

A Toxicogenomic Study of the Hepatocarcinogen Furan

by

Anna Francina Webster (Jackson)

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biology

Carleton University
Ottawa, Ontario

© 2015, Anna Francina Webster

Abstract

A major goal in human health risk assessment is the identification and management of chemicals that may cause cancer in human populations. The current gold-standard for assessing chemical carcinogenicity is the two-year rodent cancer bioassay, which is animal, time, and resource intensive. The use of toxicogenomics for chemical risk assessment was first proposed over 15 years ago because of its potential to produce toxicologically relevant data more quickly, using fewer animals, and at a lower cost than the two-year cancer bioassay. While the two-year cancer bioassay produces a detailed inventory of chemical-dependent lesions, toxicogenomics analyzes chemical-dependent changes to global gene expression. Moreover, toxicogenomics provides comprehensive mechanistic data that are not obtained using standard tests. In this thesis quantitative, predictive, and mechanistic approaches were applied to a toxicogenomic case study of the rodent hepatocarcinogen furan. Female B3C6F1 mice were exposed for three weeks to non-carcinogenic or carcinogenic doses of furan. The dose response of a variety of transcriptional endpoints produced benchmark doses (BMDs) similar to the furan-dependent cancer BMDs. Bioinformatic analysis of disease datasets showed strong similarity between global gene expression changes induced by furan and those associated with the appropriate hepatic pathologies. The molecular pathways that were enriched in the liver following furan exposure facilitated the development of a molecular mode of action (MoA) for furan-induced liver cancer. Finally, transcriptional changes in formalin-fixed and paraffin embedded (FFPE) samples were compared to high quality frozen samples in order to evaluate whether archival samples are a viable option for toxicogenomic studies. The advantage of using FFPE tissues is that they are very well

characterized (phenotypically); the disadvantage is that formalin degrades biomacromolecules, including RNA. FFPE samples as old as three decades were shown to be feasible for toxicogenomics studies using a ribo-depletion RNA-seq protocol. Taken together, this case study demonstrates the utility of toxicogenomics data in human health risk assessment and the potential of archival FFPE tissue samples, and identifies viable strategies toward the reduction of animal usage in chemical testing.

Acknowledgements

I have had a wonderful 4.5 years at Carleton University and Health Canada. For this I must thank my supervisors, Iain Lambert and Carole Yauk. To Iain, for encouraging me to massively change projects (from bioremediation of contaminated soils, to toxicogenomics and FFPE genomics); and to Carole, for taking it from there. As I think is often the case, I entered this degree expecting to do one thing, and am leaving having done something completely different.

Carole, my accomplishments are largely a product of your support, encouragement, and imagination. Thank you for giving me the opportunity to be a part of the Health Canada Genomics Lab, what a wonderful team of people you have built. You are an exceptional leader, scientist, (and the life of the party). In addition to learning about genomics, from you I learned a great deal about being a good collaborator, communicator, and writer. These important skills will, I'm sure, be invaluable as I move forward in my career.

Iain, thank you for letting me turn your office into a thesis writing headquarters over these final few months of writing. I am sure that this process would have taken much longer had I not had you as my thesis writing chaperone. Thank you also for helping me 'put fog in a box' as I waded through my massive transcriptomic datasets. I sure am glad that I followed up on my email to you back in 2010!

To my OCIB buds! Thank you so much Julie Cox, Marc Beal, and Ali Long for making my last year so memorable! I am so proud of all of our work on the 12th Annual OCIB Symposium: The Next Generation. What a great experience. I don't have enough adjectives to describe you guys: creative, enthusiastic, dedicated, hardworking, persistent,

resourceful, and wonderful; well, that's a start. Also, thank you for being such great travel buddies over the years: King's Canyon and Sequoia National Parks ('Francina stop reading and look out the window!'), San Francisco ('I'm gonna git'chu....'), Monterey (Aquarium!), Orlando (Disney!), Miami (and the Bahamas for a weekend!), Brazil ('Aaaabacackzi'), and Philadelphia and Newark (Road trip!); so many adventures! To my other Environmental Health Center (EHC) and Carleton friends (especially Sarah Labib and Matt Meier), thank you so much for your help, advice, and friendship over the years.

I need to thank my colleagues at the EHC. Ivy Moffat for training me when I started; Andrew Williams for processing all of my raw data and making some great figures with my data; Remi Gagne for coordinating with Genome Quebec; and Byron Kuo for running my data through BMDEExpress. I feel particularly lucky to have worked with Julie Buick and Andrea Rowan-Caroll, who were always available to help me out or to have a chat. I also need to thank my American collaborators at ILS¹ in NC (especially Leslie Recio and Michael Waters) and at the Weill Cornell Medical College in NY (especially Paul Zumbo and Chris Mason). Les, thank you for coordinating the animal study that produced the samples that my entire project was based on. Mike, thank you for pulling me into the EMGS community by asking me to co-chair the Toxicogenomics Symposium in Orlando, and for asking me to contribute a chapter to your book. Finally, thank you to all of the members of the ILSI HESI² FFPE³ Genomics Working Group for all of your ideas and work on this project.

¹ Integrated Laboratory Systems, Inc. (www.ils-inc.com)

² International Life Sciences Institute, Health and Environmental Sciences Institute (www.hesiglobal.org)

³ Formalin-fixed and paraffin embedded

I would like to thank my committee members, Bruce McKay, Bill Willmore, and Paul White, for supporting and encouraging me. Further, I would like to acknowledge that this work was made possible by funds provided to me by NSERC ⁴, OGS ⁵, RAP ⁶, and Carleton University, and funds provided for the project by the Health Canada GRDI ⁷, ILS, HESI, and the NTP ⁸.

