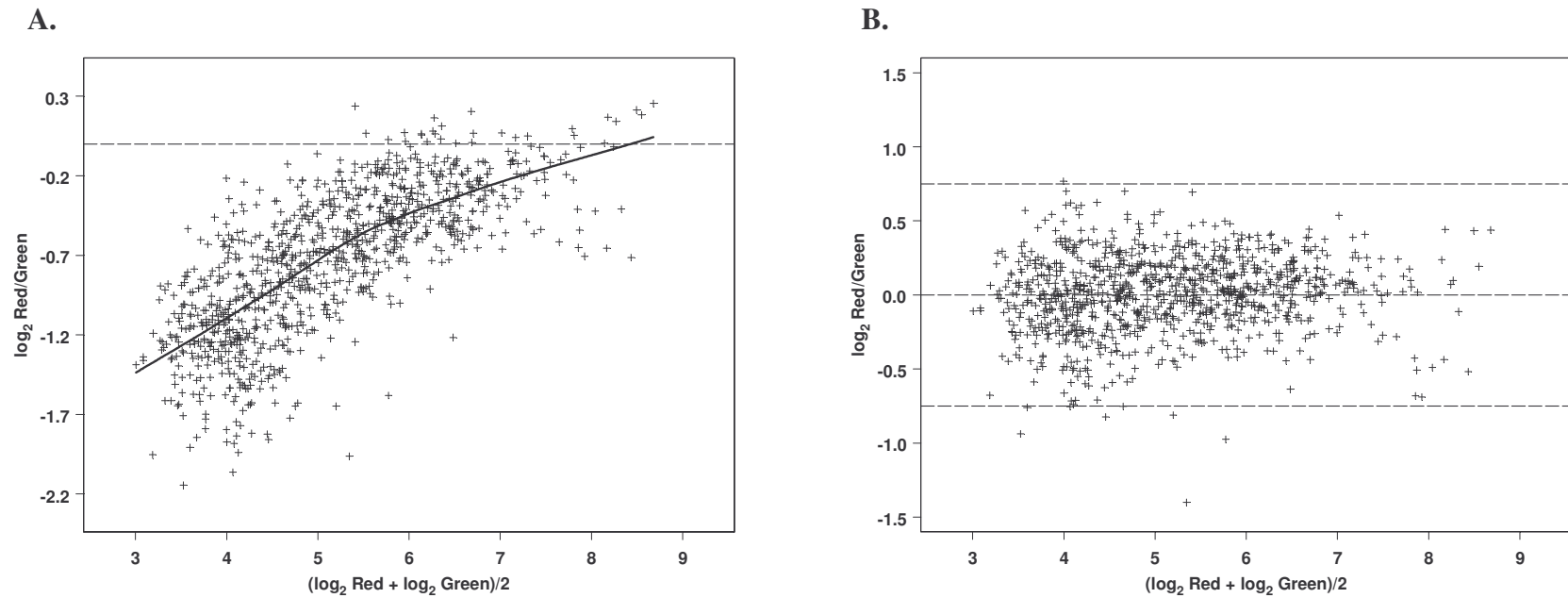


Figure 3.1. Intensity-based normalization of cDNA microarray data. Two aliquots from a pool of B6C3F1 testes were labeled with either Alexa Fluor 488 or Alexa Fluor 546 and co-hybridized onto the microarray. The average of the log of the red and green intensities (x-axis) vs. the log of the red:green ratio (y-axis) was plotted for each spot. The horizontal dashed line at a log-ratio of zero represents the expected ratio of 1. **(A)** The average log-intensity vs. log ratio for each cDNA spot on the array prior to normalization. The solid line represents the smooth curve drawn using the lowess procedure. **(B)** Data distribution following intensity-dependent normalization (see Materials and Methods section for details).

log-ratio of zero (Figure 3.1B), as expected, with little residual bias due to intensity. In replicate pooled testes vs. pooled testes hybridizations, >95% of the cDNA targets had log-ratios between -0.75 and $+0.75$, corresponding to ratios between 0.60 and 1.68. Because these normalization methods effectively minimized fluorophore- and intensity-based biases, they were applied to each microarray hybridization in this study.

To determine whether transcripts were similarly represented in the pooled testes sample and individual testis samples, expression ratio profiles for replicate individual testis vs. pooled testes hybridizations were analyzed. Gene expression ratios were not significantly different between the replicate hybridizations (data not shown); therefore, the replicate ratios for each gene were averaged (gray columns in Figure 3.2). These average gene expression ratios were similarly distributed around the expected value of 1.0 with ~98% of all individual testis vs. pooled testes ratios between 0.60 and 1.68 (sum of all columns ≤ 1.7 in Figure 3.2). Because this ratio distribution was not significantly different from the pooled testes vs. pooled testes ratio distribution, pooled testes were used as the reference tissue in subsequent hybridizations.

Animal-to-animal variation was assessed by analyzing the differences in gene expression ratios between replicate hybridizations of testis, spleen, brain or liver vs. pooled testes (Table 3.1). Average expression ratio differences at the 25th, 50th, 75th and 95th percentiles were 0.06, 0.13, 0.24, and 0.62, respectively. These differences were not significantly different from those observed for the pooled testes vs. pooled testes hybridizations. The minor amount of variation between replicate animals allowed gene expression ratios from replicate hybridizations to be averaged for subsequent analyses.

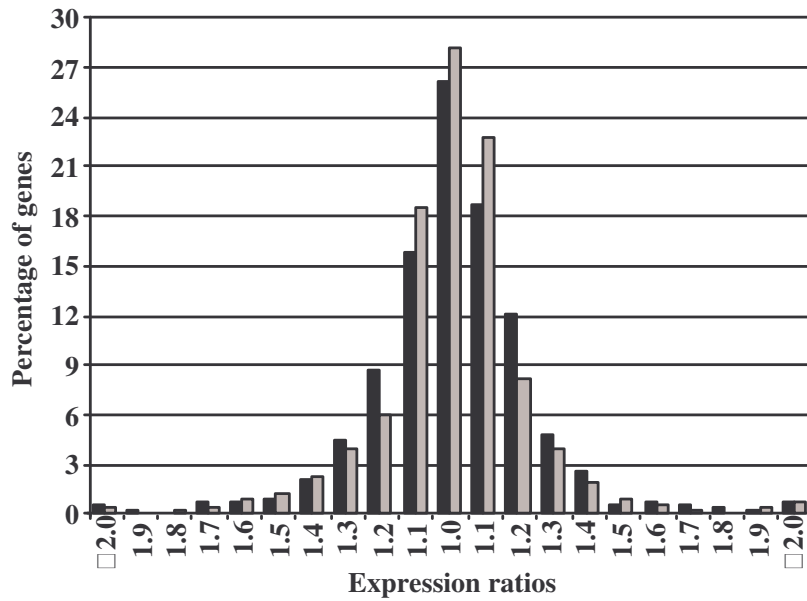


Figure 3.2. Distribution of normalized expression ratios for testes vs. testes hybridizations. The graph illustrates the percentage of genes (y-axis) with a given expression ratio (x-axis) following testes vs. testes hybridizations. Black columns represent the distribution of average expression ratios obtained from two independent hybridizations of pooled testes vs. pooled testes. Gray columns represent the distribution of average expression ratios obtained from two independent hybridizations of individual testis vs. pooled testes. Ratios that were higher in the Alexa Fluor 546-labeled pooled testes or individual testis samples are to the left of 1.0. Ratios that were higher in the Alexa Fluor 488-labeled pooled testes reference are to the right of 1.0.

Table 3.1. Differences in the gene expression ratios from replicate independent hybridizations.

<u>Percentile</u>	<u>Differences between replicate ratio measurements^a</u>					<u>Pooled testes^c</u>
	<u>Tissues from individual mice^b</u>					
	<u>Testis</u>	<u>Spleen</u>	<u>Brain</u>	<u>Liver</u>	<u>Average</u>	
25th	0.05	0.06	0.06	0.07	0.06	0.09
50th	0.10	0.12	0.13	0.16	0.13	0.17
75th	0.18	0.24	0.27	0.29	0.24	0.28
95th	0.34	0.57	0.69	0.91	0.62	0.62

^aFor each gene, the absolute value of the difference in replicate hybridization expression ratio measurements was determined. Differences were rank-ordered from lowest to highest and were examined at the 25th, 50th, 75th, and 95th percentiles.

^bTwo hybridizations were performed for individual mouse tissues (testis, spleen liver, or brain) vs. pooled testes. Absolute differences were not calculated for heart because >50% of the cDNA targets were irregularly shaped in one hybridization.

^cPooled testes vs. pooled testes hybridizations were used as an indicator of technical variation.

Comparison of gene expression profiles among healthy adult mouse tissues

Based on the results above, gene expression ratios ≤ 0.68 or ≥ 1.68 were further evaluated for differential expression among tissues. Ratios ≤ 0.60 indicated higher expression in the testis, while ratios ≥ 1.68 indicated higher expression in brain, heart, liver or spleen. Figure 3.3 summarizes the distribution of the 152 genes that were differentially expressed in at least one tissue-tissue comparison (brain, spleen, heart, or liver vs. testis). See Appendix A for a complete list of the 152 genes showing differential expression among mouse tissues. A subset of 105 genes showed significantly higher expression (unadjusted p values ≤ 0.05) in a single tissue (71 genes), in two tissues (23 genes), in three tissues (7 genes) or in four tissues (4 genes). Relatively higher baseline levels of gene expression were observed most often in heart (24 genes higher only in heart; 37 genes higher in heart and at least one other tissue) and testis (21 genes higher only in testis; 43 genes higher in testis and at least one other tissue).

Genes with expression ratios outside of the 0.60 to 1.68 range were further analyzed for tissue-specific differences in expression using cluster analysis (Figure 3.4). When a vertical line was drawn through the cluster branch points so that the joining nodes to the immediate right of the line would group only ~10% of the genes into sub-clusters of ≤ 5 genes, nine major (>5 genes) and seven minor (≤ 5 genes) sub-clusters were observed. Among these sub-clusters, we identified groups of genes whose expression was specifically higher in either heart (sub-cluster I), brain (sub-cluster III), liver (sub-cluster V), spleen (sub-cluster X), or testis (sub-clusters VII, VIII, and XI). For example, *Acrv1* and *Sycp3*, which are known to be involved in spermatogenesis and fertility, were part of the testis-specific sub-cluster XI. We also identified clusters of

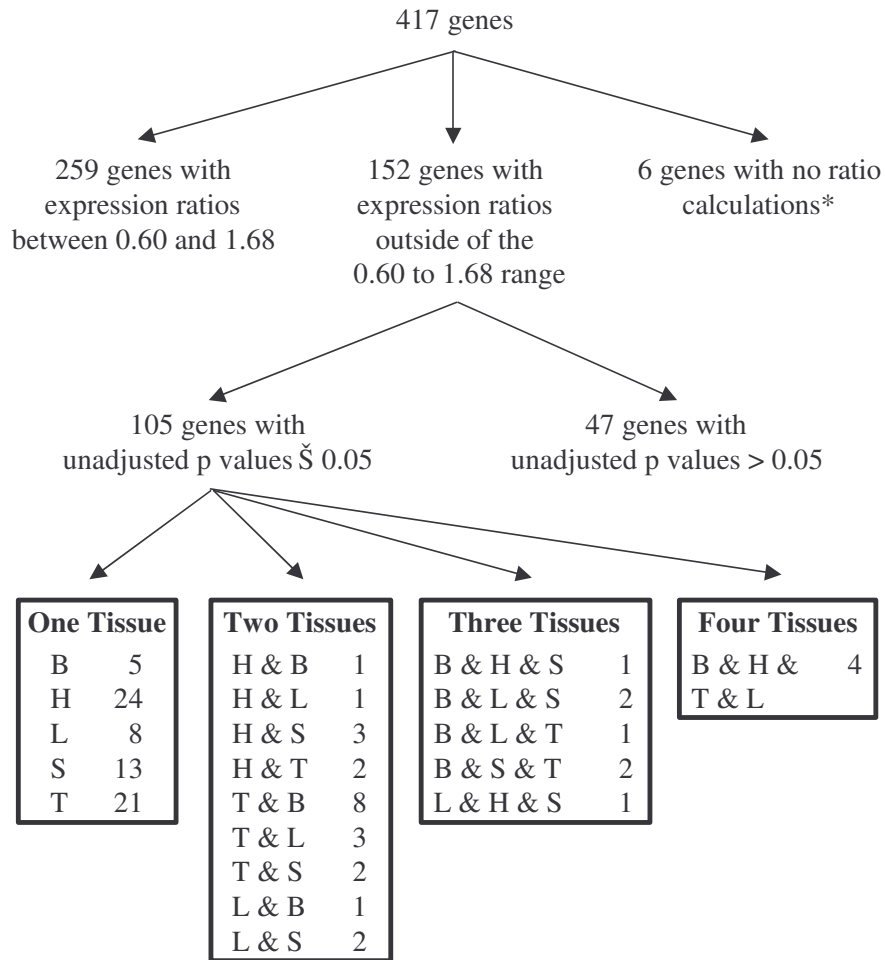


Figure 3.3. Distribution of gene expression results among tissues. Genes with ratios ≤ 0.60 are considered to be higher in testes. Genes with ratios ≥ 1.68 are considered to be higher in brain, liver, spleen, and/or heart. H = Heart, L = Liver, T = Testis, S = Spleen, B = Brain. *No ratio calculations were made for 1 or more replicate hybridizations because the target spots representing these genes were obscured by debris.

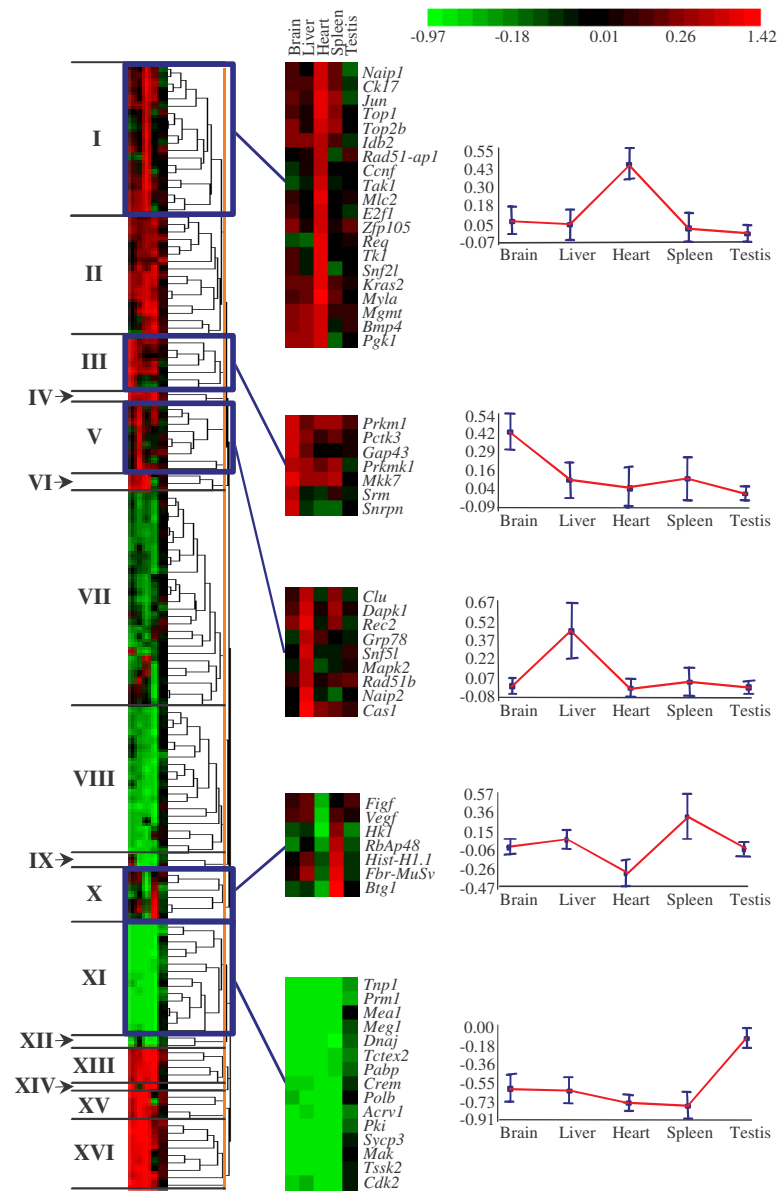


Figure 3.4. Cluster analysis of genes using CLUSFAVOR. Each column in the cluster represents one tissue (from left to right: brain, liver, heart, spleen, and testis). Each row represents one gene. Joining nodes to the immediate right of the orange vertical line determined the extent of each sub-cluster. The sixteen sub-clusters are indicated by Roman numerals I-XVI. Highlighted regions of the dendrogram illustrate examples of genes that cluster together based on tissue-specific elevated expression (e.g., top sub-cluster shows genes with elevated heart expression compared to other tissues). Graphs corresponding to each sub-cluster illustrate the mean log expression ratio for the sub-cluster (\pm the standard deviation for that sub-cluster) in each tissue.

genes with relatively higher expression in either two, three, or four tissues (e.g., *Gstp2*, which is involved in stress response, was higher in heart, brain, and liver than in spleen and testis).

Genes with expression ratios outside of the 0.60 to 1.68 range were then grouped into one of eight biological pathways (Table 3.2). The stress response pathway had the greatest percentage of genes (55%) with differential expression while genes involved in DNA repair had the lowest percentage (20%) of differential expression. Approximately 32% of the genes involved in damage control were differentially expressed among tissues.

As shown in Table 3.3, the stress response genes were further subdivided, based on function, into oxidative stress or heat shock response. Oxidative stress genes were generally expressed highest in the heart (e.g., *Sod3*) and in the liver (e.g., *Cas1*), with the exception of heme-oxygenase 2 (*Hmox2*) which exhibited the highest expression in testis. Among the heat shock response genes, the *Dnaj*-related genes had higher expression in the testis, while the *Hsp110*-related gene *Apg1b* tended to have elevated expression in non-testis tissues.

For genes involved in DNA repair, the highest levels of expression were generally observed in testis (e.g., *Pol-b*, *Pcna* and *Mre11*). However, a few genes exhibited higher expression in other tissues (i.e., *Tdg* and *Mgmt* were highly expressed in spleen and heart, respectively).

Genes with functions related to damage control (i.e., apoptosis and cell cycle) exhibited differential expression among tissues as well. The cell cycle-related genes tended to have higher expression in testis (e.g., *Cdk2*), while the genes involved in

Table 3.2. Distribution of F ratios for differentially expressed genes with respect to biological pathway.

Biological pathway	# genes arrayed	% differentially expressed genes ^a	F ratio ^b			
			<6	6-10	11-25	>25
Chromatin-related	24	46	13 (4)	21 (21)	4 (0)	8 (8)
Damage control	65	32	14 (6)	5 (5)	8 (6)	6 (6)
DNA repair	51	20	12 (2)	4 (4)	2 (2)	2 (2)
Growth regulation	80	36	16 (6)	5 (5)	11 (11)	4 (4)
Meiosis / Spermatogenesis	20	45	30 (10)	5 (0)	5 (5)	5 (5)
Stress response / Stress induced	22	55	14 (0)	5 (5)	9 (9)	27 (27)
Transcription / Translation	58	41	14 (5)	10 (9)	12 (12)	5 (5)
Miscellaneous	97	37	10 (0)	8 (8)	7 (7)	11 (11)
Total	417	36	14 (4)	7 (7)	8 (7)	7 (7)

^aGenes with expression ratios outside of the 0.60 to 1.68 range in at least one tissue.

^bPercentage of genes in each F ratio category; Number in parentheses indicate the percentage of genes with significant unadjusted p values (≤ 0.05). See Materials and Methods.

Table 3.3. Differentially expressed stress response, DNA repair and damage control genes.

Pathway & gene	Function ^a	Ratio to pooled testes					F statistic			Unadj. p value ^c
		Testis	Brain	Liver	Spleen	Heart	Num.	Denom.	Ratio ^b	
Stress response										
<i>Adk1</i>	GS	0.9	0.8	0.6	0.5*	0.7	0.19	0.09	2.1	0.204
<i>Apg1b</i>	HS	1.1	1.8*	2.2*	1.5	2.5*	0.42	0.02	9.2^	0.001
<i>Dnaj</i>	HS	0.7	0.2*	0.2*	0.1*	0.2*	1.99	0.03	43.5^	0.001
<i>Dnajl1</i>	HS	0.9	0.6	0.6	0.3*	0.5	0.94	0.02	20.7^	0.001
<i>Cas-1</i>	OS	1.1	1.0	8.3*	1.2	1.4	3.26	0.10	31.3	0.017
<i>Gadd153</i>	OS	1.1	1.8	1.9	0.6	5.4*	2.79	0.10	28.6	0.003
<i>Gpx1</i>	OS	1.1	3.4*	4.2*	2.1*	26.5*	5.93	0.13	46.2	0.003
<i>Gstp2</i>	OS	1.1	4.0*	5.2*	0.9	18.0*	6.05	0.06	106.0	0.001
<i>Hmox2</i>	OS	0.8	0.6*	0.5*	0.3*	0.3*	0.82	0.05	15.1	0.001
<i>Sod3</i>	OS	1.2	2.6*	2.6*	0.8	8.8*	3.59	0.06	62.4	0.002
DNA repair										
<i>Pol-β</i>	BER	1.0	0.5*	0.3*	0.2*	0.2*	2.32	0.05	50.8^	0.001
<i>Tdg</i>	BER	1.3	1.9*	1.8*	4.4*	1.9*	0.88	0.01	19.3^	0.001
<i>Ung</i>	BER	0.9	0.5*	0.5	0.5*	0.3*	0.56	0.29	2.0	0.221
<i>Mgmt</i>	DDR	1.1	1.4	1.6	1.1	2.8*	0.61	0.13	4.9	0.054
<i>Mre11</i>	DSBR	1.0	0.8	0.6	0.5*	0.7	0.32	0.02	7.0^	0.001
<i>Rad51B</i>	DSBR	1.3	1.1	2.0*	1.1	1.0	0.28	0.08	3.3	0.082
<i>Rfc40</i>	NER	0.9	0.6	0.5*	0.4*	0.5	0.40	0.11	3.8	0.100
<i>Pcna</i>	NER	1.0	0.5*	0.4*	0.7	0.3*	1.02	0.14	7.3	0.017
Damage control										
<i>Bmp4</i>	AP	1.1	1.5	1.8*	0.9	2.7*	0.75	0.04	16.4^	0.001
<i>Gap43</i>	AP	1.1	2.5*	1.3	1.0	1.0	0.62	0.02	13.7^	0.001
<i>Gas2</i>	AP	0.9	2.2*	3.0*	8.6*	3.7*	2.73	0.33	8.3	0.015
<i>Naip2</i>	AP	0.9	1.0	4.9*	0.7	0.9	2.53	0.12	20.5	0.074
<i>Rip1</i>	AP	1.1	3.6*	3.5*	1.0	19.4*	6.02	0.13	47.8	0.007
<i>Smp30</i>	AP	1.1	2.0*	10.5*	1.8*	2.7*	3.06	0.07	41.3	0.002
<i>Tfar15</i>	AP	0.9	1.1	0.9	0.7	0.5*	0.32	0.01	7.0^	0.001
<i>Mcl1</i>	AP, CC	1.2	2.5*	4.3*	2.4*	9.4*	2.47	0.15	16.5	0.003
<i>Btg1</i>	CC	1.0	0.8	0.8	5.7*	0.4*	3.99	0.07	59.6	0.003
<i>Cdc2</i>	CC	1.0	0.6*	0.7	0.6	0.6	0.17	0.07	2.5	0.059
<i>Cdc2a</i>	CC	1.0	0.7	0.6	0.6	0.5*	0.27	0.05	5.4	0.001
<i>Cdc42</i>	CC	1.2	2.9*	2.3*	2.9*	2.2*	0.54	0.05	10.2	0.002
<i>Cdk2</i>	CC	0.9	0.3*	0.5*	0.2*	0.2*	1.81	0.12	15.4	0.008
<i>Paga</i>	CC	0.9	1.7	12.9*	1.3	2.7*	4.43	0.03	97.1^	0.001

^aGS = Generalized Stress Response; OS = Oxidative Stress; HS = Heat-Shock; BER = Base Excision Repair; DDR = Direct Damage Reversal; DSBR = Double Strand Break Repair; NER = Nucleotide Excision Repair; AP = Apoptosis; CC = Cell Cycle

^bF ratio = Numerator/Denominator (see Materials and Methods)

^cUnadjusted *p* values for the F ratios (see Materials and Methods)

*Replicate hybridizations had expression ratios \leq 0.60 or \geq 1.68.

^Denominator was replaced by the median denominator value (0.0456) before calculating the F ratio.

apoptosis tended to be more highly expressed in non-testis tissue (e.g., *Smp30* in liver and *Rip1* in heart).

Northern blot confirmation of cDNA microarray results

Expression ratios obtained from the cDNA microarrays were verified for nine genes using northern blots. These genes were selected across the following biological pathways: stress response (*Cas1*, *Gstp2*, *Gadd153*, *Gpx1*), DNA repair (*Rad52*, *Tdg*), damage recognition (*Cdk2*), growth regulation (*Araf*), and transcription/translation (*Stat4*). Although the specific expression ratios varied slightly between the microarrays and the multiple tissue northern blots, there was strong agreement overall between the relative rank-order of expression among tissues (Figure 3.5). For example, although the microarray-based ratio measurements for *Gpx1* were consistently higher than the northern blot measurements, the relative rank-order of expression was very similar with the exception of heart, which had a much higher ratio than all of the other tissues according to the microarray. Additionally, both methods showed that the oxidative stress response gene *Cas1* exhibited the highest levels of expression in the liver (microarray: 8.3-fold higher than testis; northern blot: 9.0-fold higher than testis) with similar levels of expression among the testes, spleen, and brain (microarray range: 1.0 to 1.4-fold; northern blot: 0.8 to 1.0-fold).

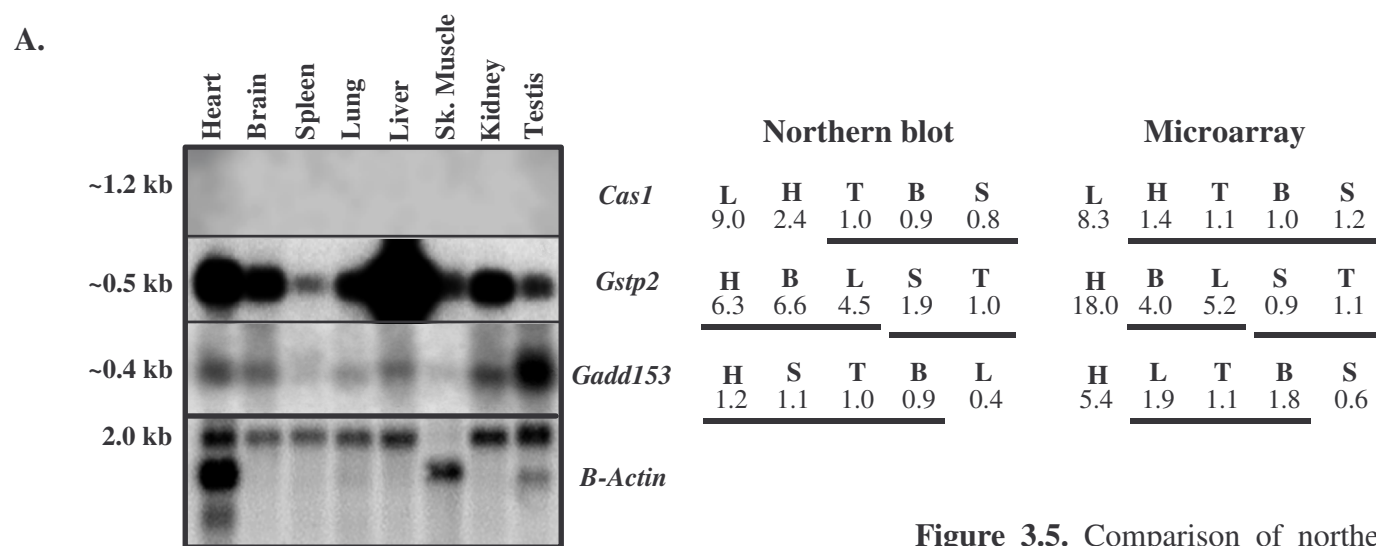


Figure 3.5. Comparison of northern blot and cDNA microarray results for selected genes. Expression ratios were calculated as the value for the tissue of interest divided by the value for testis. Tissues were rank-ordered from highest to lowest expression on the northern blot. (H = Heart, L = Liver, T = Testis, S = Spleen, B = Brain.) Genes differing by ≤ 2 -fold among tissues on the northern blot were considered to be similarly expressed (horizontal bar). Microarray-based log expression ratios for each gene were also rank-ordered (highest to lowest). If the log-ratio difference between 2 tissues was $< 2.6 \times$ (square root of the residual standard deviation), the gene was similarly expressed (horizontal bar). **(A)** Northern blot images and northern blot vs. microarray results. **(B)** Northern blot vs. microarray results for additional genes.

B.

	Northern blot					Microarray				
<i>Araf</i>	H	L	T	S	B	H	L	T	S	B
	1.3	1.1	1.0	0.9	0.7	1.0	0.9	1.0	0.7	0.9
<i>Cdk2</i>	T	S	L	B	H	T	S	L	B	H
	1.0	0.6	0.4	0.4	0.2	0.9	0.2	0.5	0.3	0.2
<i>Gpx1</i>	S	B	T	H	L	S	B	T	L	H
	1.4	1.2	1.0	0.9	0.7	2.1	3.4	1.1	4.2	26.5
<i>Rad52</i>	L	H	T	B	S	L	H	T	B	S
	1.4	1.0	1.0	0.5	0.0	0.9	0.8	1.1	0.8	0.8
<i>Stat4</i>	T	L	S	B	H	T	L	S	B	H
	1.0	0.1	0.2	0.2	0.0	0.8	1.0	0.2	0.3	0.2
<i>Tdg</i>	B	S	H	L	T	S	B	H	L	T
	2.1	1.9	1.6	1.0	1.0	4.4	1.9	1.9	1.8	1.3

3.5 *Discussion*

cDNA microarray hybridizations revealed tissue-specific differences in the basal levels of transcript abundance for chromatin-related, damage recognition, DNA repair, growth regulation, meiosis, stress response, and transcription/translation genes. Overall, 152 of the 417 arrayed genes (36%) had expression ratios outside of the 0.60 to 1.68 range, and 25% were significantly differentially expressed among healthy tissues.

Experiments with split samples and replicate hybridizations demonstrated the precision and reproducibility of the microarray protocols used in this study. The generally strong agreement between the microarray and northern blot data demonstrated the accuracy of our findings.

The unadjusted p-values used to assess the significance of differential gene expression among tissues were expected to be somewhat overly optimistic. With over 400 genes being tested, we expected ~20 genes to appear significantly different at the 0.05 level, even when no expression differences existed between tissues. However, after applying an extremely conservative indicator of differential expression (see Materials and Methods), only ~7% of the genes were still significantly differentially expressed (adjusted $p \leq 0.05$) among tissues. In fact, when evaluated by their adjusted p-values, only genes with F ratios ≥ 25 were considered to be significantly different among tissues. The challenge for future microarray studies will be to develop statistical analyses methods that compromise between the overly conservative adjusted p-values and the overly permissive unadjusted p-values.

Sixteen general patterns of differential tissue expression were observed following cluster analysis. Among the major sub-clusters, we identified genes that were similarly

expressed across tissues with the exception of higher expression in (a) one tissue (testis, heart, liver, brain, or spleen), (b) two tissues (heart and spleen), or (c) three tissues (brain, liver, and heart). We also identified minor sub-clusters of genes that had similar expression across tissues with the exception of higher expression in (a) two tissues (brain and testis), (b) three tissues (brain, liver, and heart or brain, heart, and spleen), or (c) four tissues (brain, liver, heart, and spleen or brain, heart, spleen, and testis).

After assigning genes into one of eight biological pathways, we observed that the stress response genes showed the most variation in basal gene expression among tissues, DNA repair genes had the lowest variation among tissues, and damage control genes showed intermediate levels of variation.

Stress response genes

Oxidative damage produced by reactive oxygen species (ROS) has been implicated in mutagenesis, inflammatory and autoimmune diseases, aging and cancer (Guyton et al., 1996; Lenzen et al., 1996; Loft and Poulsen, 2000; Li et al., 2001). ROS are continuously generated through the process of aerobic metabolism; however, the extent of ROS-related damage in cells depends upon the balance between ROS production, antioxidant scavenger capacity, and the efficiency of repair (Hollensworth et al., 2000; Loft and Poulsen, 2000).

We found that higher expression of oxidative stress response genes was generally observed in liver, heart, and brain compared to spleen and testis (with the exception of *Hmox2*). Similar tissue-specific expression patterns have been reported for several of these oxidative stress-related genes, including catalase and glutathione peroxidase,

(Lenzen et al., 1996). These results, considered together with the observation that certain stress genes (e.g., Gadd153) have significantly higher basal expression levels in O₂-resistant versus control cell lines (Guyton et al., 1996), raise the question of whether these tissues may have differential responses to oxidative stress.

Heat shock response provides protection against a variety of stress-inducing agents (including heat shock, oxidative stress, heavy metals, and inflammation) by either refolding or degrading damaged proteins (Jolly and Morimoto, 2000). We found that the differentially expressed heat shock response genes typically exhibited the highest expression in the testis, with the exception of *Apg1b*. Since up-regulation of heat shock gene expression occurs during stress and non-stress (e.g., cell cycle and differentiation) conditions (Jolly and Morimoto, 2000; Pirkkala et al., 2001), it is uncertain whether our finding may simply reflect variations in tissue physiology or tissue-specific variations in the ability to respond to stress stimuli.

Damage control genes

In our study, genes involved in apoptosis and/or cell cycle regulation were categorized as damage control genes. Cells that have been damaged by various environmental exposures (e.g., ionizing radiation) may be removed by apoptosis, or programmed cell death. However, apoptosis is also critical for day-to-day health maintenance in the absence of exogenous exposures. It has been shown to play an important role in processes such as physiological cell turnover (Medh and Thompson, 2000) and development (Meier et al., 2000). Our results indicate that the baseline expression of apoptosis-associated genes is usually highest in liver and heart and lowest

in testis and spleen. We found that *Smp30* had significantly higher basal gene expression levels in the liver than the other tissues evaluated. This finding correlates well with the suggestion by Fujita that SMP30 is essential for some of the highly differentiated functions of the liver (1999). Furthermore, we found that the differentially expressed genes function in different portions of the apoptotic pathway (e.g., initial signaling events, caspase substrates, etc.).

While the cell division cycle genes *Cdc2* and *Cdc2a* and the cyclin dependent kinase gene *Cdk2* exhibited the highest expression in testis (Table III), the remaining cell cycle-related genes exhibited higher expression in non-testis tissues even though some of them have similar functions. For example, both *Cdk2* and *Cdc42* are involved in progression through G1 into the S-phase of the cell cycle (Olson et al., 1995; Tsai et al., 1993). Yet, we found that these genes have very different patterns of expression among tissues.

DNA repair genes

We found that genes involved in nearly every DNA repair pathway were differentially expressed among tissues and that testis generally exhibited the highest levels of expression. Genes involved in base excision repair (BER) tended to be more highly expressed in the testis, with the exception of *Tdg* which exhibited the highest levels of expression in the spleen and other somatic tissues. These findings correlate well with the finding that BER activity initiated by uracil DNA glycosylase is highest in mixed testicular germ cell extracts (Intano et al., 2001). More specifically, Intano et al.

found that protein levels for polymerase β were much higher in mixed germ cells than in brain and liver.

We found that the direct damage reversal (DDR) enzyme *Mgmt* (or *O6-methylguanine-DNA methyltransferase*), which is responsible for removing DNA lesions caused by alkylation, had the highest expression in heart. To our knowledge, this is the first report that *Mgmt* transcripts have higher abundance in the heart than in liver, brain, spleen and testis.

Among the genes involved in double stranded DNA break repair, we found that *Mre11* showed the highest expression in testis which complements the finding of Chamankhah et al. in human testis (1998). Our finding that *Rad51B* was most highly expressed in the liver contrasts the suggestion by Albala et al. who reported that the expression of this gene is highest in actively recombining tissues (1997). However, it should also be noted liver was not among the tissues evaluated in the study by Albala et al.

Our results show that a few of the nucleotide excision repair (NER) genes represented on our cDNA microarray were differentially expressed among tissues. Genes encoding the p40 subunit of replication factor C (RFC) and *Pcna* exhibited the highest levels of expression in testis. These results are particularly interesting because the p40 subunit represents 1 of 5 RFC subunits required for *Pcna*-dependent DNA synthesis (Uhlmann et al. 1997).

Although DNA repair genes showed the least variation in expression among all of the biological pathways surveyed, we found that genes involved in four out of five DNA repair pathways were differentially expressed among tissues. A possible explanation why

differential expression was not detected for more of the DNA repair genes represented on our microarray is that low signal intensities may have affected our ability to detect expression differences for certain genes. Other groups have also reported low transcript abundance for certain DNA repair genes under normal physiological conditions (e.g., Xrcc1, Walter et al., 1994).

Our finding that transcriptional profiles for DNA repair genes vary among tissues may suggest that tissues have different DNA repair capacities. This suggestion is further supported by the finding that the induction of genotoxic damage, by chemicals such as acrylamide and acrylonitrile, may result in tissue-specific carcinogenic activity (Butterworth et al. 1992).

Summary

Through the use of cDNA microarray technology, this study takes an important step toward understanding global differences in gene expression among healthy adult tissues. Differences in basal gene expression levels among tissues may contribute to the observed differential tissue responses to DNA damaging agents such as cadmium, lead, ionizing radiation, etc. (Amundson et al. 1999; Fornace et al. 1999; Valverde et al. 2000; Valverde et al. 2002). By characterizing the differential basal expression of genes involved in DNA damage response, recognition, and repair processes among tissues, this research contributes to our understanding of tissue-specific responses to genotoxic agents and may also provide insight on the differential susceptibility to the onset and progression of cancer.

Future studies are needed to address the relative contributions of tissue microenvironments to differential basal expression levels among healthy tissues and to evaluate the involvement of these microenvironments in differential tissue responses to DNA damaging conditions. In addition, endeavors to identify groups of genes with coordinated expression (i.e., synexpression groups) and to assess transcript vs. protein abundance as well as post-transcriptional and post-translational differences among tissues will further enhance our understanding of variations in the response to and processing of DNA damage among healthy tissues.

3.6 Acknowledgements

Special thanks to Brenda Marsh for her preparation of the cDNA microarray targets, Dr. Rajiv Raja for his assistance in arraying the cDNA targets, Dorreyah Schahin-Reed for her assistance with the tissue extractions, and Dr. Leif E. Peterson for his assistance with the CLUSFAVOR analysis of the microarray data. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48 with support from NIH ES09117; NIEHS Superfund P4ZES04705; the Low Dose Radiation Research Program, Biological and Environmental Research (BER), U.S. DOE grant KP110202; the University of California and West Virginia University.

3.7 *References*

Albala, J.S., Thelen, M.P., Prange, C., Fan, W., Christensen, M., Thompson, L.H., and Lennon, G.G. (1997). Identification of a novel human *RAD51* homolog, *RAD51B*. *Genomics* 46, 476-479.

Alizadeh, A.A., Ross, D.T., Perou, C.M., and van de Rijn, M. (2001). Towards a novel classification of human malignancies based on gene expression patterns. *J. Pathol.* 195, 41-52.

Amundson, S.A., Bittner, M., Meltzer, P., Trent, J., and Fornace, A.J. (2001). Induction of gene expression as a monitor of exposure to ionizing radiation. *Radiat. Res.* 156, 657-661.

Amundson, S.A., Bittner, M., Chen, Y., Trent, J., Meltzer, P., and Fornace, A.J. (1999). Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene* 18, 3666-72.

Burns, T.F., Bernhard, E.J., and El-Deiry, W.S. (2001). Tissue specific expression of p53 target genes suggests a key role for KILLER/DR5 in p53-dependent apoptosis in vivo. *Oncogene* 20, 4601-12.

Butterworth, B.E., Eldridge, S.R., Sprankle, C.S., Working, P.K., Bentley, K.S., and Hurtt, M.E. (1992). Tissue-specific genotoxic effects of acrylamide and acrylonitrile. *Environ. Mol. Mutagen.* *20*, 148-55.

Chamankhah, M., Wei, Y.F., and Xiao, W. (1998). Isolation of *hMRE11B*: failure to complement yeast *mre11* defects due to species-specific protein interactions. *Gene* *225*, 107-116.

Coleman, M.A., Eisen, J.A., and Mohrenweiser, H.W. (2000). Cloning and characterization of HARP/SMARCAL1: A prokaryotic HepA-related SNF-2 helicase protein from human and mouse. *Genomics* *65*, 274-282.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* *14*, 457-60.

Dudoit, S., Yang, Y. H, Speed, T. P., and Callow, M. J. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*. In press.

Fornace, A.J., Amundson, S.A., Bittner, M., Myers, T.G., Meltzer, P., Weinstein, J.N., and Trent, J. (1999). The complexity of radiation stress responses: analysis by informatics and functional genomics approaches. *Gene Expr.* *7*, 387-400.

Fujita, T. (1999). Senescence marker protein-30 (SMP30): structure and biological function. *Biochem. Biophys. Res. Commun.* 254, 1-4.

Good, P. (1994). *Permutation Tests*. New York: Springer-Verlag.

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., and Smith, L.M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* 22, 5456-5465.

Guyton, K.Z., Spitz, D.R., and Holbrook, N.J. (1996). Expression of stress response genes *Gadd153*, *c-jun*, and *heme oxygenase-1* in H₂O₂- and O₂-resistant fibroblasts. *Free Radic. Biol. Med.* 20, 735-741.

Hollensworth, S.B., Shen, C.C., Sim, J.E., Spitz, D.R., Wilson, G.L., and LeDoux, S.P. (2000). Glial cell type-specific responses to menadione-induced oxidative stress. *Free Radic. Biol. Med.* 28, 1161-1174.

Intano, G.W., McMahan, C.A., Walter, R.B., McCarrey, J.R., and Walter, C.A. (2001). Mixed spermatogenic germ cell nuclear extracts exhibit high base excision repair activity. *Nucleic Acids Res.* 29, 1366-1372.

Jolly, C., and Morimoto, R.I. (2000). Role of the heat shock response and molecular chaperones in oncogenesis and cell death. *J. Nat. Can. Inst.* 92, 1564-1572.

Kegelmeyer, L.M., Tomascik-Cheeseman, L., Burnett, M.S., van Hummelen, P., and Wyrobek, A.J. (2001). A groundtruth approach to accurate quantitation of fluorescence microarrays. *SPIE Proceed.* 4266, 35-45.

Leasure, C.S., Chandler, J., Gilbert, D.J., Householder, D.B., Stephens, R., Copeland, N.G., Jenkins, N.A., and Sharan, S.K. (2001). Sequence, chromosomal location and expression analysis of the murine homologue of human *RAD51L2/RAD51C*. *Gene* 271, 59-67.

Lenzen, S., Drinkgern, J., and Tiedge, M. (1996). Low antioxidant enzyme gene expression in pancreatic islets compared with various other mouse tissues. *Free Radic. Biol. Med.* 20, 463-466.

Li, W.G., L., Miller, F.J., Zhang, H.J., Spitz, D.R., Oberley, L.W., and Weintraub, N.L. (2001). H_2O_2 -induced O_2 production by a non-phagocytic NAD(P)H oxidase causes oxidant injury. *J. Biol. Chem.* 276, 29251-29256.

Loft, S., and Poulsen, H.E. (2000). Antioxidant intervention studies related to DNA damage, DNA repair and gene expression. *Free Radic. Res.* 33, S67-83.

Medh, R.D., and Thompson, E.B. (2000). Hormonal regulation of physiological cell turnover and apoptosis. *Cell Tissue Res.* 301, 101-24.

Meier, P., Finch, A., and Evan, G. (2000). Apoptosis in development. *Nature* 407, 796-801.

Olson, M.F., Ashworth, A., and Hall, A. (1995). An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science* 269, 1270-2.

Peterson, L.E. (2002). Factor analysis of cluster-specific gene expression levels from cDNA microarrays. *Comput. Methods Programs Biomed.* In press.

Pirkkala, L., Nykanen, P., and Sistonen, L. (2001). Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J.* 15, 1118-31.

Pittman, D.L., Weinberg, L.R., and Schimenti, J.C. (1998). Identification, characterization, and genetic mapping of *Rad51d*, a new mouse and human *RAD51/RecA*-related gene. *Genomics* 49, 103-11.

Tsai, L.H., Lees, E., Faha B., Harlow, E., and Riabowol, K. (1993). The cdk2 kinase is required for the G1-to-S transition in mammalian cells. *Oncogene* 8, 1593-602.

Uhlmann, F., Gibbs, E., Cai, J., O'Donnell, M., and Hurwitz, J. (1997). Identification of regions within the four small subunits of human replication factor C required for complex formation and DNA replication. *J. Biol. Chem.* 272, 10065-10071.

Valverde, M., Fortoul, T.I., Diaz-Barriga, F., Mejia, J., and del Castillo, E.R. (2000). Induction of genotoxicity by cadmium chloride inhalation in several organs of CD-1 mice. *Mutagen.* 15, 109-114.

Valverde, M., Fortoul, T.I., Diaz-Barriga, F., Mejia, J., and del Castillo, E.R. (2002). Genotoxicity induced in CD-1 mice by inhaled lead: differential organ response. *Mutagen.* 17, 55-61.

Venables, W.N., and Ripley, B.D. (1999). *Modern Applied Statistics with S-Plus*. 3rd Ed. New York:Springer-Verlag.

Walter, C.A., Lu, J., Bhakta, M., Zhou, Z.Q., Thompson, L.H., and McCarrey, J.R. (1994). Testis and somatic *Xrcc-1* DNA repair gene expression. *Somat. Cell. Mol. Genet.* 20, 451-61.

Westfall, P.H., and Young, S.S. (1993). *Resampling-based multiple testing: examples and methods for p-value adjustment*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Yang, Y. H., Dudoit, S., Luu, P., and Speed, T. P. (2001). Normalization for cDNA microarray data. *SPIE Proceed.* 4266, 141-152.

Mitotic and Meiotic Gene Expression Profiling of Male Germ Cells

4.1 Abstract

The purpose of this research was to characterize gene expression profiles for ~10,000 genes as germ cells differentiated from spermatogonia into primary spermatocytes and to identify the genes and biological pathways associated with these periods in spermatogenesis. Expression was characterized in the testes of prepubertal B6C3F1 mice during spermatogonial mitosis (postnatal day 7), at the onset of meiosis (postnatal day 9), and during the mid-pachytene stage of meiosis I (postnatal day 14). Approximately 550 genes were found to be differentially expressed across these time points, including 428 genes that exhibited differential expression between spermatogonial mitosis and the onset of meiosis (383 with relatively higher spermatogonial expression; 45 genes with relatively higher preleptotene expression). An additional 70 genes showed differential expression between early and mid-meiosis I (25 with higher pachytene expression; 35 with higher preleptotene expression). The remaining 46 genes exhibited complex temporal expression patterns (differential expression during the mitosis to meiosis transition and also during meiosis I). These results show that a genome-scale

approach for investigating gene expression during spermatogenesis allows the (a) characterization of temporal expression patterns for genes with known roles in germ cell differentiation, (b) identification of new spermatogenesis-related roles for genes with functions previously characterized in other biological pathways, and (c) discovery of novel pre-meiosis and meiosis-enriched genes. These findings lay the foundation for understanding molecular mechanisms of the mitotic to meiotic transition and meiosis, genetic changes associated with male infertility, and susceptibility factors affecting the transmission of mutations to offspring following paternal exposure to genotoxic agents.

4.2 Introduction

Spermatogenesis can be divided into three phases: proliferative (mitosis), meiotic, and post-meiotic. Both the mitotic and meiotic phases occur within the seminiferous tubules of the testis. Spermiogenesis begins in the seminiferous tubules, and final sperm maturation occurs in the epididymis. The timing of germ cell progression through spermatogenesis has been well characterized in humans and rodents (Ham, 1974; Bellve et al., 1977).

Male germ cell development begins in the fetal mouse at 8-8.5 days post-coitum (d.p.c.) when cells in the primitive ectoderm differentiate to form primordial germ cells (PGCs). These PGCs, or prespermatogonia, divide and migrate to the genital ridge by 13.5 d.p.c where they arrest in G₁ of the mitotic cell cycle until after birth (Hogan et al., 1986). The first wave of spermatogenesis begins in the prepubertal mouse at postnatal (p.n.) day 5-6 with the initiation of spermatogonial stem cell mitosis (Bellve et al., 1977).

Mitotic divisions give rise to other spermatogonial stem cells and type A₁ spermatogonia that will subsequently differentiate and mitotically divide to form A₂-, A₃-, A₄-, Intermediate-, and B-type spermatogonia (Thomas et al., 1989). On p.n. day 9, the first type B spermatogonia divide to form spermatocytes (meiotic cells). Preleptotene spermatocytes undergo the last DNA replication (S phase) and enter the long prophase of meiosis I (Bellve et al., 1977). During prophase I, the leptotene spermatocytes differentiate through zygotene (chromosome synapsis), pachytene (recombination and increased RNA synthesis), and diplotene (synaptonemal complex degeneration and RNA synthesis). This prolonged prophase I lasts until p.n. day 18 when the primary spermatocytes undergo a reductional division to become secondary spermatocytes (Thomas et al. 1989). The first and second meiotic divisions occur in rapid succession, forming the first round spermatids by p.n. day 20 (Bellve et al., 1977). The process of spermiogenesis takes approximately 13.5 days (Gilbert, 1994). It is estimated that a new wave of mouse spermatogenesis is initiated every 8.6 days (Oakberg, 1956); therefore, the second wave of spermatogenesis is initiated at approximately p.n. day 14-15 in the prepubertal mouse.

The histophysiological characteristics of germ cell differentiation in the testis have been well documented (e.g., Bellve et al., 1977; Russell et al., 1990). Germ cells remain connected by intercellular cytoplasmic bridges as they differentiate from the type A₁ spermatogonial stage through the spermatid stage (Gilbert, 1994). This syncytial organization is thought to promote synchronous male germ cell differentiation by allowing ions, molecules and gene products to be shared among cells (Braun et al., 1989; Gilbert, 1994). A number of studies have shown that the expression of one or a small

number of genes across critical phases of germ cell development affects progression through mouse spermatogenesis. For example, Schrans-Stassen et al. showed that *c-kit* expression is a molecular marker for the transition of prespermatogonial cells into differentiating type A spermatogonial cells (1999). Additionally, it has been suggested that the stage-specific expression of certain genes, such as topoisomerase II, may be required for progression through meiosis I (Cobb et al., 1997). Recent advances in expression microarray technology provide an opportunity to develop a more complete understanding of the expression profiles of a large number of genes that contribute to the onset and progression of spermatogenesis.

The objective of this research was to investigate the expression profiles of genes in the prepubertal mouse testis as germ cells differentiate through the mitotic and meiotic compartments during the first wave of spermatogenesis. Random oligonucleotide microarrays containing ~10,000 annotated genes and unannotated sequences were used to identify clusters of genes that were differentially expressed (a) during the transition from spermatogonial mitosis to meiosis and (b) during early and mid-meiosis I (preleptotene vs. pachytene I).

4.3 *Materials and Methods*

Probe generation

B6C3F1 male mice (Harlan Sprague Dawley, Indianapolis, IN) were kept on a 12 hour light/dark cycle and were allowed food and water *ad libitum*. Mice were euthanized by cervical dislocation at each of the following postnatal time points: spermatogonial

mitosis (day 7), onset of meiosis/preleptotene (day 9), mid-meiosis/pachytene (day 14), and adult mice (8 weeks). Immediately following euthanization, testes were extracted from 4 - 6 mice per experimental time point, and microdissected under a dissection microscope to remove the tunica albuginea and isolate the seminiferous tubules. The tubules were stored at -80°C until RNA isolation was performed.

Total RNA was isolated by homogenization (Omni Tissue Homogenizer, Warrenton, VA) in TRIzol Reagent (Life Technologies, Rockville, MD). RNA was ethanol precipitated twice, resuspended in RNase-free water (Sigma, St. Louis, MO), and purified using RNeasy (Qiagen, Valencia, CA). Purified RNA was quantitated using the GeneQuant spectrophotometer (Amersham Pharmacia Biotech, Piscataway, NJ).

First and second strand cDNA synthesis was performed using a starting quantity of 8 μg of total RNA. To synthesize first strand cDNA, RNA was primed with an HPLC purified T7-(dT)₂₄ primer (final concentration: 100 pmol; Integrated DNA Technologies, Coralville, IA) and reverse transcribed at 42°C using Superscript II reverse transcriptase (Life Technologies) in the presence of unlabeled dNTPs (Amersham Pharmacia Biotech). Second strand synthesis was performed using DNA polymerase I (Life Technologies) and DNA ligase (Life Technologies) in the presence of RNase H (Life Technologies) to incorporate unlabeled dNTPs (Amersham Pharmacia Biotech). The synthesized cDNA was purified using the QIAquick Purification Kit (Qiagen). RNA was biotin-labeled during *in vitro* transcription reactions using the BioArray HighYield RNA Transcript Labeling Kit (Enzo Diagnostics, Inc., Farmingdale, NY). The amplified, labeled RNA was then purified using the RNeasy Mini Kit (Qiagen). Prior to hybridization, approximately 20 μg of labeled RNA was fragmented in a buffer comprised of 200 mM

tris-acetate (Sigma), 500 mM potassium acetate (Sigma), and 150 mM magnesium acetate (Sigma).

Hybridization and image capture

Two pools of independently prepared testes were hybridized per experimental group. The fragmented labeled probe was hybridized to Affymetrix Murine Genome U74Av.1 microarrays (Affymetrix, Inc., Santa Clara, CA) according to manufacturer's directions. Briefly, the probe mixture for each hybridization was comprised of 5 µg fragmented probe, Affymetrix eukaryotic hybridization controls (*BioB*: 1.5 pmol; *BioC*: 5 pmol; *BioD*: 25 pmol; *cre*: 100 pmol), 10 µg herring sperm DNA (Life Technologies), 50 µg acetylated bovine serum albumin (Life Technologies), and the Affymetrix hybridization buffer (100 mM MES/1M sodium salt/20 mM EDTA/0.01% Tween 20). Hybridizations were performed for 16 hours at 45°C in an Affymetrix hybridization oven set at 60 rpm. The hybridized arrays were then washed and stained with streptavidin phycoerythrin (SAPE) on the Affymetrix Fluidics Station according to manufacturer's directions (Affymetrix, Inc.). Images were captured using the Agilent Gene Array laser scanner (Agilent Technologies, Palo Alto, CA).

Data normalization and analysis

To normalize the array data, a quantile normalization scheme was applied at the probe level (Irizarry et al., 2002) using an R statistical package that is publicly available at: <http://biosun01.biostat.jhsph.edu/~ririzarr/Raffy>. Following normalization, all genes with MicroArray Suite 5.0 (Affymetrix, Inc.) detection p-values ≤ 0.005 (in at least one

hybridization) were selected for pairwise comparisons of gene expression between time points. Using the replicate \log_2 intensity values obtained for each gene, a two-sample student's t distribution was performed, assuming unequal variance (Dowdy and Wearden, 1991), for each of the following comparisons: spermatogonial mitosis vs. onset of meiosis, onset of meiosis vs. mid-pachytene of MI, and mid-pachytene of MI vs. adult. Genes with t-test p-values ≤ 0.05 and expression ratios ≥ 1.8 were considered to be differentially expressed between time points and were further characterized according to their function in a given biological pathway.

Cluster analysis of expression

Expression trends were characterized for all genes with a MicroArray Suite 5.0 detection p-value ≤ 0.005 using the unsupervised hierarchical cluster analysis program CLUSFAVOR (CLUSter and Factor Analysis Using Varimax Orthogonal Rotation; Peterson, 2002). Briefly, \log_{10} gene expression intensities were clustered using the centroid average Euclidean distance between joining nodes. The resulting dendograms identified groups of genes that were differentially expressed between specific time points during the first wave of spermatogenesis in the prepubertal mouse and full spermatogenesis in the adult mouse.

4.4 Results

Temporal Gene Expression During the First Wave of Mouse Spermatogenesis

As shown in Figure 4.1, gene expression profiles were investigated in the seminiferous tubules obtained from prepubertal mice euthanized at times that correspond to when the most advanced cell types of the first wave of spermatogenesis were at the following: spermatogonial mitosis (p.n. day 7), onset of meiosis (preleptotene; p.n. day 9), and mid-pachytene stage of meiosis I (MI; p.n. day 14). Of approximately 10,000 genes evaluated on the Affymetrix Murine U74Av.1 microarray, 3986 genes had significant signal intensities (Affymetrix MAS 5.0 detection p-value ≤ 0.005) at one or more experimental time points, indicating that they were expressed. Of these, 544 genes (231 annotated genes and 313 unannotated sequences) exhibited significant temporal changes in expression (Student's t-test p-value ≤ 0.05) in the prepubertal mouse testis, as further described below. Among the 544 differentially expressed genes, 428 were differentially expressed between spermatogonial mitosis and preleptotene (but were not differentially expressed between preleptotene and pachytene), 70 were differentially expressed between preleptotene and pachytene (but were not differentially expressed between spermatogonial mitosis and preleptotene), and 46 exhibited complex patterns of expression (i.e., differential expression between spermatogonial mitosis and preleptotene and also between preleptotene and pachytene). The temporally expressed annotated genes have a variety of biological functions, as shown in Tables 4.1 - 4.3 and Appendix B. Therefore, each gene was also classified according to whether or not it had a previously described role in germ cell development or differentiation, meiosis, spermatogenesis, and/or fertilization.

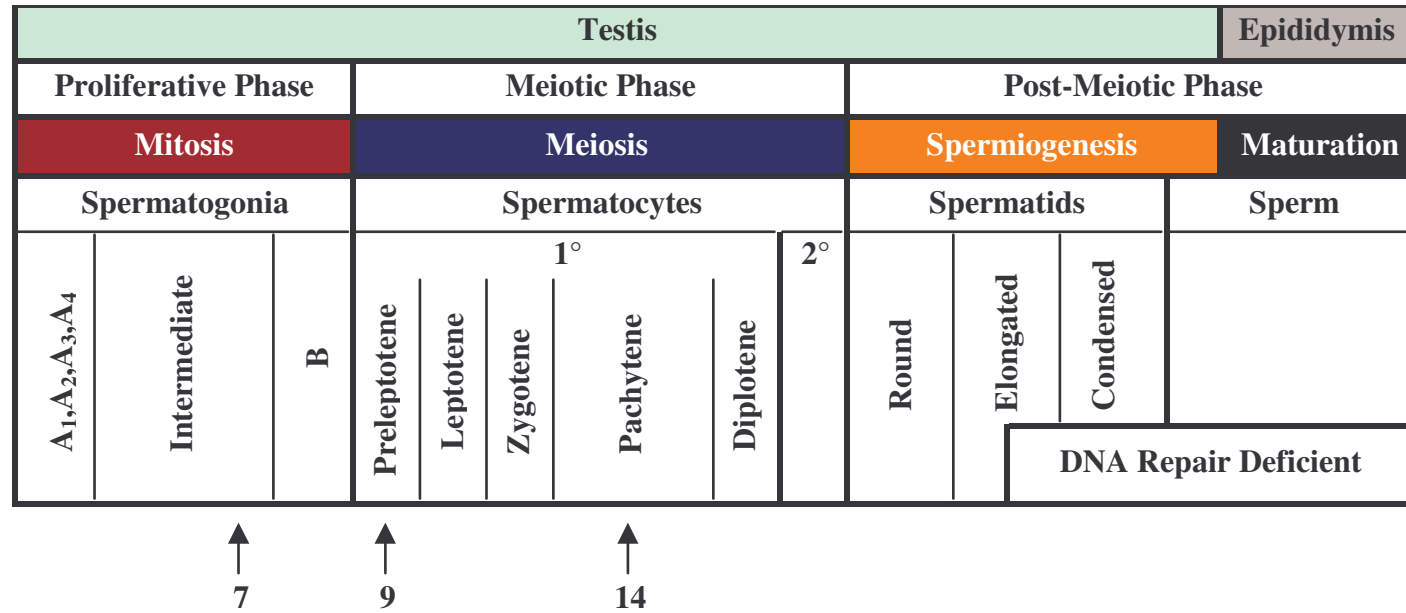


Figure 4.1 Germ cell progression through mouse spermatogenesis. Male germ cells differentiate through three phases: proliferative, meiotic, and post-meiotic (left to right). The proliferative phase includes the mitotic divisions of spermatogonial cells. The meiotic phase includes the first and second divisions of primary and secondary spermatocytes, respectively. The post-meiotic phase includes the morphological transformation of the round haploid spermatid into testicular sperm, and biochemical maturation of sperm which occurs in the epididymis.

The genes that were differentially expressed between spermatogonial mitosis and the onset of meiosis may have important functions during the transition from mitosis to meiosis. As shown in Figure 4.2, 187 of the 428 genes were annotated while the remaining 241 were unannotated sequences. Among the 187 annotated genes, 14 exhibited higher expression at the onset of meiosis, and the remaining 173 genes had higher expression during spermatogonial mitosis. The expression data for these 187 annotated genes is listed in Appendix B. Specific roles in spermatogenesis were previously reported for only 37 (~20%) of the annotated genes (Appendix B). Examples of the annotated genes that exhibited relatively higher expression levels during the onset of meiosis are listed in Table 4.1. These genes may be important for the transition into the meiotic phase of spermatogenesis. For example, *Zfx* which is important for cellular differentiation during spermatogenesis is expressed ~2.7-fold higher in preleptotene cells than in the mitotic cells ($p = 0.05$) and this higher level of expression is still observed in pachytene cells ($p = 0.2$; data not shown). Examples of genes that exhibited relatively higher expression during spermatogonial mitosis are also listed in Table 4.1. These genes may have important roles in spermatogonial mitosis; however, it is also possible that their decreased expression is necessary for the transition into meiosis. For example, the cell cycle-related gene *Plkl* had higher expression during mitosis than in preleptotene cells (~2.6-fold increase; $p = 0.03$) and the lower level of expression in preleptotene cells was still observed during pachytene I ($p = 0.60$; data not shown). The 241 differentially expressed unannotated sequences (31 with higher preleptotene expression; 210 with higher mitotic expression), which may represent novel genes with roles in the mitosis to meiosis transition, are listed in Appendix C.

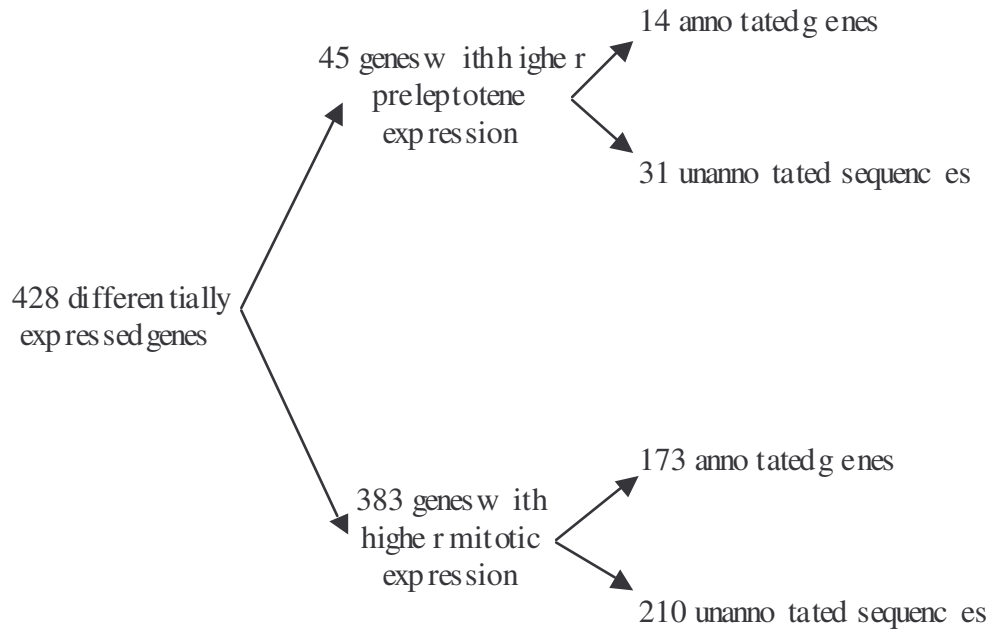


Figure 4.2. Genes that were differentially expressed between mitosis and the onset of meiosis (but not between early and mid-meiosis).

Table 4.1. Examples of genes with differential expression between mitosis and the onset of meiosis.

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b	
				Ratio	P-value ^a		
Genes with increased expression as germ cells progress from mitosis to the onset of meiosis							
Ig germline H-chain gene V-region	<i>IgH (germline)</i>	Immune response	M16724	6.8	0.001	Yes	(Kerr and Burrows, 1991)
Zinc finger protein X-linked	<i>Zfx</i>	Cellular differentiation	M32309	2.7	0.05	Yes	(Luoh et al., 1997)
DNA methyltransferase	<i>Dnmt1</i>	Transcription	AF036008	1.7	0.01	Yes	(Mertineit et al., 1998)
Methylmalonyl coenzyme A mutase	<i>Mut</i>	Metabolism	X51941	11.9	0.04	No	
Sprouty-4	<i>Spry4</i>	Signal transduction	AB019280	5.6	0.02	No	
Transactivating transcription factor 3	<i>Sp3</i>	Transcription	AF062567	3.8	0.01	No	
ADP-ribosyltransferase 5	<i>Art5</i>	Protein modification	U60881	2.3	0.03	No	
Cryptochrome 1	<i>Cry1</i>	Molecular clock	AB000777	2.1	0.01	No	
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis							
Glutathione peroxidase 4	<i>Gpx4</i>	Stress response	D87896	44.1	0.03	Yes	(Giannattasio et al., 1997)
DEAD box polypeptide, Y chromosome	<i>Dby</i>	RNA helicase	AJ007376	11.4	0.03	Yes	(Mazeyrat et al., 2001)
Flap structure specific endonuclease 1	<i>Fen1</i>	DNA repair	L26320	7.0	0.04	Yes	(Harrington and Lieber, 1994)
Bcl2-associated X protein	<i>Bax</i>	Apoptosis	L22472	6.4	0.01	Yes	(Russell et al., 2002)
Breast cancer 2	<i>Brca2</i>	DNA repair	U89652	3.5	0.02	Yes	(Connor et al., 1997)
Polo-like kinase 1	<i>Plk1</i>	Cell cycle	U73170	2.6	0.03	Yes	(Matsubara et al., 1995)
RNA binding motif 3	<i>Rbm3</i>	Stress response	AB016424	1.9	0.04	Yes	(Danno et al., 2000)
Myeloid cell leukemia sequence 1	<i>Mcl1</i>	Apoptosis	U35623	14.9	0.02	No	
Thioredoxin	<i>Txn1</i>	Stress response	X77585	10.9	0.001	No	
Cyclin F	<i>Ccnf</i>	Cell cycle	Z47766	9.0	0.04	No	
Programmed cell death 6	<i>Pdcd6</i>	Apoptosis	U49112	4.4	0.01	No	
DNA polymerase zeta catalytic subunit	<i>Rev3l</i>	DNA repair	AF083464	2.8	0.02	No	

^aStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points (see Materials and Methods)

^bGenes with described roles in germ cell development or differentiation, meiosis, spermatogenesis, or fertilization

As shown in Figure 4.3, there were 70 genes that were differentially expressed between preleptotene and pachytene of meiosis I (25 with higher pachytene expression; 35 with higher preleptotene expression). Only 25 of the 70 genes were annotated, including 10 with known roles related to spermatogenesis (Table 4.2). Nine of the 25 annotated genes showed higher expression during pachytene I (e.g., the recombination-associated gene *Sycp1* was ~ 5.7-fold higher in pachytene compared to preleptotene; $p = 0.02$). The remaining 16 annotated genes exhibited higher expression during preleptotene (e.g., *Zfp49* had ~3.8-fold higher expression in preleptotene compared to pachytene; $p = 0.01$). The expression of these genes may be important for spermatogonial mitosis and the onset of meiosis and/or their decreased expression may be necessary for progression through meiosis I. The 45 unannotated sequences that were also differentially expressed during meiosis I (16 with higher pachytene expression; 29 with higher preleptotene expression; Figure 4.3) are listed in Appendix C.

Forty-six genes (19 annotated and 27 unannotated) showed complex patterns of expression. They were differentially expressed between spermatogonial mitosis and the onset of meiosis and also between the onset of meiosis and mid-meiosis I (Figure 4.4). As shown in Figure 4.4 and Table 4.3, three temporal patterns of expression were observed: (1) increased expression from mitosis to preleptotene and decreased expression from preleptotene to pachytene (1 annotated gene; 2 unannotated sequences); (2) decreased expression from mitosis to preleptotene and increased expression from preleptotene to pachytene (10 annotated genes; 15 unannotated sequences); and (c) continual decrease in expression from mitosis to preleptotene and from preleptotene to

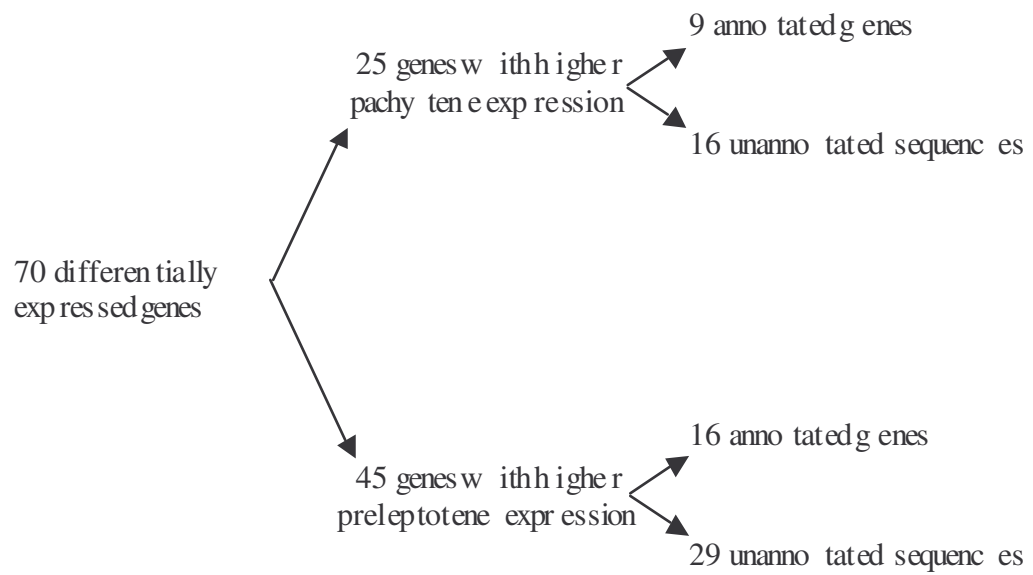


Figure 4.3. Genes that were differentially expressed between the onset of meiosis and mid-meiosis I (but not between mitosis and preleptotene).

Table 4.2. Genes with significant differential expression between the onset and middle of meiosis I.

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with increased expression as germ cells progress from the onset of meiosis to mid-meiosis I						
Hematopoietic cell transcript	<i>HemT</i>	Cellular differentiation	AJ242830	78.1	0.02	Yes (Xue et al., 1999)
Lipase, hormone sensitive	<i>Lipe</i>	Metabolism	U69543	11.2	0.02	Yes (Chung et al., 2001)
Lactate dehydrogenase 3	<i>Ldh3</i>	Metabolism	M17587	9.2	0.04	Yes (Kao et al., 1988)
Synaptonemal complex protein 1	<i>Sycp1</i>	Recombination	D88539	5.7	0.02	Yes (Tureci et al., 1998)
Acid phosphatase 2, lysosomal	<i>Acp2</i>	Signal transduction	X57199	8.9	0.02	No
Ganglioside-induced differentiation 3	<i>Gdap3</i>	Unknown	Y17852	3.9	0.04	No
Protein kinase inhibitor p58	<i>Ipk</i>	Protein modification	U28423	3.2	0.05	No
Glutamyl-tRNA synthetase	<i>Glut-tRNA synthetase</i>	Biosynthesis	X54327	2.4	0.01	No
House-keeping protein 1	<i>Hkp1</i>	Misc.	M74555	2.2	0.02	No
Genes with decreased expression as germ cells progress from the onset of meiosis to mid-meiosis I						
Hydroxysteroid (17-beta) dehydrogenase 3	<i>Hsd17b3</i>	Metabolism	U66827	6.6	0.01	Yes (Sha et al., 1996)
Zinc finger protein 35	<i>Zfp35</i>	Cellular differentiation	M36146	3.8	0.01	Yes (Cunliffe et al., 1990)
Integrin alpha 6	<i>Iga6</i>	Cell adhesion	X69902	3.4	0.005	Yes (Husen et al., 1999)
Glutathione-S-transferase, alpha 1	<i>Gsta1</i>	Stress response	L06047	2.9	0.01	Yes (Benbrahim-Tallaa et al., 2002)
Diazepam-binding inhibitor	<i>Dbi</i>	Metabolism	X61431	2.7	0.01	Yes (Kolmer et al., 1997)
Zinc finger protein 49	<i>Zfp49</i>	Cellular differentiation	AB013357	2.6	0.02	Yes (Cunliffe et al., 1990)
GA repeat binding protein, beta 1	<i>Gabpb1</i>	Transcription	M74516	7.0	0.03	No
Tissue factor pathway inhibitor	<i>Tfpi</i>	Misc.	AF004833	6.3	0.03	No
N-myristoyltransferase 2	<i>Nmt2</i>	Biosynthesis	AF043327	5.1	0.04	No
Monocyte macrophage 19	<i>Mmrp19</i>	Immune response	AB028863	4.9	0.001	No
Smoothed homolog	<i>Smoh</i>	Signal transduction	AF089721	4.1	0.01	No
Prolyl oligopeptidase	<i>Prep</i>	Protein modification	AB022053	3.6	0.001	No
Lysosomal-associated, transmembrane 4	<i>Laptm4</i>	Intercellular transport	U34259	3.1	0.05	No
Fas-associated death domain	<i>Fadd</i>	Apoptosis	U50406	2.8	0.03	No
Annexin III	<i>Anx3</i>	Metabolism	AJ001633	2.4	0.004	No
Ryanodine receptor 3	<i>Ryr3</i>	Signal transduction	D38218	2.0	0.04	No

^aStudent's t-test p-values ≤ 0.05 indicate significant differences in expression across time points (see Materials and Methods)

^bGenes with described roles in germ cell development or differentiation, meiosis, spermatogenesis, or fertilization

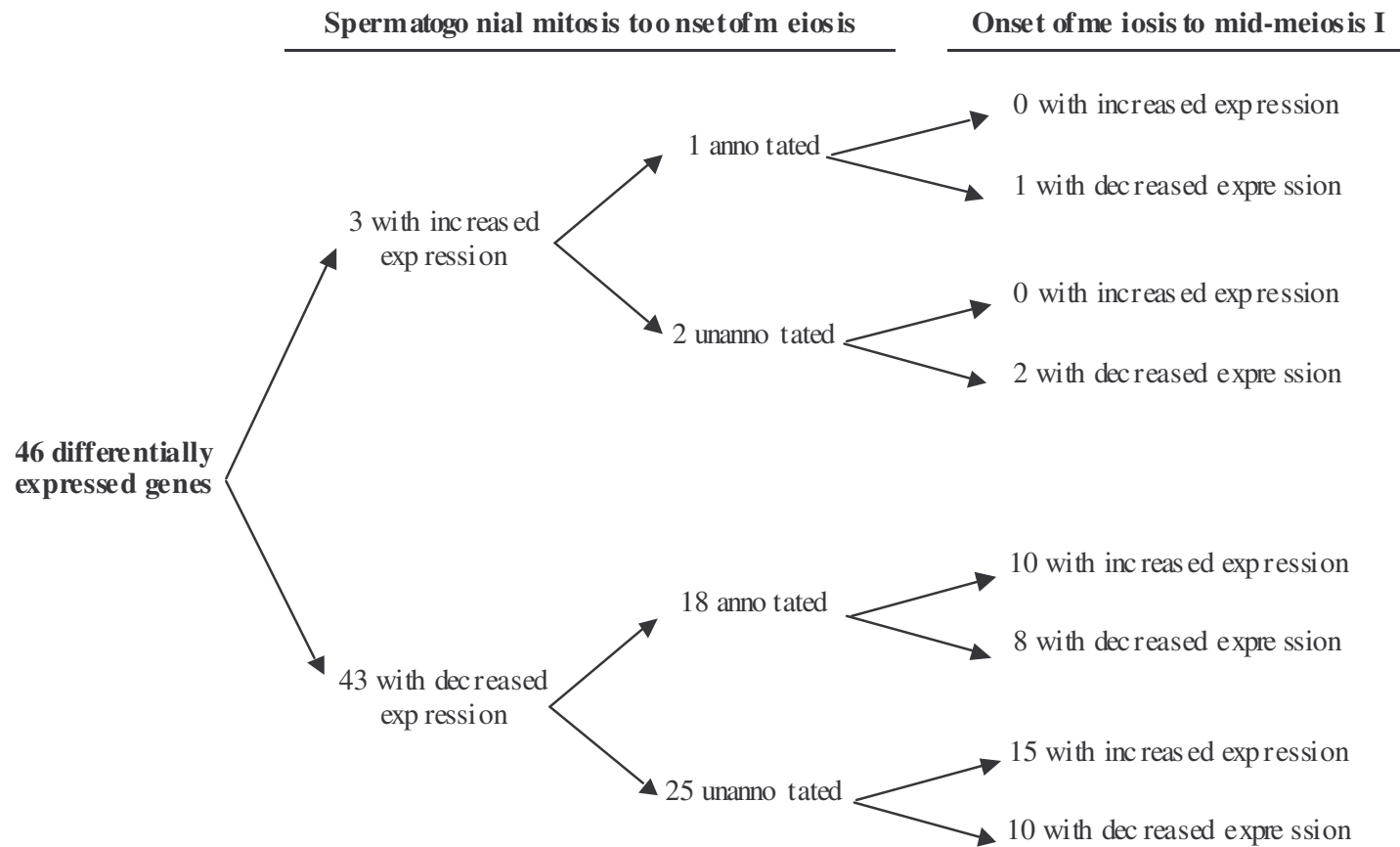


Figure 4.4. Genes with complex patterns of differential expression from mitosis to mid-meiosis I. These genes show expression changes both between mitosis and the onset of meiosis and between the onset of meiosis and mid-meiosis I.

Table 4.3. Genes with complex patterns of differential expression from mitosis to mid-meiosis I.

Name	Symbol	Function	GenBank Accession ID	Relative Expression Pattern ^a	Preleptotene / Spermatogonial Mitosis		Pachytene / Preleptotene		Reference
					Ratio ^b	P-value ^c	Ratio ^b	P-value ^c	
Genes with described roles in germ cell development or differentiation, meiosis, spermatogenesis, or fertilization									
Degenerative spermatocyte homolog	<i>Mdes</i>	Meiosis	Y08460	DD	-7.5	0.01	-2.1	0.02	(Endo et al., 1997)
Growth arrest specific 6	<i>Gas6</i>	Growth regulation	X59846	DD	-7.3	0.04	-3.6	0.02	(Chan et al., 2000)
Ornithine decarboxylase antienzyme 2	<i>Oaz2</i>	Misc.	D78643	DI	-23.4	0.01	2.9	0.03	(Ivanov et al., 2000)
Secreted protein acidic and rich in cysteine	<i>Sparc</i>	Extracellular matrix	X04017	DI	-4.6	0.02	1.9	0.02	(Vernon and Sage, 1989)
Outer dense fiber 1	<i>Odf1</i>	Cytoskeletal element	X79446	ID	4.8	0.01	2.3	0.04	(Burmester et al., 1996)
Genes with no known role in germ cell development or differentiation, meiosis, spermatogenesis, or fertilization									
Cardiac troponin T isoform A3b	<i>CtnTA3b</i>	Metabolism	L47600	DD	-11.6	0.02	-2.3	0.04	
DEAD box polypeptide 4	<i>Dbp4</i>	RNA helicase	D14859	DD	-2.4	0.02	-1.8	0.04	
Glutathione transferase mu 3	<i>Gstm3</i>	Stress response	J03953	DD	-2.1	0.02	-2.3	0.02	
Inosine monophosphate dehydrogenase	<i>Impdh</i>	Biosynthesis	M33934	DD	-5.1	0.01	-2.6	0.02	
Suppressor of cytokine signaling 2	<i>Socs2</i>	Growth regulation	U88327	DD	-4.5	0.01	-5.9	0.04	
Transcytosis associated protein p115	<i>Tap</i>	Intracellular transport	AF096868	DD	-2.6	0.01	-1.8	0.01	
Adenylosuccinate synthetase	<i>Adss</i>	Biosynthesis	L24554	DI	-5.6	0.004	1.9	0.05	
Cop9 complex subunit 6	<i>Cops6</i>	Protein modification	AF071315	DI	-32.8	0.01	3.3	0.01	
Fibroblast secretory protein	<i>Fsp</i>	Growth regulation	M70642	DI	-12.2	0.02	1.6	0.01	
Fibulin extracellular matrix glycoprotein	<i>Fbln1</i>	Cell adhesion	X70853	DI	-6.9	0.001	5.1	0.002	
Peroxisomal integral membrane protein 34	<i>Pmp34</i>	Intracellular transport	AJ006341	DI	-3.8	0.04	3.7	0.01	
Spastin	<i>Spast</i>	Cytoskeletal element	AJ246002	DI	-7.4	0.03	5.2	0.04	
TEA domain family member 4	<i>Tead4</i>	Transcription	X94441	DI	-2.7	0.03	3.8	0.05	
Translocase of inner membrane 23	<i>Tim23</i>	Intracellular transport	AB021122	DI	-4.7	0.01	2.5	0.05	

^aDD = Decreased expression from spermatogonial mitosis to preleptotene and decreased expression from preleptotene to pachytene; DI = Decreased expression from spermatogonial mitosis to preleptotene and increased expression from preleptotene to pachytene; ID = Increased expression from spermatogonial mitosis to preleptotene and decreased expression from preleptotene to pachytene

^bFold difference in expression for preleptotene vs. spermatogonial mitosis or pachytene vs. preleptotene; Negative numbers indicate decreased expression as germ cells differentiate

^cStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points (see Materials and Methods)

pachytene (8 annotated genes; 10 unannotated sequences). Only 5 of the 19 annotated genes have described roles related to germ cell development or progression through spermatogenesis. The 27 unannotated sequences that have significant expression differences across all time points evaluated and, therefore, may be important for progression through the first wave of spermatogenesis are listed in Appendix C.

Cluster Analysis of Temporal Gene Expression

Unsupervised hierarchical cluster analysis was performed to identify clusters of genes with similar patterns of expression (CLUSFAVOR; Peterson 2002). These similar expression patterns may indicate that these genes belong to common molecular/biochemical pathways. All genes with significant differential expression were analyzed by cluster analysis. Expression data for the adult mouse was also included in order to compare the onset and progression of meiosis in the prepubertal mouse with expression during full spermatogenesis. Cluster analysis revealed the presence of 44 major (≥ 4 genes) and 83 minor (< 4 genes) sub-clusters. Examples of the major sub-clusters, which contain both annotated genes and unannotated sequences, are shown in Figure 4.5. Clusters were identified that contain genes which are highly expressed during spermatogonial mitosis (Figure 4.5 A). It also identified groups of unannotated sequences that exhibit the highest or lowest expression levels at the onset of meiosis (Figure 4.5 B and C, respectively). Groups of known genes and unannotated sequences that have increased (Figure 4.5 D) or decreased (Figure 4.5 E) expression across meiosis I were also observed.

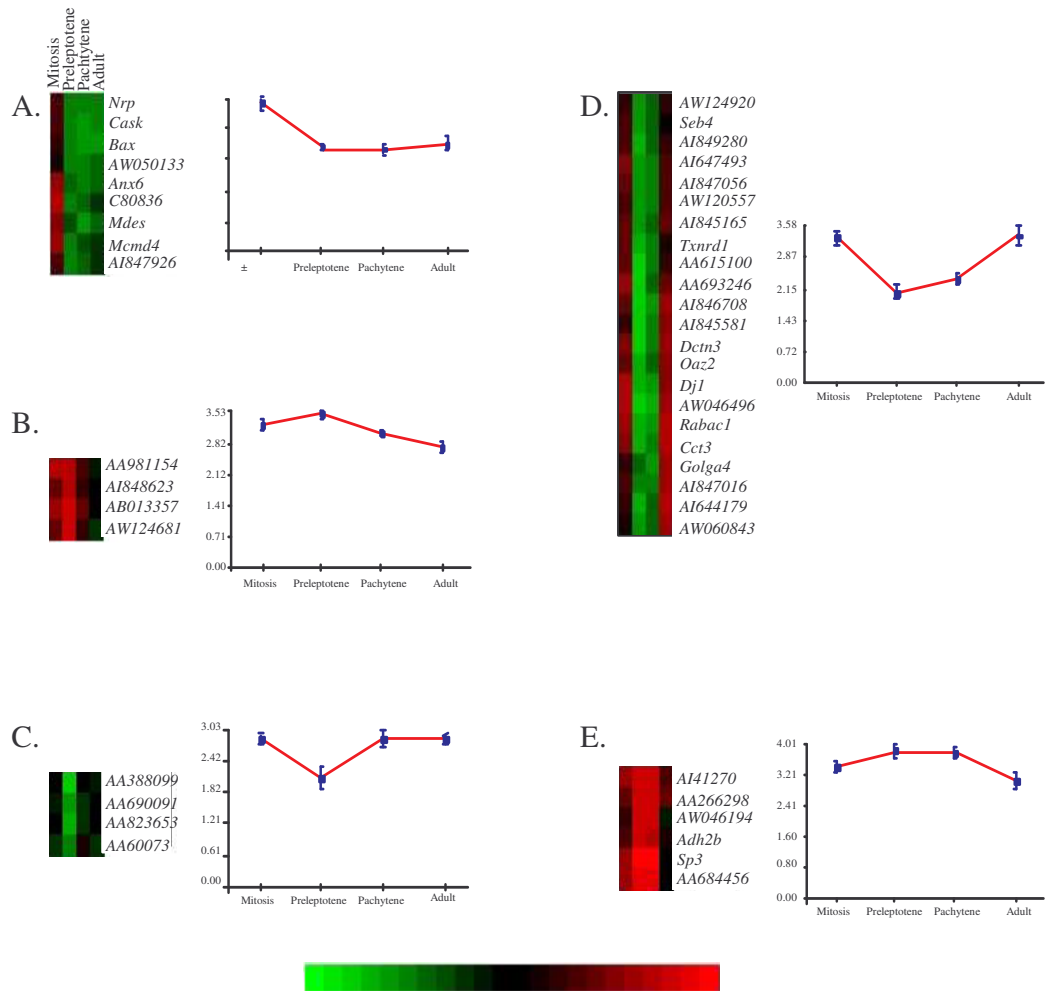


Figure 4.5. Examples of gene sub-clusters obtained using CLUSFAVOR. Each colt in the sub-cluster represents one time point (from left to right: spermatogonial mitosis, preleptotene, pachytene I, and adult). Each row represents one gene. Graphs corresponding to each sub-cluster illustrate the mean log expression ratio for that sub-cluster (\pm the standard deviation for that sub-cluster) at each time point. **A.**) Genes exhibiting relatively elevated expression during spermatogonial mitosis. **B.**) Genes with slightly elevated expression in preleptotene. **C.**) Genes with decreased expression during preleptotene. **D.**) Genes that have decreased expression during both meiosis I samplings (preleptotene and pachytene). **E.**) Genes showing relative increased expression during both meiosis I samplings.

4.5 *Discussion*

Spermatogenesis is a complex and highly synchronized differentiation process. During spermatogenesis, the expression of genes that encode proteins with key roles during specific periods of germ cell development is strictly regulated (Hecht, 1998; Grootegoed et al., 2000). Temporal changes in gene expression as specific germ cell types appear during spermatogenesis have been reported for a small number of genes (e.g., Willison and Ashworth 1987; Thomas et al., 1989; Hecht 1998). However, a global analysis of gene expression across critical timepoints during spermatogenesis (e.g., transition from mitosis to meiosis) is necessary to thoroughly understand germ cell progression through spermatogenesis and to identify genes that play critical roles in controlling the development and formation of mature germ cells.

In the present study, random oligonucleotide microarrays were used to characterize temporal changes in gene expression in the prepubertal mouse testis as germ cells differentiated from spermatogonia to pachytene spermatocytes. Differential gene expression was observed between every stage analyzed in this study. This finding supports the suggestion that all stages of spermatogenesis are characterized by selective gene expression (Thomas et al. 1989). Out of ~10,000 genes investigated, we identified 544 genes (231 annotated genes and 313 unannotated sequences) that were differentially expressed during the transition from mitosis to meiosis and/or during meiosis I. Among the 231 differentially expressed annotated genes, only 53 were known to have functions related to spermatogenesis. Therefore, the remaining 178 annotated genes and 313 unannotated sequences represent candidate genes with novel roles in spermatogenesis.

Differential gene expression between spermatogonial mitosis and the onset of meiosis

Approximately 79% of the genes (428 out of 544) showed differential expression during the mitotic to meiotic transition. As shown in Figure 4.2, the majority of these genes showed a decrease in expression as spermatogonia committed to meiosis. Among the annotated genes with decreased expression were several apoptosis-related genes such as *Bax*, *Mcl1* and *Pdcd6* (Table 4.1). Apoptosis plays a key role during spermatogenesis to regulate sperm production. During the first wave of spermatogenesis in particular, an early wave of apoptosis, which is necessary for the development of functional spermatogenesis, is coincident with a temporarily high expression of *Bax* (Rodriguez et al 1997). Recent data indicate that *Bax* promotes cell death specifically in type A spermatogonia (Russell et al 2002). Therefore, the reduced expression of *Bax* as spermatogonia move to meiosis is in agreement with the known involvement of *Bax* in regulating sperm production.

Several genes associated with DNA repair and stress response were also shown to decrease in expression as cells proceeded from mitosis to meiosis (Table 4.1). Differential expression of DNA repair and stress response genes among the various spermatogenic phases has been reported in the rat, and it has been suggested that such differential expression contributes to the selective susceptibilities of germ cells to stress (Aguilar-Mahecha et al., 2001). The present results show that a similar trend is present in the mouse.

Only 45 of the 428 differentially expressed annotated genes and unannotated sequences genes showed higher expression in meiotic cells with respect to mitotic cells. Their higher expression suggest an important role in meiosis for these genes. The limited

number of genes that was detected as having higher expression in preleptotene cells is probably related to the fact that the Affymetrix Murine U74Av.1 microarray used in this study is highly enriched for genes with functions in somatic tissues.

Differential gene expression between the onset of meiosis and mid-meiosis I

As shown in Figure 4.3, there was also a trend toward down-regulated gene expression as cells proceeded through meiosis. Of the 70 genes that were differentially expressed during meiosis I, approximately 2/3 had relatively higher expression at the onset of meiosis compared to mid-meiosis. Among the 25 annotated genes and unannotated sequences with higher expression during pachytene I, we identified *Sycp1* which exhibited a 5.7-fold increase in expression. *Sycp1* is an integral component of the synaptonemal complex which is assembled during pachytene I (Tureci et al 1998).

Complex temporal patterns of gene expression

Forty-six genes had a complex pattern of expression across the periods of spermatogenesis evaluated in this study (Table 4.3). Further analysis of the temporal patterns of expression for these genes may provide clues to their function. For example, genes with increased expression as spermatogonia differentiate into preleptotene spermatocytes and decreased expression as the preleptotene spermatocytes differentiate into pachytene I spermatocytes, (i.e., *Odf* increased 4.8-fold and subsequently decreased ~2.3-fold), may have roles in the onset of meiosis. Genes with decreased expression from mitosis to preleptotene followed by increased expression from preleptotene to pachytene, (e.g., *Spast* decreased ~7.1-fold and then increased ~5.2-fold), may represent

genes whose repression is necessary for the transition from mitosis to meiosis. Genes whose expression decreased between spermatogonial mitosis and preleptotene and then decreased further from preleptotene to pachytene, (e.g., *Mdes* expression decreased by ~7.7-fold from spermatogonial mitosis to preleptotene and decreased ~2.1-fold from preleptotene to pachytene I), may be mitosis-associated genes.

Cluster analysis of gene expression

Cluster analysis has found widespread use in functional genomics studies for classifying genes based on gene expression levels from cDNA microarrays. Several clustering methods have been developed for analyzing gene expression data (Eisen et al 1998; Sneath and Sokol, 1973; Peterson, 2002). We used hierarchical cluster analysis to identify groups of genes (sub-clusters) that shared identical expression patterns across the various spermatogenic periods (Figure 4.5). A total of 44 major (≥ 4 genes) and 83 minor (< 4 genes) sub-clusters were identified. Genes within each group may have similar cellular functions. For example, four of the known genes in sub-cluster A in Figure 4.5 (*Nrp*, *Cask*, *Anx6* and *Mdes*) encode for plasma membrane-associated proteins suggesting that they may all be part of the cellular modification processes associated with spermatogonial proliferation and differentiation. Genes belonging to the same cluster can also be used to search for common regulatory elements that may be responsible for their coordinated expression (Werner, 2001). This approach has been successfully used in yeast to identify transcription factors involved in cell cycle control (Jelinsky et al 2000; Futcher, 2000).

Summary

By characterizing the temporal coordination of gene expression in the prepubertal testis as germ cells differentiate from spermatogonia to pachytene spermatocytes, this research has identified ~550 genes whose function may be associated with male germ cell differentiation. Characterizing the function of these genes will greatly expand our understanding of the molecular mechanisms associated with the onset and progression of spermatogenesis. The prepubertal mouse model utilized in this study provides a unique method for future investigations of (a) the molecular basis of male infertility (e.g., meiotic arrest in mutant mice) and (b) the genetic causes of adverse reproductive outcomes following paternal exposure to various genotoxic agents (e.g., acrylamide, ENU, ionizing radiation, etc.).

4.6 Acknowledgements

Special thanks to Dorreyah Schahin-Reed for her assistance with the testis extractions and Shea Gardner for normalizing the Affymetrix data. Work was performed under the auspices of the US DOE by the University of California, LLNL under contract W-7405-ENG-48 with support from NIH ES09117-02; NIEHS Superfund P42ES04705; the Low Dose Radiation Research Program, Biological and Environmental Research (BER), U.S. DOE grant KP110202; the University of California and West Virginia University.

4.7 References

Adamali, H. I., Somani, I. H., Huang, J. Q., Gravel, R. A., Trasler, J. M., and Hermo, L. (1999a). II. Characterization and development of the regional- and cellular- specific abnormalities in the epididymis of mice with beta- hexosaminidase A deficiency. *J. Androl.* *20*, 803-24.

Adamali, H. I., Somani, I. H., Huang, J. Q., Mahuran, D., Gravel, R. A., Trasler, J. M., and Hermo, L. (1999b). I. Abnormalities in cells of the testis, efferent ducts, and epididymis in juvenile and adult mice with beta-hexosaminidase A and B deficiency. *J. Androl.* *20*, 779-802.

Aguilar-Mahecha, A., Hales, B. F., and Robaire, B. (2001). Expression of stress response genes in germ cells during spermatogenesis. *Biol. Reprod.* *65*, 119-127.

Bellve, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M., and Dym, M. (1977). Spermatogenic cells of the prepuberal mouse: isolation and morphological characterization. *J. Cell Biol.* *74*, 68-85.

Benbrahim-Tallaa, L., Tabone, E., Tosser-Klopp, G., Hatey, F., and Benahmed, M. (2002). Glutathione S-transferase alpha expressed in porcine Sertoli cells is under the control of follicle-stimulating hormone and testosterone. *Biol. Reprod.* *66*, 1734-42.

Bernard, P., Maure, J. F., and Javerzat, J. P. (2001). Fission yeast Bub1 is essential in setting up the meiotic pattern of chromosome segregation. *Nat. Cell. Biol.* 3, 522-6.

Bo, J. and Wensink, P. C. (1989). The promoter region of the *Drosophila* alpha 2-tubulin gene directs testicular and neural specific expression. *Development* 106, 581-7.

Braun, R.E., Behringer, R.R., Peschon, J.J., Brinster, R.L., and Palmiter, R.D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature* 337, 373-6.

Broceno, C., Ruiz, P., Reina, M., Vilaro, S., and Pons, G. (1995). The muscle-specific phosphoglycerate mutase gene is specifically expressed in testis during spermatogenesis. *Eur. J. Biochem.* 227, 629-35.

Bucci, L. R., Brock, W. A., Goldknopf, I. L., and Meistrich, M. L. (1984). Characterization of high mobility group protein levels during spermatogenesis in the rat. *J. Biol. Chem.* 259, 8840-6.

Burmester, S. and Hoyer-Fender, S. (1996). Transcription and translation of the outer dense fiber gene (*Odf1*) during spermiogenesis in the rat. A study by in situ analyses and polysome fractionation. *Mol. Reprod. Dev.* 45, 10-20.

Cai, L., Chen, S., Evans, T., Deng, D. X., Mukherjee, K., and Chakrabarti, S. (2000). Apoptotic germ-cell death and testicular damage in experimental diabetes: prevention by endothelin antagonism. *Urol. Res.* 28, 342-7.

Castellano, L. E., Martinez-Cadena, G., Lopez-Godinez, J., Obregon, A., and Garcia-Soto, J. (1997). Subcellular localization of the GTP-binding protein Rho in the sea urchin sperm. *Eur. J. Cell Biol.* 74, 329-35.

Chan, M. C., Mather, J. P., McCray, G. and Lee, W. M. (2000). Identification and regulation of receptor tyrosine kinases Rse and Mer and their ligand Gas6 in testicular somatic cells. *J. Androl.* 21, 291-302.

Chung, S., Wang, S. P., Pan, L., Mitchell, G., Trasler, J., and Hermo, L. (2001). Infertility and testicular defects in hormone-sensitive lipase-deficient mice. *Endocrinology* 142, 4272-81.

Cobb, J., Reddy, R.K., Park, C., and Handel, M.A. (1997). Analysis of expression and function of topoisomerase I and II during meiosis in male mice. *Mol. Reprod. Dev.* 46, 489-98.

Connor, F., Bertwistle, D., Mee, P. J., Ross, G. M., Swift, S., Grigorieva, E., Tybulewicz, V. L., and Ashworth, A. (1997). Tumorigenesis and a DNA repair defect in mice with a truncating Brca2 mutation. *Nat. Genet.* 17, 423-30.

Cunliffe, V., Koopman, P., McLaren, A., and Trowsdale, J. (1990). A mouse zinc finger gene which is transiently expressed during spermatogenesis. *EMBO J.* *9*, 197-205.

Cunningham, D.B., Segretain, D., Arnaud, D., Rogner, U.C., and Avner, P. (1998). The mouse *Tsx* gene is expressed in Sertoli cells of the adult testis and transiently in premeiotic germ cells during puberty. *Dev. Biol.* *15*, 345-60.

Danno, S., Itoh, K., Matsuda, T., and Fujita, J. (2000). Decreased expression of mouse *Rbm3*, a cold-shock protein, in Sertoli cells of cryptorchid testis. *Am. J. Pathol.* *156*, 1685-92.

Desseyn, J. L., Burton, K. A., and McKnight, G. S. (2000). Expression of a nonmyristylated variant of the catalytic subunit of protein kinase A during male germ-cell development. *Proc. Natl. Acad. Sci. USA* *97*, 6433-8.

Di Agostino, S., Rossi, P., Geremia, R., and Sette, C. (2002). The MAPK pathway triggers activation of *Nek2* during chromosome condensation in mouse spermatocytes. *Development* *129*, 1715-27.

Dowdy, S. and Wearden, S. (1991). Student's *t* distribution. *Statistics for Research*. 2nd Ed. New York: Wiley-Interscience. pp. 193-228.

Edelmann, L., Stankiewicz, P., Spiteri, E., Pandita, R. K., Shaffer, L., Lupski, J. R., and Morrow, B. E. (2001). Two functional copies of the DGCR6 gene are present on human chromosome 22q11 due to a duplication of an ancestral locus. *Genome Res* *11*, 208-17.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* *95*, 14863-14868.

Endo, K., Matsuda, Y., and Kobayashi, S. (1997). Mdes, a mouse homolog of the *Drosophila* degenerative spermatocyte gene is expressed during mouse spermatogenesis. *Dev. Growth Differ.* *39*, 399-403.

Fouquet, J., Kann, M., Soues, S., and Melki, R. (2000). ARP1 in Golgi organisation and attachment of manchette microtubules to the nucleus during mammalian spermatogenesis. *J. Cell Sci.* *113*, 877-86.

Fulcher, K. D., Welch, J. E., Klapper, D. G., O'Brien, D. A., and Eddy, E. M. (1995). Identification of a unique mu-class glutathione S-transferase in mouse spermatogenic cells. *Mol. Reprod. Dev.* *42*, 415-24.

Futcher, B. (2000). Microarrays and cell cycle transcription in yeast. *Curr. Opin. Cell Biol.* *12*, 710-715.

Giannattasio, A., Girotti, M., Williams, K., Hall, L., and Bellastella, A. (1997). Puberty influences expression of phospholipid hydroperoxide glutathione peroxidase (GPX4) in rat testis: probable hypophysis regulation of the enzyme in male reproductive tract. *J. Endocrinol. Invest.* 20, 439-44.

Gilbert, S.F. (1994). The saga of the germ line. *Developmental Biology*. 4th Ed. Sunderland: Sinauer Associates, Inc. pp. 798-801.

Grootegoed, J. A., Siep, M., and Baarends, W. M. (2000). Molecular and cellular mechanisms in spermatogenesis. *Baill. Clin. Endocrin. Metab.* 14, 331-343.

Ham, A.W. (1974). The male reproductive system. *Histology*. 7th Ed. Philadelphia: J.B. Lippincott Company, pp. 900-35.

Hansis, C., Jahner, D., Spiess, A. N., Boettcher, K., and Ivell, R. (1998). The gene for the Alzheimer-associated beta-amyloid-binding protein (ERAB) is differentially expressed in the testicular Leydig cells of the azoospermic by w/w(v) mouse. *Eur. J. Biochem.* 258, 53-60.

Harrington, J. J. and Lieber, M. R. (1994). The characterization of a mammalian DNA structure-specific endonuclease. *EMBO J.* 13, 1235-46.

Hogan, B., Costantini, F., and Lacy, E. (1986). Summary of mouse development. *Manipulating the Mouse Embryo: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory. pp. 19-27.

Husen, B., Giebel, J., and Rune, G. (1999). Expression of the integrin subunits alpha 5, alpha 6 and beta 1 in the testes of the common marmoset. *Int. J. Androl.* 22, 374-84.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2002). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Submitted to: *Biostatistics* and available at: <http://biosun01.biostat.jhsph.edu/~ririzarr/papers/index.html>.

Ivanov, I. P., Rohrwasser, A., Terreros, D. A., Gesteland, R. F., and Atkins, J. F. (2000). Discovery of a spermatogenesis stage-specific ornithine decarboxylase antizyme: antizyme 3. *Proc. Natl. Acad. Sci. USA* 97, 4808-13.

Jelinsky, S. A., Estep, P., Church, G. M., and Samson, L. D. (2000). Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol. Cell. Biol.* 20, 8157-67.

Kanai, Y., Kanai-Azuma, M., Tajima, Y., Birk, O. S., Hayashi, Y., and Sanai, Y. (2000). Identification of a stromal cell type characterized by the secretion of a soluble integrin-binding protein, MFG-E8, in mouse early gonadogenesis. *Mech. Dev.* 96, 223-7.

Kao, F. T., Wu, K. C., Law, M. L., Hartz, J. A., and Lau, Y. F. (1988). Assignment of human gene encoding testis-specific lactate dehydrogenase C to chromosome 11, region p14.3-p15.5. *Somat. Cell Mol. Genet.* *14*, 515-8.

Kerr, W. G. and Burrows, P. D. (1991). Stage-specific transcription of germline IgH C gamma and C alpha regions during human B cell differentiation. *Int. Immunol.* *3*, 1059-65.

Kilpatrick, D. L., Zinn, S. A., Fitzgerald, M., Higuchi, H., Sabol, S. L., and Meyerhardt, J. (1990). Transcription of the rat and mouse proenkephalin genes is initiated at distinct sites in spermatogenic and somatic cells. *Mol. Cell. Biol.* *10*, 3717-26.

Kim, E., Waters, S. H., Hake, L. E., and Hecht, N. B. (1989). Identification and developmental expression of a smooth-muscle gamma-actin in postmeiotic male germ cells of mice. *Mol. Cell. Biol.* *9*, 1875-81.

Kolmer, M., Pelto-Huikko, M., Parvinen, M., Hoog, C., and Alho, H. (1997). The transcriptional and translational control of diazepam binding inhibitor expression in rat male germ-line cells. *DNA Cell Biol.* *16*, 59-72.

Komada, M., McLean, D. J., Griswold, M. D., Russell, L. D., and Soriano, P. (2000). E-MAP-115, encoding a microtubule-associated protein, is a retinoic acid-inducible gene required for spermatogenesis. *Genes Dev.* *14*, 1332-42.

Kurth, B. E., Klotz, K., Flickinger, C. J., and Herr, J. C. (1991). Localization of sperm antigen SP-10 during the six stages of the cycle of the seminiferous epithelium in man. *Biol. Reprod.* *44*, 814-21.

Le Douarin, B., Nielsen, A. L., Garnier, J. M., Ichinose, H., Jeanmougin, F., Losson, R., and Chambon, P. (1996). A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J.* *15*, 6701-15.

Lindner, K., Gregan, J., Montgomery, S., and Kearsey, S. E. (2002). Essential role of MCM proteins in premeiotic DNA replication. *Mol. Biol. Cell* *13*, 435-44.

Luoh, S. W., Bain, P. A., Polakiewicz, R. D., Goodheart, M. L., Gardner, H., Jaenisch, R., and Page, D. C. (1997). Zfx mutation results in small animal size and reduced germ cell number in male and female mice. *Development* *124*, 2275-84.

Manandhar, G., Moreno, R. D., Simerly, C., Toshimori, K., and Schatten, G. (2000). Contractile apparatus of the normal and abortive cytokinetic cells during mouse male meiosis. *J. Cell Sci.* *113*, 4275-86.

Matsubara, N., Yanagisawa, M., Nishimune, Y., Obinata, M., and Matsui, Y. (1995). Murine polo like kinase 1 gene is expressed in meiotic testicular germ cells and oocytes. *Mol. Reprod. Dev.* *41*, 407-15.

Matsumoto, M., Kurata, S., Fujimoto, H., and Hoshi, M. (1993). Haploid specific activations of protamine 1 and hsc70t genes in mouse spermatogenesis. *Biochim. Biophys. Acta* *1174*, 274-8.

Mazeyrat, S., Saut, N., Grigoriev, V., Mahadevaiah, S. K., Ojarikre, O. A., Rattigan, A., Bishop, C., Eicher, E. M., Mitchell, M. J., and Burgoyne, P. S. (2001). A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nat. Genet.* *29*, 49-53.

Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M., and Bestor, T. H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* *125*, 889-97.

Mulholland, D. J., Dedhar, S., and Vogl, A. W. (2001). Rat seminiferous epithelium contains a unique junction (Ectoplasmic specialization) with signaling properties both of cell/cell and cell/matrix junctions. *Biol. Reprod.* *64*, 396-407.

Oakberg, E.F. (1956). Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. *Am. J. Anat.* *99*, 507-16.

Pei, L. (1999). Pituitary tumor-transforming gene protein associates with ribosomal protein S10 and a novel human homologue of DnaJ in testicular cells. *J. Biol. Chem.* 274, 3151-8.

Peterson, L.E. (2002). Factor analysis of cluster-specific gene expression levels from cDNA microarrays. *Comput. Methods Programs Biomed.* In press.

Ramalho-Santos, J. and Moreno, R. D. (2001). Targeting and fusion proteins during mammalian spermiogenesis. *Biol. Res.* 34, 147-52.

Rodriguez, I., Ody, C., Araki, K., Garcia, I., and Vassalli, P. (1997). An early and massive wave of germinal cell apoptosis is required for the development of functional spermatogenesis. *EMBO J.* 16, 2262-2270.

Roscoe, W. A., Barr, K. J., Mhawi, A. A., Pomerantz, D. K., and Kidder, G. M. (2001). Failure of spermatogenesis in mice lacking connexin43. *Biol. Reprod.* 65, 829-38.

Russell, D. L. and Kim, K. H. (1996). Expression of triosephosphate isomerase transcripts in rat testis: evidence for retinol regulation and a novel germ cell transcript. *Biol. Reprod.* 55, 11-8.

Russell, L. D., Chiarini-Garcia, H., Korsmeyer, S. J., and Knudson, C. M. (2002). Bax-dependent spermatogonia apoptosis is required for testicular development and spermatogenesis. *Biol. Reprod.* *66*, 950-8.

Russell, L., Ettl, R., Sinha-Hikim, A., and Clegg, E. (1990). Mammalian spermatogenesis. *Histological and Histopathological Evaluation of the Testis*. Illinois:Cache River Press. pp. 1-58.

Schrans-Stassen, B.H., van de Kant, H.J., de Rooij, D.G., and van Pelt, A.M. (1999). Differential expression of c-kit in mouse undifferentiated and differentiating type A spermatogonia. *Endocrinology* *140*, 5894-900.

Sha, J., Baker, P., and O'Shaughnessy, P. J. (1996). Both reductive forms of 17 beta-hydroxysteroid dehydrogenase (types 1 and 3) are expressed during development in the mouse testis. *Biochem. Biophys. Res. Commun.* *222*, 90-4.

Shih, D. M. and Kleene, K. C. (1992). A study by in situ hybridization of the stage of appearance and disappearance of the transition protein 2 and the mitochondrial capsule seleno-protein mRNAs during spermatogenesis in the mouse. *Mol. Reprod. Dev.* *33*, 222-7.

Sneath, P. H. A. and Sokol, R. R. (1973). *Numerical Taxonomy*. W. H. Freeman Co., New York.

Tanaka, K., Tamura, H., Tanaka, H., Katoh, M., Futamata, Y., Seki, N., Nishimune, Y., and Hara, T. (2002). Spermatogonia-dependent expression of testicular genes in mice. *Dev. Biol.* *246*, 466-79.

Thomas, K.H., Wilkie, T.M., Tomashefsky, P., Bellve, A.R., and Simon, M.I. (1989). Differential gene expression during mouse spermatogenesis. *Biol. Reprod.* *41*, 729-39.

Tosaka, Y., Tanaka, H., Yano, Y., Masai, K., Nozaki, M., Yomogida, K., Otani, S., Nojima, H., and Nishimune, Y. (2000). Identification and characterization of testis specific ornithine decarboxylase antizyme (OAZ-t) gene: expression in haploid germ cells and polyamine-induced frameshifting. *Genes Cells* *5*, 265-76.

Tsuruta, J. K., O'Brien, D. A., and Griswold, M. D. (1993). Sertoli cell and germ cell cystatin C: stage-dependent expression of two distinct messenger ribonucleic acid transcripts in rat testes. *Biol. Reprod.* *49*, 1045-54.

Tureci, O., Sahin, U., Zwick, C., Koslowski, M., Seitz, G., and Pfreundschuh, M. (1998). Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens. *Proc. Natl. Acad. Sci. USA* *95*, 5211-6.

Vernon, R. B. and Sage, H. (1989). The calcium-binding protein SPARC is secreted by Leydig and Sertoli cells of the adult mouse testis. *Biol. Reprod.* *40*, 1329-40.

Wakayama, T., Ohashi, K., Mizuno, K., and Iseki, S. (2001). Cloning and characterization of a novel mouse immunoglobulin superfamily gene expressed in early spermatogenic cells. *Mol. Reprod. Dev.* *60*, 158-64.

Welch, J. E., Schatte, E. C., O'Brien, D. A., and Eddy, E. M. (1992). Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells. *Biol. Reprod.* *46*, 869-78.

Werner, T. (2001). Target gene identification from expression array data by promoter analysis. *Biomol. Eng.* *17*, 87-94.

Wine, R. N. and Chapin, R. E. (1999). Adhesion and signaling proteins spatiotemporally associated with spermiation in the rat. *J. Androl.* *20*, 198-213.

Xue, H., O'Neill, D., Wang, X., Wolgemuth, D. J., and Bank, A. (1999). HemT-3, an alternative transcript of mouse gene HemT specific to male germ cells. *Gene* *240*, 193-9.

Conclusions

Characterizing variations in gene expression is essential for thoroughly understanding the molecular events associated with diverse biological functions including: normal cell and tissue physiology, tissue response to DNA damaging conditions, susceptibility to genetic disease, and cellular differentiation during development. The large and continuing increase in genome sequence information has facilitated the development of expression microarrays for analyzing the differential expression of hundreds to thousands of genes in parallel. The goal of this research was to use expression microarray technology to (a) understand the natural tissue variation in stress response and DNA repair-associated gene expression and (b) characterize differential gene expression during male germ cell differentiation from mitotic to meiotic cells and discover new candidate genes associated with the onset and progression of spermatogenesis.

To address these issues, this research utilized two separate approaches. Custom cDNA microarrays comprised of genes selected *a priori* were used to characterize the variation in gene expression among tissues, and random oligonucleotide microarrays were used to discover new genes associated with early male germ cell differentiation. First, cDNA microarray hybridization and image analysis techniques were optimized (Chapter 2). A custom cDNA microarray was built to detect the differential basal

expression of 417 genes associated with chromatin remodeling, damage control, DNA repair, growth regulation, meiosis, stress response, transcription, and translation among adult mouse tissues (testis, brain, liver, spleen, and heart). Approximately 25% of the arrayed genes were differentially expressed among healthy tissues. Genes associated with stress response showed the most variation among tissues, while DNA repair genes showed the least variation. Damage control genes showed intermediate variation (Chapter 3). These findings identify candidate stress response, DNA repair, and damage control genes whose variation in expression among tissues may contribute to tissue-specific differences in the response to DNA damage and in genetic disease susceptibility.

A random oligonucleotide microarray was then utilized to identify genes whose expression profile changed as germ cells differentiated from spermatogonia into primary spermatocytes during the first wave of spermatogenesis (Chapter 4). Expression profiling of the prepubertal mouse testis characterized the temporal modulations in gene expression that underlie cellular differentiation during spermatogenesis. Specifically, differential expression was observed for 231 annotated genes and 313 unannotated sequences (544 total). Spermatogenesis-related roles were previously reported for only 53 of these 231 annotated genes (~23%). Thus, this research has identified potential spermatogenesis-related functions for 178 annotated genes and 313 unannotated sequences that had not been previously associated with male germ cell development. The identification of ~500 new candidate genes associated with the onset and progression of spermatogenesis opens the door to a better understanding of the molecular events controlling male germ cell development.

The findings of this dissertation research suggest that future work is needed to (a)

link tissue-specific variations in gene expression with genetic disease phenotypes, (b) characterize the specific functions of candidate genes associated with germ cell progression through spermatogenesis, (c) identify groups of genes with coordinated expression (i.e., syn-expression groups), (d) assess differences in transcript vs. protein abundance, and (e) investigate the relative contributions of tissue microenvironments to differences in expression.

APPENDIX A

Symbols, names, biological pathways, I.M.A.G.E. clone IDs, and cDNA microarray data for 152 genes with differential expression among healthy adult mouse tissues

Complete list of microarray information for genes identified as having differential expression among healthy, adult mouse tissues ^a

Gene Symbol	Gene Name	Pathways (General)	I.M.A.G.E. ID	Ratio to Pooled Testis					F Statistic ^b			P value ^c
				Testis	Spleen	Heart	Liver	Brain	Num.	Denom.	Ratio	
14-3-3 gamma	14-3-3 gamma	Growth regulation	636783	0.95	0.66	0.74	0.60	0.78	0.13	0.16	0.8	1.00
14-3-3 tau	14-3-3 tau	Growth regulation	849982	1.00	0.66	0.43	0.54	1.43	0.95	*0.03870	20.9	0.12
Acetyltransferase	Acetyltransferase	Chromatin related	577211	0.89	0.47	0.89	1.66	0.66	0.92	0.10	9.3	0.20
Acrv1	Acrosomal vesicle protein 1	Other	515873	0.65	0.15	0.20	0.34	0.28	1.27	0.10	12.8	0.13
Actb	Actin, beta	Housekeeping	992475, 989387	1.13	2.86	2.06	0.69	0.66	1.78	0.20	8.8	0.22
Adk1	Adenylate kinase isozyme 1	Stress response/ Stress inducible	602526	0.92	0.52	0.73	0.65	0.77	0.19	0.09	2.1	0.98
Apg-1b	Apg-1b	Stress response/ Stress induced	901984	1.11	1.49	2.48	2.18	1.81	0.42	*0.02340	9.2	0.21
B2m	Beta-2 microglobulin	Growth regulation	596438	1.06	9.90	2.68	5.87	1.79	3.36	0.24	13.8	0.12
Bmp4	Bone morphogenetic protein 4	Damage recognition, cell cycle, apoptosis	888656, 865694, 873328	1.08	0.92	2.68	1.82	1.53	0.75	*0.04070	16.4	0.12
Bp-h5	Brain protein h5	Other	876821	1.33	0.58	6.52	12.30	2.13	6.23	*0.03120	137.0	0.00
Btg1	B-cell translocation gene 1	Damage recognition, cell cycle, apoptosis	634804	0.99	5.66	0.41	0.83	0.81	3.99	0.07	59.6	0.01
Cas1	Catalase 1	Stress response/ Stress induced	573840	1.10	1.23	1.43	8.28	0.98	3.26	0.10	31.3	0.03
Ctnnb1	Catenin, beta 1	Transcription/ Translation	777028	1.00	1.86	2.47	1.46	1.73	0.46	0.05	10.1	0.18
Cdc2	Cell division cycle 2	Damage recognition, cell cycle, apoptosis	375194, 763260	0.96	0.60	0.60	0.71	0.60	0.17	0.07	2.5	0.91
Cdc2a	Cell division cycle 2a	Damage recognition, cell cycle, apoptosis	582109	1.04	0.64	0.54	0.59	0.69	0.27	0.05	5.4	0.52
Cdc42	Cell division cycle 42	Damage recognition, cell cycle, apoptosis	977455	1.19	2.88	2.15	2.28	2.86	0.54	0.05	10.2	0.18
Cdk2	Cyclin dependent kinase 2	Damage recognition, cell cycle, apoptosis	876724	0.93	0.19	0.20	0.46	0.33	1.81	0.12	15.4	0.12
Pki	Camp-dependent protein kinase inhibitor	Growth regulation	514418	0.66	0.18	0.19	0.25	0.28	1.11	0.06	18.2	0.12
Clu	Clusterin	Other	617298	0.87	1.44	0.87	2.04	1.09	0.55	0.11	4.8	0.58
Crem	CAMP responsive element modulator	Growth regulation	917671	0.76	0.25	0.33	0.50	0.57	0.80	0.36	2.2	0.96
Crem-like	CAMP responsive element modulator-like	Growth regulation	917671	0.85	0.18	0.26	0.30	0.38	1.39	0.13	10.6	0.17
Cut3	Cut3	Chromatin related	643940	0.79	0.23	0.22	0.87	0.50	1.83	0.18	9.9	0.19
Ccnf	Cyclin-F	Damage recognition, cell cycle, apoptosis	888861	0.99	0.95	2.20	1.03	0.87	0.59	0.78	0.8	1.00
Ck17	Cytokeratin-17	Housekeeping	975849	0.86	1.25	2.42	1.16	1.20	0.60	0.34	1.8	1.00
Dapk1	DAP-kinase related protein 1	Other	403602	1.06	1.68	1.05	2.46	1.13	0.58	0.12	4.9	0.58
Dby	DEAD/H box polypeptide, Y chromosome	Meiosis/ Spermatogenesis	573643	1.04	1.51	1.43	1.33	1.69	0.14	*0.01640	3.0	0.83
Dnaj	Dnaj	Stress response/ Stress inducible	602316, 602961	0.67	0.11	0.16	0.17	0.20	1.99	*0.03350	43.5	0.01
Dnaj-like	Dnaj-like	Stress response/ Stress inducible	514436	0.93	0.25	0.47	0.61	0.57	0.94	*0.01690	20.7	0.12
E2f1	E2f transcription factor 1	Transcription/ Translation	605037	0.87	0.94	2.24	0.93	1.17	0.64	*0.00859	13.9	0.12
E2f3	E2f transcription factor 3	Transcription/ Translation	539249	1.02	0.46	0.48	0.54	0.78	0.49	0.08	6.2	0.38
Faf1	Fas(TNFRSF6)-associated factor 1	Damage recognition, cell cycle, apoptosis	574610	0.91	0.55	0.81	0.85	0.77	0.16	*0.01930	3.4	0.75
Fbr-MuSV	Finkel-Biskis-Reilly murine sarcoma virus	Other	805511	0.87	2.88	0.67	1.61	1.07	1.37	0.05	28.4	0.03
Figf	C-fos induced growth factor	Transcription/ Translation	614347	1.18	1.01	0.48	1.32	1.09	0.66	0.56	1.2	1.00
Gadd153	Growth arrest and DNA damage inducible 153	Stress response/ Stress inducible	903718	1.07	0.59	5.44	1.88	1.77	2.79	0.10	28.6	0.03
Gap43	Growth accentuating protein 43	Damage recognition, cell cycle, apoptosis	479228	1.06	1.02	1.02	1.26	2.52	0.62	*0.01960	13.7	0.12
Gas2	Growth arrest specific 2	Damage recognition, cell cycle, apoptosis	520303	0.92	8.57	3.70	3.01	2.20	2.73	0.33	8.3	0.23
Gli	Glioma-associated oncogene homolog	Transcription/ Translation	386514	1.08	1.09	0.73	0.60	0.65	0.33	0.14	2.5	0.91
Gpx1	Glutathione peroxidase 1	Stress response/ Stress induced	861820	1.10	2.09	26.50	4.15	3.38	5.93	0.13	46.2	0.01
Grp78	Glucose regulated protein, 78 kDa	Stress response/ Stress inducible	935093	0.89	0.96	1.17	1.88	0.82	0.46	0.12	3.8	0.70
Gstp2	Glutathione S-transferase, pi 2	Stress response/ Stress inducible	864333	1.13	0.95	18.00	5.18	3.95	6.05	0.06	106.0	0.00
Hdac1	Histone deacetylase 1	Chromatin related	641105	1.02	0.94	0.53	0.63	0.80	0.32	0.05	6.0	0.41
Histone-1	Histone-1	Chromatin related	493091	1.28	2.58	2.31	1.41	1.63	0.39	0.23	1.7	1.00
Histone-H1.1	Histone-H1.1	Chromatin related	483515	0.84	2.63	0.74	1.34	1.00	1.06	0.12	8.9	0.22

Gene Symbol	Gene Name	Pathways (General)	I.M.A.G.E. ID	Ratio to Pooled Testis					F Statistic ^b			P value ^c
				Testis	Spleen	Heart	Liver	Brain	Num.	Denom.	Ratio	
Histone-H3.3A	Histone-H3.3A	Chromatin related	961662	0.97	0.49	0.39	0.26	0.44	0.95	0.14	6.9	0.33
Histone-H4	Histone-H4	Chromatin related	493223	0.97	1.73	14.80	4.43	9.97	5.47	0.09	60.6	0.01
Hk1	Hexokinase 1	Growth regulation	973936	0.65	1.38	0.27	0.81	0.79	1.46	0.07	19.5	0.12
Hmox2	Heme oxygenase 2	Stress response/ Stress induced	602116	0.82	0.33	0.28	0.54	0.63	0.82	0.05	15.1	0.12
hnRNP-A2	heterogenous nuclear ribonucleoprotein A2/B1	Transcription/ Translation	721779	1.52	1.71	1.18	1.21	1.51	0.11	0.11	1.0	1.00
hnRP-A1	heterogenous nuclear ribonucleoprotein A1	Transcription/ Translation	386370	1.14	3.66	1.55	1.11	0.93	1.24	0.09	14.2	0.12
hTAFII18-like	PolII transcription factor TFIID chain hTAFII18	Transcription/ Translation	523974	0.78	0.42	0.70	0.76	0.89	0.36	0.07	4.9	0.58
Idb2	Inhibitor of DNA binding 2	Transcription/ Translation	608134	0.88	1.19	1.95	1.39	1.46	0.36	*0.02120	7.8	0.24
Jun	v-jun sarcoma virus 17 oncogene	Growth regulation	975691	0.80	1.61	3.24	1.10	1.24	1.16	1.11	1.0	1.00
			975676, 975855,									
Keratin-like	Keratin-like	Housekeeping	608390, 608109	1.08	1.58	2.71	2.06	1.49	0.50	*0.02260	11.0	0.15
Kif1a	Kinesin heavy chain member 1A	Growth regulation	492514	1.12	15.90	4.84	2.76	1.95	4.26	0.45	9.5	0.19
Kif1b	Kinesin heavy chain member 1B	Growth regulation	560049	1.08	1.83	1.51	1.49	1.03	0.25	*0.00513	5.5	0.50
Kif2	Kinesin heavy chain member 2	Growth regulation	602145	0.94	0.34	0.90	1.08	1.06	0.97	*0.03390	21.2	0.11
Kif3b	Kinesin family member 3b	Growth regulation	1108812	1.08	0.76	0.48	0.52	0.75	0.44	0.06	7.8	0.24
Kinesin-2-related	Kinesin-2-related	Growth regulation	990126	0.71	0.95	0.53	0.75	0.85	0.21	0.06	3.7	0.70
Krag	Kras oncogene-associated gene	Growth regulation	484140	1.06	2.80	1.34	2.14	1.13	0.75	0.10	7.8	0.24
Kras2	Kirsten rat sarcoma oncogene 2	Growth regulation	572995	1.04	1.17	3.84	1.24	1.31	1.18	2.31	0.5	1.00
Krt1-10	Keratin complex 1, acidic, gene 10	Housekeeping	976351	0.93	0.37	0.26	0.41	0.48	0.90	0.42	2.1	0.96
Krt1-13	Keratin complex 1, acidic, gene 13	Housekeeping	521007	1.11	2.37	2.88	1.28	0.95	1.00	*0.03720	21.9	0.11
Lama2	Laminin, alpha 2	Other	493133	1.45	8.51	5.90	2.14	1.89	2.49	0.07	37.2	0.01
Lama4	Laminin, alpha 4	Other	584781	1.10	0.87	15.40	2.67	3.27	5.36	*0.00823	118.0	0.00
Mak	Male germ cell associated kinase	Growth regulation	602281	0.94	0.16	0.19	0.24	0.27	2.06	*0.02610	45.2	0.01
			749454, 439956,									
Mapk2	Mitogen activated protein kinase 2	Growth regulation	554209	1.07	0.91	0.79	2.05	0.90	0.59	*0.02690	12.9	0.13
Mcl1	Myeloid cell leukemia sequence 1	Damage recognition, cell cycle, apoptosis	493170	1.20	2.38	9.42	4.32	2.46	2.47	0.15	16.5	0.12
Mea1-like	Male enhanced antigen 1-like	Other	315756	0.64	0.40	0.40	0.56	0.82	0.40	*0.01120	8.8	0.22
Mea1	Male enhanced antigen 1	Other	608800	1.01	0.11	0.18	0.20	0.21	2.89	0.18	16.1	0.12
Meg1	Meiosis expressed gene 1	Other	514389	0.75	0.11	0.11	0.18	0.22	2.52	*0.04480	55.3	0.01
Map2k1	Mitogen activated protein kinase kinase 1	Meiosis/ Spermatogenesis	616850	1.17	2.03	1.50	1.24	1.37	0.19	*0.03230	4.2	0.58
Map3k1	Mitogen activated protein kinase kinase kinase 1	Growth regulation	875311	0.90	0.25	0.49	0.44	0.43	0.89	0.16	5.5	0.51
Mgmt	0-6-methylguanine-DNA methyltransferase	DNA repair (direct reversal)	493108	1.12	1.11	2.83	1.56	1.38	0.61	0.13	4.9	0.58
Mig-2	Mitogen inducible gene 2	Meiosis/ Spermatogenesis	988186	1.24	1.07	5.50	18.40	3.14	5.67	0.88	6.4	0.38
Mkk7	Mitogen-activated protein kinase kinase 7	Growth regulation	821017	0.96	1.87	1.60	1.63	4.90	1.48	*0.00408	32.4	0.02
Mlc2	Myosin light chain 2	Housekeeping	556208, 604573	1.04	0.96	2.91	0.94	1.11	0.95	*0.04520	20.8	0.12
Mns1	Meiosis-specific nuclear structural protein 1	Meiosis/ Spermatogenesis	538140	1.11	0.23	0.36	0.27	0.37	1.58	0.32	5.0	0.58
Kzf2	Kruppel zinc finger protein 2	Other	789990	1.11	0.56	0.70	0.51	0.87	0.42	*0.00970	9.2	0.20
RbAp48	Retinoblastoma-binding protein	Other	660074	0.77	1.93	0.57	0.95	0.66	0.96	0.15	6.5	0.38
Mre11b	Meiotic recombination 11 homolog b	DNA Repair (RR-Non-Homologous end joi	524361	0.96	0.46	0.74	0.62	0.77	0.32	*0.02260	7.0	0.32
Msk2	Mitogen- and stress-activated protein kinase-2	Growth regulation	873975	1.03	1.42	1.73	1.26	0.99	0.22	0.10	2.2	0.96
Mt2	Metallothionein 2	Other	963382	0.80	0.26	0.24	0.34	1.11	1.99	0.07	27.9	0.03
Mybl2	Myeloblastosis oncogene-like 2	Growth regulation	532188	1.02	0.54	0.69	0.84	0.69	0.24	*0.03350	5.3	0.54
Myla	Myosin light chain, alkali	Housekeeping	586078	0.94	1.24	5.22	1.32	1.21	1.94	0.07	27.2	0.05

Gene Symbol	Gene Name	Pathways (General)	I.M.A.G.E. ID	Ratio to Pooled Testis					F Statistic ^b			P value ^c
				Testis	Spleen	Heart	Liver	Brain	Num.	Denom.	Ratio	
Naip1	Neuronal apoptosis inhibitory protein 1	Damage recognition, cell cycle, apoptosis	864370	0.73	1.23	2.38	1.01	1.15	0.78	1.43	0.5	1.00
Naip2	Neuronal apoptosis inhibitory protein 2	Damage recognition, cell cycle, apoptosis	737787	0.93	0.73	0.89	4.93	0.99	2.53	0.12	20.5	0.12
Odf1	Outer dense fiber of sperm tails 1	Other	603127	0.86	0.40	0.50	0.63	0.51	0.35	0.23	1.5	1.00
p18	p18	Transcription/ Translation	474080	0.90	0.31	0.50	0.49	0.65	0.65	0.14	4.6	0.58
p40	p40	DNA repair (NER)	931450	0.90	0.38	0.53	0.54	0.62	0.40	0.11	3.8	0.70
p52	p52	DNA repair (NER)	443359	1.15	0.82	0.60	0.92	0.80	0.24	0.05	5.0	0.58
Pabp	PolyA binding protein	Transcription/ Translation	891504, 977971	0.70	0.54	0.33	0.29	0.37	0.55	*0.04060	12.1	0.15
Pabp (testis)	PolyA binding protein (testis-enriched isoform)	Transcription/ Translation	516680	0.67	0.29	0.15	0.26	0.21	1.32	0.20	6.5	0.38
Paga	Proliferation-associated gene A	Damage recognition, cell cycle, apoptosis	617734	0.93	1.31	2.67	12.90	1.68	4.43	*0.03300	97.1	0.00
Pcna	Proliferating cell nuclear antigen	DNA repair (NER)	907695	1.04	0.72	0.31	0.37	0.48	1.02	0.14	7.3	0.28
Pcsk2	Proprotein convertase subtilisin/kexin type 2	Other	477515	0.91	1.18	0.56	0.83	0.59	0.41	0.05	8.1	0.23
Pctk3	PCTAIRE-motif protein kinase 3	Damage recognition, cell cycle, apoptosis	579416	1.07	1.35	1.08	1.42	2.34	0.42	0.14	2.9	0.87
Pgk1	Phosphoglycerate kinase 1	Growth regulation	822652	0.98	0.74	2.47	1.70	1.68	0.97	0.06	15.8	0.12
Phb	Prohibitin	Other	584178	0.94	0.47	1.42	1.07	0.66	0.78	0.06	13.0	0.13
Pkcq	Protein kinase C, theta	Growth regulation	582973	0.86	0.59	0.99	0.83	1.06	0.21	0.07	3.0	0.84
Pla2g2c	Phospholipase A2, group IIC	Housekeeping	513783	1.13	1.52	21.00	6.50	5.30	5.83	*0.02490	128.0	0.00
Pol-β	DNA polymerase beta	DNA repair (BER)	918389	1.00	0.17	0.18	0.26	0.46	2.32	*0.04520	50.8	0.01
Mapk1	Mitogen activated protein kinase 1	Growth regulation	736825, 634946	1.21	1.64	1.65	1.26	2.72	0.44	0.10	4.2	0.60
Prkmk1	Prkmk1	Growth regulation	585802	0.87	1.67	1.22	1.76	2.23	0.55	0.06	8.8	0.22
Prm1	Protamine 1	Meiosis/ Spermatogenesis	918252	0.41	0.17	0.18	0.16	0.13	0.79	0.29	2.7	0.87
Prm2	Protamine 2	Meiosis/ Spermatogenesis	917852	0.94	0.36	0.33	0.61	0.53	0.75	0.06	12.1	0.15
Prm3	Protamine 3	Meiosis/ Spermatogenesis	907309	1.11	2.16	1.42	0.88	1.01	0.52	0.31	1.7	1.00
Prp22	Pre-mRNA splicing factor RNA helicase	Transcription/ Translation	604173	1.05	1.10	0.29	0.42	0.45	1.45	0.18	8.2	0.23
Ptgds	Prostaglandin D2 synthase	Other	571621	1.46	1.31	16.70	5.93	4.98	4.68	0.09	54.4	0.01
Rad51-ap1	Rad51 associated protein 1	DNA repair related	849369	1.13	0.80	2.23	1.06	0.96	0.64	0.86	0.7	1.00
Rad51b	Rad51b	DNA repair (Recombinational repair-Homo	1225890, 1246004	1.33	1.13	1.04	1.97	1.11	0.28	0.08	3.3	0.77
Rara	Retinoic acid receptor, alpha	Other	475996	1.02	0.74	0.74	0.61	0.48	0.32	*0.02520	7.1	0.31
Rbm3	RNA binding motif protein 3	Transcription/ Translation	872105	1.16	1.78	14.40	6.90	3.07	4.30	*0.04340	94.2	0.00
Rec2	RecA-like protein	Other	1246004	0.87	1.45	0.91	2.94	1.44	1.00	0.06	17.6	0.12
Req	Requiem	Damage recognition, cell cycle, apoptosis	573346	1.03	0.95	3.30	0.70	0.75	1.66	1.14	1.5	1.00
Rip1	Ral-interacting protein 1	Damage recognition, cell cycle, apoptosis	849333	1.11	0.98	19.40	3.52	3.59	6.02	0.13	47.8	0.01
RNA pol transcrip. reg. med.	RNA polymerase transcriptional regulation mediator	Transcription/ Translation	635118	0.95	0.68	0.53	0.67	0.98	0.28	0.07	4.2	0.58
Rxrb	Retinoid X receptor beta	Other	493651	1.27	4.79	6.07	1.92	1.65	1.98	0.70	2.8	0.87
Rxrg	Retinoid X receptor gamma	Other	479866	1.03	3.51	3.16	1.54	1.07	1.43	*0.00852	31.4	0.03
Smp30	Senescence marker protein-30	Damage recognition, cell cycle, apoptosis	521164, 747954	1.07	1.82	2.72	10.50	1.97	3.06	0.07	41.3	0.01
Snf2l	SWI/SNF-related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	Transcription/ Translation	960224	0.98	0.73	3.64	0.92	1.15	1.65	*0.02490	36.3	0.01
Snf5l	SWI/SNF-related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	Transcription/ Translation	580874	1.08	0.91	0.90	2.51	0.94	0.80	*0.00704	17.5	0.12
Snrpn	Small nuclear ribonucleoprotein N	Transcription/ Translation	977306	1.01	0.72	0.68	0.84	2.36	1.07	0.05	19.5	0.12
Sod1	Superoxide dismutase 1	Stress response/ Stress induced	762216	0.98	0.59	0.63	1.06	0.64	0.32	0.14	2.3	0.93
Sod3	Superoxide dismutase 3	Stress response/ Stress induced	776821	1.16	0.78	8.78	2.57	2.58	3.59	0.06	62.4	0.01

Gene Symbol	Gene Name	Pathways (General)	I.M.A.G.E. ID	Ratio to Pooled Testis					F Statistic ^b			P value ^c
				Testis	Spleen	Heart	Liver	Brain	Num.	Denom.	Ratio	
Sox6	SRY-box containing gene 6	Transcription/ Translation	1068053	1.11	0.95	4.38	2.21	1.74	1.54	0.05	31.1	0.03
Srm	Spermidine-synthase	Other	582569	0.89	1.09	0.88	0.90	2.04	0.54	0.13	4.3	0.58
Spnr	Spermatid perinuclear RNA-binding protein	Transcription/ Translation	764543	0.87	0.34	0.30	0.40	0.78	0.99	0.07	14.8	0.12
Stat3	Signal transducer and activator of transcription 3	Growth regulation	975718	0.97	0.76	1.85	0.66	0.58	0.88	0.64	1.4	1.00
Stat4	Signal transducer and activator of transcription 4	Transcription/ Translation	577343	0.77	0.19	0.18	0.98	0.34	2.55	0.19	13.5	0.12
Sycp3	Synaptonemal complex protein 3	Chromatin related	874980, 1230236	0.92	0.19	0.18	0.26	0.27	1.81	0.07	25.8	0.06
Tak1	TGF-beta-activated kinase	Growth regulation	865301	0.96	0.92	2.24	0.91	0.81	0.71	0.86	0.8	1.00
Tb4y	Thymosin, beta 4, Y chromosome	Meiosis/ Spermatogenesis	273224, 325197	0.93	7.43	3.07	0.88	5.26	4.01	0.09	43.8	0.01
Tbp	TATA box binding protein	Transcription/ Translation	573420	0.99	0.61	0.61	0.60	0.70	0.19	0.15	1.3	1.00
Tctex2	T-complex testis-expressed 2	Meiosis/ Spermatogenesis	514791, 515753	0.65	0.34	0.20	0.15	0.22	1.36	0.27	5.0	0.58
Tdgb	T:G mismatch-specific thymine-DNA glycosylase	DNA repair (BER)	990644	1.25	4.37	1.87	1.80	1.89	0.88	*0.00642	19.3	0.12
Tenr	Testis nuclear RNA-binding protein	Transcription/ Translation	602129	0.93	0.39	0.86	1.14	1.09	0.80	0.14	5.8	0.46
Testin-2	Testin-2	Other	917479	0.90	0.35	0.91	0.48	0.45	0.78	0.21	3.8	0.70
Pcd10	Programmed cell death 10	Damage recognition, cell cycle, apoptosis	822588	0.90	0.71	0.54	0.88	1.11	0.32	*0.01190	7.0	0.32
TIFII-B	Transcription initiation factor IIB	Transcription/ Translation	315324	0.98	0.77	0.55	0.61	0.87	0.24	*0.04490	5.2	0.55
Timp2	Tissue inhibitor of metalloproteinase 2	Adhesion/ Extracellular matrix	535104	1.47	1.62	1.82	1.18	1.22	0.14	0.11	1.4	1.00
Tk1	Thymidine kinase 1	Other	556061	0.97	0.99	3.93	0.90	1.14	1.60	*0.01860	35.0	0.01
Tnp1	Transition protein 1	Chromatin related	602551	0.58	0.16	0.18	0.17	0.15	1.37	0.11	12.1	0.15
Top1	Topoisomerase (DNA) I	Chromatin related	873675	0.94	1.36	2.89	1.07	1.21	0.80	0.39	2.1	0.98
Top2b	Topoisomerase (DNA) II beta	Chromatin related	652824	0.95	1.62	2.27	1.04	1.47	0.52	0.14	3.6	0.70
Tssk2	Testis specific serine kinase 2	Growth regulation	602020	0.93	0.14	0.22	0.28	0.26	2.00	0.12	16.6	0.12
Tuba2	Tubulin alpha 2	Housekeeping	919504	1.10	1.42	0.89	0.60	1.10	0.43	*0.04300	9.4	0.20
Ube2b	Ubiquitin-conjugating enzyme E2B	Other	577631	0.98	0.50	1.22	0.73	0.64	0.52	0.09	6.0	0.40
Ube3a	Ubiquitin conjugating enzyme E3A	Other	538653	0.93	0.38	0.52	0.60	0.91	0.59	0.13	4.5	0.58
Ung	Uracil-DNA glycosylase	DNA repair (BER)	355462, 406824, 931428	0.92	0.50	0.33	0.49	0.53	0.56	0.29	2.0	0.99
Vegf	Vascular endothelial growth factor	Damage recognition, cell cycle, apoptosis	988133	1.03	1.19	0.39	1.23	1.20	1.03	0.66	1.6	1.00
Wnt10b	Wingless related MMTV integration site 10b	Growth regulation	439485	1.10	0.33	0.23	0.23	1.47	3.28	0.07	47.5	0.01
Zfp105	Zinc finger protein 105	Damage recognition, cell cycle, apoptosis	536078	1.23	1.06	2.38	1.04	1.45	0.48	0.20	2.4	0.92

^aData for the 152 genes with expression ratios ≥ 0.60 or ≤ 1.68 in one or more tissues

^bDenominators with an asterisk were replaced by the median of the denominators (0.0456) before calculating the ratio

^cPermutation p value

APPENDIX B

All annotated genes showing significant differences in expression during the transition from mitosis to meiosis in the prepubertal mouse testis

All annotated genes showing significant differences in expression between spermatogonial mitosis and preleptotene

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with increased expression as germ cells progress from mitosis to the onset of meiosis						
Seleno-protein	Sep	Misc.	M88463	9.8	0.04	Yes (Shih and Kleene, 1992)
Ig germline H-chain gene V-region	IgH (germline)	Immune response	M16724	6.8	0.001	Yes (Kerr and Burrows, 1991)
Immunosuperfamily protein B12	Igsf4	Immune response	AF061260	4.0	0.04	Yes (Wakayama et al., 2001)
Zinc finger protein X-linked	Zfx	Cellular differentiation	M32309	2.7	0.05	Yes (Luoh et al., 1997)
Adipocyte-specific protein adipoQ	AdipoQ	Unknown	U49915	37.0	0.02	No
Glioblastoma amplified sequence	Gbas	Unknown	AJ001261	22.2	0.002	No
Methylmalonyl coenzyme A mutase	Mut	Metabolism	X51941	11.9	0.04	No
Sprouty-4	Spry4	Signal transduction	AB019280	5.6	0.02	No
Short stature homeobox 2	Shox2	Transcription	U66918	4.1	0.04	No
Transactivating transcription factor 3	Sp3	Transcription	AF062567	3.8	0.01	No
Histone H2a(A), H2a(B), and H2b	H2A-B	Chromatin-related	U62673	2.8	0.03	No
ADP-ribosyltransferase 5	Art5	Protein modification	U60881	2.3	0.03	No
Cryptochrome 1	Cry1	Molecular clock	AB000777	2.1	0.01	No
Intracisternal A-particle type IIB	Iap2	Growth regulation	X16672	2.1	0.03	No
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis						
DiGeorge syndrome critical region 6	Dgcr6	Unknown	AF021031	143.8	0.02	Yes (Edelmann et al., 2001)
Glutathione peroxidase 4	Gpx4	Stress response	D87896	44.1	0.03	Yes (Giannattasio et al., 1997)
Budding inhibited by benzimidazoles 1	Bub1	Cell cycle	AF002823	22.6	0.02	Yes (Bernard et al., 2001)
Hydroxysteroid (17-beta) dehydrogenase 10	Hsd17b10	Metabolism	U96116	20.4	0.002	Yes (Hansis et al., 1998)
High mobility group 1	Hmgb1	Cellular differentiation	U00431	20.2	0.0003	Yes (Bucci et al., 1984)
Hexosaminidase A	Hexa	Metabolism	U05837	20.1	0.02	Yes (Adamali et al., 1999a,b)
Integrin linked kinase	Ilk	Signal transduction	U94479	15.7	0.05	Yes (Mulholland et al., 2001)
Actin-related protein 2/3 1A	Arpc1a	Cytoskeletal element	AB024984	13.2	0.01	Yes (Fouquet et al., 2000)
Glutathione s-transferase, MU	Gstm5	Cytoskeletal element	U24428	13.0	0.02	Yes (Fulcher et al., 1995)
DEAD box polypeptide, Y chromosome	Dby	RNA helicase	AJ007376	11.4	0.03	Yes (Mazeyrat et al., 2001)
Protein kinase, cAMP dependent, catalytic, alpha	Prkaca	Metabolism	M12303	11.1	0.005	Yes (Desseyn et al., 2000)

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis (continued)						
Milk fat globule-EGF factor 8	Mfge8	Cell adhesion	M38337	9.5	0.002	Yes (Kanai et al., 2000)
Acrosomal vesicle protein 1	Acvp1	Fertilization	U31992	8.9	0.02	Yes (Kurth et al., 1991)
NIMA-related expressed kinase 2	Nek2	Cell cycle	AF013166	8.8	0.004	Yes (Di Agostino et al., 2002)
Endothelin 1	Edn1	Signal transduction	U35233	7.7	0.01	Yes (Cai et al., 2000)
Cystatin 3	Cst3	Cellular differentiation	U10098	7.6	0.02	Yes (Tsuruta et al., 1993)
Flap structure specific endonuclease 1	Fen1	DNA repair	L26320	7.0	0.04	Yes (Harrington and Lieber, 1994)
Triosephosphate isomerase	Tpi	Metabolism	L31777	6.7	0.03	Yes (Russell and Kim, 1996)
Bcl2-associated X protein	Bax	Apoptosis	L22472	6.4	0.01	Yes (Russell et al., 2002)
Catenin alpha 1	Catn1	Cell adhesion	X59990	4.8	0.01	Yes (Wine and Chapin, 1999)
Phosphoglycerate mutase muscle-specific	Pgam-m	Metabolism	AF029843	4.6	0.05	Yes (Broceno et al., 1995)
Tubulin alpha 2	Tuba2	Cytoskeletal element	M28727	4.2	0.01	Yes (Bo and Wensink, 1989)
Mcm2	Mcm2	DNA replication	D86725	3.9	0.02	Yes (Lindner et al., 2002)
TIF1 beta	Tif1b	Transcription	X99644	3.6	0.04	Yes (Le Douarin et al., 1996)
Breast cancer 2	Brca2	DNA repair	U89652	3.5	0.02	Yes (Connor et al., 1997)
Microtubule-associated protein 7	Mtap7	Cytoskeletal element	Y15197	3.0	0.02	Yes (Komada et al., 2000)
Rab7	Rab7	Intracellular transport	Y13361	2.7	0.03	Yes (Ramalho-Santos and Moreno, 2001)
Polo-like kinase 1	Plk1	Cell cycle	U73170	2.6	0.03	Yes (Matsubara et al., 1995)
RhoB	Arhb	Cytoskeletal element	X99963	2.4	0.002	Yes (Castellano et al., 1997)
Preproenkephalin related	Penk-rs	Signal transduction	M55181	2.0	0.02	Yes (Kilpatrick et al., 1990)
Heat shock cognate 70 (testis)	Hsc70t	Stress response	AF109905	2.0	0.02	Yes (Matsumoto et al., 1993)
RNA binding motif 3	Rbm3	Stress response	AB016424	1.9	0.04	Yes (Danno et al., 2000)
Actin, gamma 2	Actg2	Cytoskeletal element	U20365	1.8	0.05	Yes (Kim et al., 1989)
Peroxiredoxin protein 2	Prdx2	Stress response	AF093853	1752.9	0.03	No
LIM protein-1	Lhx1	Transcription	D88792	105.0	0.02	No
TPR-containing, SH2-binding phosphoprotein	Tsbp	Misc.	L49502	66.9	0.02	No
Calponin 2	Cnn2	Cytoskeletal element	Z19543	43.1	0.01	No
Isocitrate dehydrogenase 3 (NAD+)	Idh3	Metabolism	U68564	41.9	0.02	No
Dynactin 3	Dctn3	Cytoskeletal element	AF098508	39.0	0.01	No
Diacylglycerol acyltransferase	Dgat	Metabolism	AF078752	36.3	0.004	No
RNA binding protein regulatory subunit	Djl	Growth regulation	AB015652	34.7	0.05	No
Rab6/rab5-associated protein	Rabac1	Intracellular transport	L40934	33.2	0.002	No
SRY-box containing gene 18	Sox18	Transcription	L35032	32.9	0.04	No
Peptidylprolyl isomerase B	Ppib	Protein modification	X58990	27.3	0.05	No
Chaperonin subunit 3	Cct3	Protein modification	L20509	24.5	0.01	No

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis (continued)						
Ubiquitin c-terminal hydrolase related	Uchrp	Protein modification	D84096	24.3	0.03	No
Viral envelope like protein (<i>G7e</i>)	G7e	Unknown	U69488	23.0	0.05	No
RNA polymerase I associated factor	Paf53	Transcription	D14336	22.5	0.05	No
Centromere protein A	Cenp-a	Chromatin-related	AF012710	22.2	0.04	No
Thioredoxin reductase 1	Txnrd1	Stress response	AB027565	20.8	0.01	No
Serine protease inhibitor 4	Serpine2	Misc.	X70296	20.7	0.02	No
Alpha-actinin-2 associated LIM protein	Alp	Cytoskeletal element	AF002283	20.0	0.04	No
Tuberous sclerosis 2	Tsc2	Growth regulation	U39818	19.6	0.02	No
Baf53a	Baf53a	Chromatin-related	AF041476	18.7	0.05	No
NIK-related kinase	Nrk	Signal transduction	AB020741	18.2	0.05	No
Adaptor protein complex gamma	Ap1g1	Intracellular transport	X54424	17.6	0.04	No
Cell surface antigen AA4	AA4	Unknown	AF081789	15.8	0.04	No
P35b	P35b	Immune response	X53619	15.8	0.04	No
ClpP protease	ClpP	Protein modification	AJ005253	15.0	0.03	No
X chromosome-linked phosphoglycerate kinase	Pgk1	Metabolism	M15668	14.9	0.01	No
Elongation factor 2	Eef2	Biosynthesis	M76131	14.9	0.04	No
Myeloid cell leukemia sequence 1	Mcl1	Apoptosis	U35623	14.9	0.02	No
Ribosomal protein L8	Rpl8	Biosynthesis	U67771	14.8	0.004	No
Dynactin 1	Dctn1	Cytoskeletal element	U60312	14.4	0.04	No
Mitogen activated protein kinase kinase 5	Map2k5	Signal transduction	AB019374	14.2	0.01	No
Anti-corynebacterium kutscheri	Ack	Immune response	AF037260	13.5	0.04	No
Solute carrier 20, member 1	Slc20a1	Intercellular transport	M73696	13.3	0.02	No
Transaldolase	Taldol	Metabolism	U67611	13.3	0.01	No
N-myristoyltransferase 1	Nmt1	Protein modification	AF043326	13.1	0.05	No
Mini chromosome maintenance deficient 4	Mcmd4	Cell cycle	D26089	13.0	0.01	No
Seb4	Seb4l	Post-transcriptional reg.	X75316	12.8	0.04	No
Ferredoxin-NADP reductase	Fdxr	Metabolism	D49920	12.3	0.05	No
Annexin VI, p68	Anx6	Metabolism	X13460	11.9	0.05	No
Rab24	Rab24	Unknown	Z22819	11.8	0.04	No
Kryn	Kryn	Unknown	D89677	11.8	0.02	No
Histidyl-tRNA synthetase	Hars	Biosynthesis	U39473	11.8	0.01	No
SKD3	Skd3	Intercellular transport	U09874	11.0	0.003	No
Retinal S-antigen	Sag	Misc.	M24086	11.0	0.05	No
Thioredoxin	Txn1	Stress response	X77585	10.9	0.001	No
Sec22b	Sec221	Intracellular transport	U91538	10.3	0.02	No
Capping protein beta 1	Cappb1	Cytoskeletal element	U10407	9.9	0.02	No

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis (continued)						
Alcohol dehydrogenase-B2	Adhb2	Metabolism	M84147	9.5	0.01	No
Ubiquitin-like enzyme 1a	Uble1a	Protein modification	AB024303	9.5	0.01	No
Transcription factor S-II	SII	Transcription	M18209	9.3	0.01	No
Histone 1-0	H1fo	Chromatin-related	M29260	9.3	0.02	No
Cyclin F	Ccnf	Cell cycle	Z47766	9.0	0.04	No
Calcium/calmodulin serine protein kinase	Cask	Signal transduction	Y17138	8.9	0.04	No
Glutaredoxin	Glrx1	Metabolism	AB013137	8.8	0.01	No
WSB-1	Wsb	Unknown	AF033186	8.1	0.05	No
Neuropilin	Nrp	Cell adhesion	D50086	8.0	0.02	No
Procollagen, type XVIII, alpha 1	Col18a1	Cell adhesion	L22545	7.8	0.01	No
Ribosomal protein L29	Rpl29	Biosynthesis	X05021	7.6	0.03	No
CMP-N-acetylneuraminic acid synthetase	Cmas	Biosynthesis	AJ006215	7.5	0.04	No
Phosphatidylserine synthase-1	Ptdss1	Biosynthesis	AF042731	7.4	0.04	No
Chloride channel protein 3	Clcn3	Intracellular transport	AF029347	6.8	0.03	No
NfiX1-protein	Nfix	Transcription	Y07688	6.5	0.05	No
Lumican	Lum	Extracellular matrix	AF013262	6.4	0.05	No
Kinesin heavy chain member 2	Kif2	Intracellular transport	D12644	6.4	0.04	No
Glycerol-3-phosphate dehydrogenase	G3pd	Metabolism	D50430	6.3	0.02	No
Catechol-O-methyltransferase	Comt	Metabolism	AF076156	6.3	0.02	No
S100 calcium-binding protein A13	S100a13	Misc.	X99921	6.2	0.01	No
Cleft lip and palate transmembrane 1	Clptm1	Intracellular transport	D67067	6.1	0.003	No
Hemoglobin, beta adult major chain	Hbb	Misc.	J00413	6.1	0.01	No
Cdc6-related protein	Cdc6p	Cell cycle	AJ223087	5.8	0.03	No
TFIIH, 62 kD subunit	Gtf2h1	Transcription	AJ002366	5.7	0.01	No
H2A histone, member X	H2afx	Chromatin-related	Z35401	5.5	0.02	No
MHC class III region RD	RD	Unknown	AF109906	5.5	0.01	No
Transcription factor 21	Tf21	Transcription factor	AF035717	5.4	0.01	No
Golgi autoantigen subfamily a4	Golga4	Growth regulation	AF051357	5.4	0.04	No
Wingless-related MMTV integration site 6	Wnt6	Signal transduction	M89800	5.3	0.02	No
DNA polymerase alpha 1	Pola1	DNA replication	D13543	5.2	0.03	No
Polyubiquitin TI-225	Ti-225	Protein modification	D50527	4.9	0.004	No
Ribosomal protein S16	Rps16	Biosynthesis	M11408	4.9	0.002	No
ATP binding cassette D3	Abcd3	Biosynthesis	L28836	4.8	0.02	No
Unc-119	Unc119h	Misc.	AF030169	4.6	0.02	No
Acid beta glucosidase	Gba	Metabolism	M24119	4.5	0.02	No
Rp2	Rp2	Misc.	X04097	4.5	0.05	No

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis (continued)						
Programmed cell death 6	Pdcd6	Apoptosis	U49112	4.4	0.01	No
Discoidin domain receptor 1	Ddr1	Signal transduction	L57509	4.4	0.004	No
Complement component 1 inhibitor	C1inh	Misc.	AF010254	4.4	0.01	No
Defender against Apoptotic Death	Dad1	Apoptosis	U81052	4.3	0.01	No
Annexin 1	Anxa1	Growth regulation	M69260	4.2	0.001	No
Tubulin, beta	Tbb	Cytoskeletal element	X04663	4.2	0.02	No
Ribosomal protein S2	Rps2	Biosynthesis	M20632	4.2	0.01	No
Lantibiotic synthetase component C	Lanc11	Signal transduction	Y16518	4.2	0.04	No
Sialyltransferase 10	Siat10	Metabolism	D28941	4.1	0.003	No
Siva	Siva	Apoptosis	AF033115	4.1	0.01	No
Pur-alpha	Pura	Transcription	U02098	3.8	0.03	No
Phosphatidylinositol 4-phosphate 5-kinase I-beta	Pip5k1b	Signal transduction	D86177	3.8	0.02	No
Activin receptor (ActR IIB)	AcvrIIB	Signal transduction	M84120	3.8	0.01	No
Acidic ribosomal phosphoprotein PO	Arbp	Biosynthesis	X15267	3.8	0.03	No
Ndr1 related protein 3	Ndr3	Cellular differentiation	AB033922	3.7	0.03	No
Calpain small subunit	Capn4	Protein modification	AF058298	3.7	0.04	No
G protein beta 2	Gbeta2	Signal transduction	U34960	3.6	0.01	No
Cystatin B	Cstb	Apoptosis	U59807	3.5	0.05	No
CD1d1 antigen	Cd1d1	Immune response	M63695	3.4	0.02	No
Protein tyrosine phosphatase epsilon	Ptpe	Signal transduction	D83484	3.4	0.03	No
Ribosomal protein L28	Rpl28	Biosynthesis	X74856	3.4	0.001	No
Trans-golgi network protein 1	Tgn1	Intracellular transport	D50031	3.3	0.05	No
Mitochondrial ribosomal protein S31	Mrps31	Biosynthesis	Z46966	3.2	0.02	No
Poliovirus receptor homolog	Plvr	Signal transduction	D26107	3.2	0.04	No
Ndr1 related protein	Ndr2	Cellular differentiation	AB033921	3.1	0.05	No
Fibroblast growth factor inducible 13	Fgf13	Signal transduction	U42383	3.1	0.05	No
Dia	Dia	Cytoskeletal element	Y15910	3.1	0.05	No
Rab9	Rab9	Intracellular transport	AB027290	3.1	0.05	No
Sara	Sara	Intracellular transport	L20294	3.0	0.01	No
E46	E46	Unknown	X61506	2.9	0.02	No
Carnitine palmitoyltransferase 1	Cpt1	Metabolism	AF017175	2.9	0.01	No
Sin3B	Sin3b	Transcription	AF038848	2.9	0.03	No
Signal peptidase complex 18	Spc18	Intracellular transport	AB025405	2.9	0.02	No
Ribosomal protein S3	Rps3	Biosynthesis	X76772	2.9	0.01	No
Enoyl coenzyme A hydratase 1	Ech1	Metabolism	AF030343	2.8	0.01	No
DNA polymerase zeta catalytic subunit	Rev3l	DNA repair	AF083464	2.8	0.02	No

Name	Symbol	Function	GenBank Accession ID	Fold-difference in expression		Previously described role in reproduction? ^b
				Ratio	P-value ^a	
Genes with decreased expression as germ cells progress from mitosis to the onset of meiosis (continued)						
DEAD box polypeptide 6	Ddx6	RNA helicase	AF038995	2.7	0.01	No
Nucleolin	Ncl	Transcription	X07699	2.7	0.001	No
Ribosomal protein S8	Rps8	Biosynthesis	X73829	2.6	0.001	No
Metallothionein 1	Mt1	Stress response	V00835	2.6	0.05	No
CD98 antigen	Cd98	Immune response	AB017189	2.6	0.02	No
Actin, alpha 2, smooth muscle	Acta2	Cytoskeletal element	X13297	2.5	0.004	No
Calnexin	Canx	Protein modification	L18888	2.5	0.04	No
Ribosomal protein S6 kinase 3	Rps6ka2	Biosynthesis	AJ131021	2.4	0.03	No
JTB	Jtb	Misc.	AB016490	2.4	0.03	No
Bat-4	Bat4	Unknown	L76155	2.4	0.02	No
Ribosomal protein S4, X-linked	Rps4x	Biosynthesis	M73436	2.3	0.001	No
Cyclin T1	Ccnt1	Cell cycle	AF095640	2.1	0.02	No
Ribosomal protein L19	Rpl19	Biosynthesis	M62952	2.1	0.03	No
Procollagen, type VI, alpha 1	Col6a1	Cell adhesion	X66405	2.1	0.03	No
Lymphocyte antigen 6 locus E	Ly6e	Immune response	U47737	2.1	0.03	No
Golgi autoantigen a5	Golga5	Growth regulation	AB016784	2.0	0.03	No
Ornithine decarboxylase antizyme	Oaz1	Metabolism	U52823	2.0	0.005	No
Dystroglycan 1	Dag1	Cytoskeletal element	U43512	1.9	0.04	No
Antigen identified by monoclonal antibodies 4F2	4F2	Unknown	X14309	1.9	0.01	No
CD-1 cardiac troponin I	Tnni3	Cytoskeletal element	U09181	1.8	0.03	No

^aStudent's t-test p-values ≤ 0.05 indicate significant differences in expression across time points (see Materials and Methods in Chapter 4)

^bGenes with described roles in germ cell development or differentiation, meiosis, spermatogenesis, or fertilization

APPENDIX C

Unannotated sequences with significant expression differences during the first wave of spermatogenesis in the prepubertal mouse

Complete list of unannotated sequences with differential expression during spermatogenesis in the prepubertal mouse

Unannotated sequences showing differential expression between spermatogonial mitosis and preleptotene

GenBank Accession ID	Preleptotene / Spermatogonial Mitosis		GenBank Accession ID	Preleptotene / Spermatogonial Mitosis	
	Ratio ^a	P-value ^b		Ratio ^a	P-value ^b
C81612	185.0	0.04	AW125224	-1.9	0.05
AW123801	63.1	0.03	AI835883	-2.0	0.03
AI451558	45.4	0.05	AI847904	-2.0	0.01
C79052	36.6	0.04	AW124711	-2.0	0.05
AI451008	34.5	0.04	AI839770	-2.1	0.05
AI853703	26.8	0.02	AI463490	-2.1	0.03
AI837621	18.3	0.02	AI552528	-2.1	0.05
AW048038	18.1	0.001	AI854432	-2.2	0.05
AI324972	14.1	0.02	AW122284	-2.3	0.003
C80919	11.0	0.02	AA388099	-2.3	0.02
AI197161	10.3	0.01	AI842095	-2.3	0.01
AI183202	7.7	0.03	AI841629	-2.4	0.05
AW121930	5.6	0.04	AW061222	-2.4	0.02
AI843895	5.3	0.03	AI847269	-2.5	0.03
AF045953	5.1	0.02	AI854771	-2.5	0.03
AI846484	4.7	0.01	AW123032	-2.5	0.05
AW046194	3.9	0.01	AI854624	-2.5	0.05
AA986782	3.6	0.02	AI255450	-2.5	0.01
AA266298	3.0	0.0004	AA822413	-2.6	0.01
AI842066	2.8	0.04	AW049194	-2.6	0.03
AW125669	2.6	0.02	AI843895	-2.6	0.04
AI842066	2.6	0.05	AA760073	-2.6	0.04
C76063	2.5	0.004	AI843655	-2.7	0.04
AW048944	2.5	0.05	AA879764	-2.7	0.04
AI892206	2.4	0.01	AW123953	-2.7	0.04
AW047476	2.4	0.04	AW050353	-2.8	0.03
AA684456	2.4	0.01	AW122615	-2.8	0.04
AW124681	2.3	0.04	AI843655	-2.8	0.05
AW124735	2.1	0.02	AI852808	-2.8	0.04
C77386	2.0	0.04	AI851539	-2.9	0.04
AI842259	1.9	0.01	AI843682	-3.0	0.02
AI843417	-1.8	0.05	AI844853	-3.0	0.05
AW125739	-1.8	0.03	AI847904	-3.0	0.02
L29441	-1.8	0.03	AA895984	-3.1	0.02
AI647612	-1.8	0.05	AI851218	-3.1	0.05
AW121892	-1.8	0.02	AI851046	-3.1	0.03
AI462105	-1.9	0.02	AA919208	-3.2	0.01

^aFold difference in expression for preleptotene vs. spermatogonial mitosis; Negative numbers indicate decreased expression as germ cells differentiate

^bStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points

Unannotated sequences showing differential expression between spermatogonial mitosis and preleptotene (continued)

GenBank Accession ID	Preleptotene / Spermatogonial Mitosis		GenBank Accession ID	Preleptotene / Spermatogonial Mitosis	
	Ratio ^a	P-value ^b		Ratio ^a	P-value ^b
AI845182	-3.2	0.03	AA655199	-4.7	0.02
AW050086	-3.2	0.03	AA675468	-4.8	0.02
AA688938	-3.3	0.02	AW046443	-4.8	0.01
AI846628	-3.3	0.03	AI836686	-4.9	0.03
AW048038	-3.3	0.05	AA612483	-4.9	0.004
AW045665	-3.3	0.04	AI844011	-4.9	0.01
AI851237	-3.3	0.02	AW050305	-4.9	0.01
AI839988	-3.3	0.05	AI845882	-5.0	0.02
AI852340	-3.4	0.04	AA596710	-5.0	0.05
AW124044	-3.4	0.04	AI846127	-5.0	0.03
AI854043	-3.4	0.02	AA165759	-5.1	0.05
AW048272	-3.5	0.01	AI845580	-5.1	0.01
AW124187	-3.5	0.03	AW045481	-5.3	0.03
C88243	-3.7	0.01	AI842524	-5.4	0.02
AI852457	-3.7	0.04	AI851356	-5.5	0.001
AA967263	-3.8	0.02	AI853331	-5.5	0.04
AW227778	-3.8	0.03	AI843586	-5.6	0.02
AI840579	-3.8	0.01	AW125634	-5.6	0.004
AI843679	-3.9	0.02	AW121930	-5.7	0.04
AW046793	-3.9	0.01	AW122551	-5.7	0.004
AI840267	-3.9	0.04	AI850356	-5.7	0.05
AI843709	-3.9	0.002	AI851160	-5.9	0.01
AW061255	-3.9	0.05	AW050133	-5.9	0.01
AA709672	-4.0	0.01	AI846078	-5.9	0.02
AI843335	-4.0	0.01	AC002397	-6.2	0.002
AA122714	-4.1	0.03	AI838398	-6.4	0.05
AI853269	-4.1	0.02	AI854771	-6.4	0.01
AW047926	-4.1	0.02	AI849180	-6.4	0.05
AW122935	-4.2	0.01	AW124933	-6.4	0.01
AI851542	-4.2	0.05	AW046273	-6.4	0.002
AI848471	-4.3	0.04	AA615161	-6.5	0.05
AI851441	-4.3	0.01	AW210320	-6.6	0.02
AI841629	-4.4	0.02	AW121162	-6.6	0.03
AW124778	-4.4	0.02	AI853331	-6.9	0.04
AI846396	-4.4	0.02	AI414051	-7.2	0.001
AW123802	-4.4	0.04	AA823653	-7.3	0.03
AA693125	-4.4	0.03	AW124920	-7.4	0.01
AW048347	-4.4	0.04	AW125420	-7.4	0.04
AI849271	-4.4	0.004	AI846078	-7.5	0.05
AI153693	-4.5	0.01	AW124582	-7.5	0.02
AA285446	-4.5	0.05	AA879709	-8.0	0.01
AA755260	-4.6	0.04	AW045481	-8.0	0.03
AA690091	-4.6	0.02	AW045710	-8.1	0.01
AW124194	-4.6	0.01	AI847926	-8.1	0.05

^aFold difference in expression for preleptotene vs. spermatogonial mitosis; Negative numbers indicate decreased expression as germ cells differentiate

^bStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points

Unannotated sequences showing differential expression between spermatogonial mitosis and preleptotene (continued)

GenBank Accession ID	Preleptotene / Spermatogonial Mitosis		GenBank Accession ID	Preleptotene / Spermatogonial Mitosis	
	Ratio ^a	P-value ^b		Ratio ^a	P-value ^b
AW060819	-8.3	0.001	C80836	-16.6	0.02
AI844089	-8.6	0.03	AI152779	-16.6	0.04
AI839286	-8.7	0.001	AW212479	-16.8	0.01
AA796831	-8.8	0.05	AI839681	-18.1	0.03
AW121136	-8.8	0.02	AW122530	-18.1	0.03
AI838398	-8.9	0.01	AI845581	-18.2	0.002
L29441	-8.9	0.04	AW124401	-19.6	0.01
AI838053	-9.0	0.01	AA866668	-19.9	0.01
AW259199	-9.1	0.04	AW050287	-21.3	0.04
AI838915	-9.2	0.01	AI647493	-22.6	0.05
AW121767	-9.2	0.01	AW125224	-24.1	0.04
AI788543	-9.4	0.02	U38981	-24.6	0.01
AI845593	-9.6	0.05	AI835592	-27.2	0.02
AI847314	-10.0	0.003	AA615100	-27.2	0.002
AI847564	-10.2	0.05	AW120691	-30.6	0.04
AV380793	-10.3	0.04	AI853173	-31.9	0.04
AA517835	-10.8	0.02	AI115399	-34.0	0.01
AI843709	-10.8	0.05	AW125649	-34.4	0.04
AA657164	-11.1	0.04	AW227345	-34.8	0.01
AI845165	-11.3	0.02	AI843884	-35.7	0.01
AI006319	-11.5	0.04	AW060257	-37.1	0.04
AW123952	-11.6	0.01	AA285446	-37.7	0.05
AW123801	-11.7	0.05	AW060324	-38.3	0.02
AA388099	-12.5	0.05	AA693246	-39.3	0.01
AW123880	-12.6	0.04	AI846708	-41.3	0.04
AI644179	-12.8	0.02	AW046496	-43.5	0.03
AI747444	-12.8	0.05	AW049122	-44.0	0.003
AA624336	-13.1	0.01	AW120683	-46.1	0.0004
AI849620	-13.3	0.03	AW046038	-59.5	0.05
AA657164	-13.4	0.04	AW120986	-60.0	0.05
AA940430	-13.5	0.03	AI843074	-67.7	0.02
AW120557	-13.6	0.001	AW061222	-75.1	0.003
AW047012	-14.2	0.01	AI846773	-82.4	0.03
AW213777	-14.6	0.05	AW124874	-104.4	0.05
AW124735	-14.7	0.01	AI847972	-107.3	0.02
AI849280	-14.7	0.03	AI844737	-108.8	0.04
AI842095	-15.4	0.01	AI843401	-132.2	0.02
AI060709	-15.7	0.03	AW120739	-132.9	0.04
AW125880	-16.1	0.05	AI843417	-146.5	0.004
AI847056	-16.2	0.03			

^aFold difference in expression for preleptotene vs. spermatogonial mitosis; Negative numbers indicate decreased expression as germ cells differentiate

^bStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points

Unannotated sequences showing differential expression between preleptotene and pachytene

GenBank Accession ID	Pachytene / Preleptotene		GenBank Accession ID	Pachytene / Preleptotene	
	Ratio ^a	P-value ^b		Ratio ^a	P-value ^b
AI852571	34.8	0.04	AW125043	-5.4	0.03
AW124239	7.0	0.04	AI006228	-6.7	0.05
AW060971	6.5	0.04	AI850793	-8.7	0.04
AI842544	6.2	0.04	AW046661	-11.4	0.05
AI120844	5.9	0.04	AI851220	-12.1	0.01
AA921481	5.0	0.01	AI195392	-14.8	0.05
AW120890	4.5	0.002	AA763918	-24.5	0.03
AW049598	3.3	0.05	AI891475	-30.7	0.05
AW049513	2.5	0.01			
AI847766	2.5	0.02			
AA989957	2.5	0.03			
AI594427	2.4	0.002			
AW124128	2.2	0.02			
AA839465	2.1	0.04			
AI413179	1.9	0.004			
AI842603	1.9	0.05			
AW124656	-1.8	0.03			
AI843396	-1.8	0.02			
AI851081	-1.9	0.01			
AW120746	-2.0	0.03			
AI839690	-2.0	0.05			
AI843802	-2.1	0.01			
AI838859	-2.2	0.04			
AI837107	-2.3	0.01			
AA981154	-2.3	0.02			
AI226368	-2.3	0.04			
AI836182	-2.4	0.01			
AI844939	-2.5	0.04			
AU044050	-2.5	0.001			
AA266467	-2.6	0.01			
AI848623	-2.9	0.01			
AW125218	-2.9	0.005			
AA866768	-2.9	0.01			
AI848671	-3.6	0.002			
AW124926	-4.2	0.04			
AI852665	-5.0	0.03			
AA409481	-5.2	0.04			

^aFold difference in expression for pachytene vs. preleptotene; Negative numbers indicate decreased expression as germ cells differentiate

^bStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points

Unannotated sequences with differential expression between spermatogonial mitosis and preleptotene and also between preleptotene and pachytene

GenBank Accession ID	Preleptotene / Spermatogonial Mitosis		Pachytene / Preleptotene		Expression Pattern ^c
	Ratio ^a	P-value ^b	Ratio ^a	P-value ^b	
AI848888	6.4	0.02	-5.0	0.02	ID
AW047237	3.7	0.04	-8.5	0.04	ID
AI854214	-1.8	0.05	-2.6	0.05	DD
AI854144	-2.4	0.04	-1.8	0.01	DD
AI837369	-2.4	0.02	2.1	0.03	DI
AI836082	-2.9	0.03	-2.2	0.02	DD
AI849082	-3.0	0.02	3.0	0.01	DI
AW124044	-3.0	0.02	2.5	0.03	DI
AI848968	-3.1	0.005	2.3	0.01	DI
AW122609	-3.4	0.01	-3.2	0.04	DD
AW049326	-3.4	0.001	2.2	0.05	DI
AI847054	-3.7	0.04	-4.2	0.02	DD
AW125431	-4.1	0.03	-1.8	0.03	DD
AW049897	-4.5	0.02	4.8	0.01	DI
AI415065	-4.8	0.001	1.8	0.01	DI
AI840458	-5.2	0.03	2.3	0.05	DI
AI850953	-5.5	0.03	1.9	0.02	DI
AW060843	-6.2	0.04	1.7	0.05	DI
AI843650	-6.7	0.01	-2.8	0.02	DD
AI844549	-6.8	0.04	2.2	0.05	DI
AI837302	-13.3	0.04	-3.0	0.01	DD
AW047625	-20.60	0.01	3.2	0.04	DI
AI467390	-23.50	0.02	-3.8	0.03	DD
AW213883	-28.40	0.002	-2.2	0.05	DD
AA794350	-39.60	0.02	9.6	0.001	DI
AW046708	-46.50	0.002	5.8	0.01	DI
AI836446	-81.20	0.001	2.4	0.03	DI

^aFold difference in expression for preleptotene vs. spermatogonial mitosis and pachytene vs. preleptotene; Negative numbers indicate decreased expression as germ cells differentiate

^bStudent's t-test p-values \leq 0.05 indicate significant differences in expression across time points

^cDD = Decreased expression from spermatogonial mitosis to preleptotene and decreased expression from preleptotene to pachytene; DI = Decreased expression from spermatogonial mitosis to preleptotene and increased expression from preleptotene to pachytene; ID = Increased expression from spermatogonial mitosis to preleptotene and decreased expression from preleptotene to pachytene

APPENDIX D

Functional genomics-related websites

Websites utilized for functional genomics-related research

Website Name	URL	Website Description
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/	Public databases for biological research
PubMed	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi	Searchable scientific publication database
Entrez Nucleotide Query	http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.html	Searchable nucleotide sequence database
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/	Searchable non-redundant gene cluster database
Basic Local Alignment Search Tool (BLAST)	http://www.ncbi.nlm.nih.gov/BLAST/	Compares DNA and/or protein sequences
CloneRanger	https://www.resgen.com/resources/apps/cloneranger/index.php3	Searchable cDNA clone database
I.M.A.G.E. Consortium at LLNL	http://image.llnl.gov/	Searchable cDNA clone database
Large-Scale Gene Expression & Microarray Resources	http://industry.ebi.ac.uk/~alan/MicroArray/	Non-commercial microarray-related resources
Affymetrix	http://www.affymetrix.com/index.affx	Commercial GeneChip resource
Mouse Genome Informatics: Gene Expression	http://www.informatics.jax.org/menus/expression_menu.shtml	Searchable gene expression databases
Human Gene Nomenclature Database	http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl	Searchable database for obtaining gene symbols
Spidey	http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/	Aligns mRNA and genomic sequences
BLAST 2 Sequences	http://www.ncbi.nlm.nih.gov/gorf/bl2.html	Aligns 2 user-selected nucleotide sequences against each other
CLUSFAVOR	http://mbr.bcm.tmc.edu/genepi/home.html	Unsupervised hierarchical cluster and principal component analyses
MicroArray Explorer	http://www.lecb.ncifcrf.gov/MAExplorer/	Analyzes the expression of individual genes and gene families; Compares expression patterns; Provides access to other genomic databases
Gene Microarray Pathway Profiler (GenMAPP)	http://www.genMAPP.org/	Maps gene expression data to biological pathways
Gene Ontology Project	http://www.geneontology.org	Searchable database of molecular functions, biological processes, and cellular components that are similar across organisms
Genomatix PromoterInspector	http://www.genomatix.de/cgi-bin/promoterinspector/promoterinspector.pl	Predicts promoter regions in mammalian genomic sequences
Methods for Affymetrix Oligonucleotide Arrays	http://biosun01.biostat.jhsph.edu/~ririzarr/Raffy	Interactive Affymetrix oligonucleotide array data analysis
National Library of Medicine Gateway Search	http://gateway.nlm.nih.gov/gw/Command	Simultaneously searches multiple databases at the U.S. National Library of Medicine

APPENDIX E

Curriculum Vitae

Lisa Tomascik-Cheeseman

5381 Moonflower Way

Livermore, CA 94550

(925)456-5354

Education

West Virginia University, Morgantown, WV

Graduate student in the interdisciplinary program of Genetics and Developmental Biology
(Aug. 1995 - Present)

Passed written and oral comprehensive qualifying examinations for Ph.D. degree candidacy
(Sept. 1997)

Ph.D. dissertation research in the Biology and Biotechnology Research Program at LLNL in the
laboratory of Dr. A.J. Wyrobek (Dec. 1997 - Present)

Delaware Valley College of Science and Agriculture, Doylestown, PA

B.S. in Biology with Chemistry minor (May 1995)

Work Experience

7/98–Present: **Life Sciences Scholar**, Genetic Damage and Disease Team, Biology and
Biotechnology Research Program, Lawrence Livermore National
Laboratory, Livermore, CA.

Characterize the baseline expression of genes essential for DNA repair, stress
response, cell cycle regulation, apoptosis, signal transduction and chromosome
segregation in various mouse and human tissues using cDNA microarrays.

Develop a mouse model to investigate the molecular mechanisms of meiosis and
the genetic causes of male infertility.

Determine the temporal expression of DNA repair, chromosome segregation and
recombination-associated genes during the first wave of mouse spermatogenesis.

Develop and optimize cDNA microarray protocols for RNA isolation, probe
labeling and hybridization in order to maximize hybridization specificity and
sensitivity.

Optimize microarray slide chemistries to maximize binding of target cDNA and
to minimize background fluorescence during image analysis.

12/97-7/98: **Participating Guest**, Biology and Biotechnology Research Program, Genetic Damage and Disease Team, Lawrence Livermore National Laboratory, Livermore, CA.

Conducted research on the molecular mechanisms of paternally transmitted genetic disease.

Investigated the role of centromere protein B (CENP-B) in male meiotic chromosome segregation using multicolor fluorescence *in situ* hybridization (FISH).

Developed and applied several multicolor FISH assays to assess the frequency of autosomal and sex-chromosomal non-disjunction in sperm from a father of multiple aneuploid pregnancies.

Used immunochemistry to identify protamine expression in specific cell types in the mouse testis.

Performed small animal surgery.

8/95-5/97: **Graduate Teaching Assistant**, Department of Biology, West Virginia University, Morgantown, WV.

Taught 4-5 undergraduate general biology laboratory classes per semester.

Developed lectures to complement every laboratory exercise.

Supervised independent student research projects.

Promoted critical thinking in a cooperative learning environment.

Evaluated student performance based on independent research and examinations.

Research Articles

Tomascik-Cheeseman LM, Coleman MA, Nelson DO, Kegelmeyer LM, Marchetti F, Nath J and AJ Wyrobek. 2002. Differential basal expression of genes associated with stress response, damage control and DNA repair among mouse tissues. *Molecular Biology of the Cell*. Submitted.

Tomascik-Cheeseman LM, Marchetti F, Lowe X, Shamanski F, Nath J and AJ Wyrobek. 2002. CENP-B is not critical for meiotic chromosome segregation in male mice. *Mutation Research* (513):197-203.

Kegelmeyer LM, **Tomascik-Cheeseman L**, Burnett MS, VanHummelen P and AJ Wyrobek. 2001. A groundtruth approach to accurate quantitation of fluorescence microarrays. *Microarrays: Optical Technologies and Informatics, Proceedings of SPIE* (4266):35-45.

Meeting Presentations and Published Abstracts

Tomascik-Cheeseman LM, Marchetti F, Coleman MA, Yin E, Nath J and AJ Wyrobek. 2002. Stage-dependent transcriptional profiling of male germ cells. *Environmental and Molecular Mutagenesis* (39)A203.

Velazquez-Wong AC, Slotter E, **Tomascik-Cheeseman LM**, Araujo-Solis MA, Wyrobek AJ and F Salamanca-Gomez. 2002. Structural and numerical chromosomal abnormalities in sperm following accidental occupational exposure to ¹⁹²Iridium. *Environmental and Molecular Mutagenesis* (39)A210.

Tomascik-Cheeseman LM, Marchetti F, Coleman MA, Yin E, Rankin K, Nath J and AJ Wyrobek. 2001. Microarray transcriptional profiling of mammalian male meiosis. *American Journal of Human Genetics*, 69 (4):953.

Tomascik-Cheeseman LM, Coleman MA, Marchetti F, Mascio-Kegelmeyer L, Nelson D, Nath J and AJ Wyrobek. 2001. Baseline expression of DNA repair-related genes differs among healthy mouse tissues. *Environmental and Molecular Mutagenesis* (37)A225.

Yin E, Coleman MA, **Tomascik-Cheeseman L**, Mascio-Kegelmeyer L, Wyrobek AJ. 2001. Radiation-induced changes in the expression of stress response and DNA repair genes in human and mouse cells. *Proceedings of the American Association for Cancer Research* (42)A713.

Tomascik-Cheeseman LM, Coleman MA, Kegelmeyer LM, Marchetti F, Raja R, Nath J and AJ Wyrobek. 2000. cDNA microarray analyses reveal differential expression of DNA-repair related genes among normal tissues. *American Journal of Human Genetics*, 67 (4):1466.

Coleman MA, Yin E, Sorensen K, Marsh BJ, Mabery S, **Tomascik-Cheeseman LM**, Liu S, Gregg J, Kegelmeyer LM, Attix T, Tucker J and AJ Wyrobek. 2000. Radiation-induced changes in the expression of stress response and DNA repair genes in human and mouse cells. *American Journal of Human Genetics*, 67 (4):1019.

Tomascik-Cheeseman LM, Raja R, Mabery S, Marsh B, Marchetti F, Nath J and AJ Wyrobek. 2000. Parallel expression analyses of DNA repair genes in adult mouse tissues using cDNA microarrays. *Environmental and Molecular Mutagenesis*, 35 (31):212.

Raja RG, **Tomascik-Cheeseman LM**, Pera RR and AJ Wyrobek. 2000. Representational Difference Analysis (RDA) of DNA repair genes using cDNA microarrays. Cambridge Healthtech Institute's Fourth Annual Genome 2000 Tri-Conference: Gene Functional Analysis, San Francisco, CA.

Raja RG, Kegelmeyer LM, **Tomascik-Cheeseman LM**, VanHummelen P, Marchetti F and AJ Wyrobek. 1999. Quantitation of cDNA expression microarrays for DNA repair and damage response genes. *Proceedings of The Microarray Meeting: Technology, Application and Analysis, Nature Genetics*, 70.

Tomascik-Cheeseman LM, Raja RG, Lowe X, Wilson D, Nath J and AJ Wyrobek. 1999. Parallel expression analyses of DNA repair genes in adult mouse brain and testis using cDNA microarrays. *American Journal of Human Genetics*, 65 (4):421.

Tomascik-Cheeseman LM, Lowe X, Kidd S, Eskenazi B, Nath J and AJ Wyrobek. 1999. Persistence of autosomal and sex-chromosomal disomy in a father of four aneuploid pregnancies. *Environmental and Molecular Mutagenesis*, 33 (30):216.

Honors

Young Investigator Achievement Award, Environmental Mutagen Society, 2002
Best Student/New Investigator Presentation, 32nd Annual Meeting of the Environmental Mutagen Society, 2001
Achievement Award, Biology & Biotechnology Research Program at Lawrence Livermore National Laboratory, 2001
Young Investigator Achievement Award, Environmental Mutagen Society, 2001
Student Travel Award, Environmental Mutagen Society, 2000
Student Travel Award, Environmental Mutagen Society, 1999
Cum laude graduate, Delaware Valley College of Science and Agriculture, 1995
Outstanding Senior Student in Biology, Delaware Valley College of Science and Agriculture, 1995
National Dean's List, 1994
Who's Who Among Students in American Colleges and Universities, 1994
National Student Government, 1994
Full Tuition Scholarship, Delaware Valley College of Science and Agriculture, 1991
William C. Dougherty National Scholarship, 1991

Activities

Student Committee, Environmental Mutagen Society
Education Award Subcommittee Chair, Environmental Mutagen Society
Special Interest Group on Animal Genetics/Aneuploidy/Germ Cells, Environmental Mutagen Society
Tri-Valley Animal Rescue Foster Parent
Lawrence Livermore National Laboratory S.T.O.M.P. (Scientists of Tomorrow: Oakland Mentoring Project) Mentor
Granada High School Supporters Club Scholarship Judge
Mervyn's Child Spree Volunteer

Professional Society Memberships

The American Society of Human Genetics
Environmental Mutagen Society
Genetic and Environmental Toxicology Association of Northern California

APPENDIX F

Published research articles