Finally, I need to thank my best friend and husband, Richard Webster. Thank you for keeping me fed and on track during this final dash. I am looking forward to getting back out onto the water with you!

Thank you everyone!

⁴ The Natural Sciences and Engineering Research Council of Canada (www.nserc-crsng.gc.ca)

⁵ The Ontario Graduate Scholarship

⁶ The Federal Government of Canada's Research Affiliate Program

⁷ Genomics Research and Development Initiative (grdi-irdg.collaboration.gc.ca)

⁸ National Toxicology Program (<http://ntp.niehs.nih.gov/>)

Table of Contents (brief)

Chapter		Page
<u>1</u>	Introduction	2
<u>2</u>	Case Study on the Utility of Hepatic Global Gene Expression Profiling in the Risk Assessment of the Carcinogen Furan	31
<u>3</u>	Bromodeoxyuridine treatment to measure Hepatocellular Proliferation does not Mask Chemically Induced Gene Expression Changes in Mouse Liver	75
<u>4</u>	Impact of Platform and Statistical Filtering on Transcriptional Benchmark Doses and Multiple Approaches for Selection of Point of Departure	94
<u>5</u>	Preparation of Archival FFPE Mouse Samples for use with the Agilent Gene Expression Microarray Platform	127
<u>6</u>	Mining the Archives: A Cross-Platform Analysis of Gene Expression Profiles in Archival Formalin-Fixed Paraffin-Embedded Tissue	155
<u>7</u>	Final Discussion	187
Appendix		
<u>A</u>	Gene Expression Analysis of Livers from Female B6C3F1 Mice Exposed to Carcinogenic and Noncarcinogenic Doses of Furan, with or without Five Days Bromodeoxyuridine Treatment	238
<u>B</u>	Differential Expression of Long Non-Coding RNAs in the Livers of Female B6C3F1 Mice Exposed to the Carcinogen Furan	251
<u>C</u>	Supplemental Material	278

Table of Contents (full)

ABSTRACT.....	II
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS (BRIEF).....	VII
TABLE OF CONTENTS (FULL)	VIII
LIST OF TABLES	XVI
LIST OF FIGURES	XVIII
LIST OF APPENDICES	XXI
LIST OF ABBREVIATIONS	XXII
STATEMENT OF CONTRIBUTIONS.....	XXV
1 CHAPTER: INTRODUCTION.....	2
1.1 Chemical Testing and Toxicogenomics	2
1.2 Reducing Animal Use in Toxicity Testing.....	5
1.3 Formalin-Fixed, Paraffin-Embedded (FFPE) Archival Tissue.....	10
1.4 The Liver	12
1.4.1 Liver Physiology	12

1.4.2	Liver Regeneration.....	13
1.4.3	Liver Cancer.....	14
1.5	Test Article: Furan	16
1.6	Thesis Overview.....	21
1.6.1	Rationale	21
1.6.2	Chapter Objectives.....	22
1.7	Tables and Figures.....	26
2	CHAPTER: CASE STUDY ON THE UTILITY OF HEPATIC GLOBAL GENE EXPRESSION PROFILING IN THE RISK ASSESSMENT OF THE CARCINOGEN FURAN.....	31
2.1	Preamble.....	31
2.2	Abstract	32
2.3	Introduction.....	33
2.4	Materials and Methods.....	33
2.4.1	Chemical	33
2.4.2	Animals	34
2.4.3	Immunohistochemistry for BrdU Nuclear Staining.....	35
2.4.4	RNA Extraction.....	37
2.4.5	Microarray.....	37
2.4.6	Bioinformatics.....	38
2.4.6.1	Ingenuity Pathway Analysis (IPA)	38
2.4.6.2	Disease Prediction and Chemical Profile Comparison	39
2.4.6.3	PCA in R.....	39

2.4.6.4	Cluster Analysis.....	40
2.4.7	Benchmark Dose (BMD) Calculations.....	40
2.4.8	Margin of Exposure.....	42
2.5	Results.....	43
2.5.1	Cellular Proliferation (BrdU), Cytotoxicity, and Cancer.....	43
2.5.2	Gene Expression Analysis.....	43
2.5.3	Cluster Analysis	44
2.5.4	Pathways and Upstream Regulators of Transcription Analyses	45
2.5.5	Disease and Chemical Prediction	46
2.5.6	GO Terms.....	47
2.5.7	Margin of Exposure.....	47
2.6	Discussion	47
2.6.1	Furan MoA: Oxidative Stress.....	50
2.6.2	Furan MoA: Cytotoxicity	52
2.6.3	Furan MoA: Inflammation	55
2.6.4	Furan MoA: Regenerative Proliferation.....	57
2.6.5	Furan MoA: Summary	60
2.6.6	Margin of Exposure.....	61
2.6.7	Conclusions	62
2.7	Tables and Figures.....	64
3	CHAPTER: BROMODEOXYURIDINE TREATMENT TO MEASURE HEPATOCELLULAR PROLIFERATION DOES NOT MASK CHEMICALLY INDUCED GENE EXPRESSION CHANGES IN MOUSE LIVER	75
3.1	Preamble.....	75

3.2	Abstract	76
3.3	Introduction.....	77
3.4	Materials and Methods.....	79
3.4.1	Animals	79
3.4.2	RNA Extraction and Microarray	79
3.4.3	Identification of Unique and Shared Probes.....	80
3.4.4	Bioinformatic Analysis	80
3.5	Results and Discussion.....	81
3.5.1	Effects of BrdU Treatment on Gene Expression (within dose groups)	82
3.5.2	Detection of Furan-Induced Gene Expression Changes in the Presence of BrdU.....	82
3.5.3	Detection of Furan’s Molecular MoA	84
3.6	Conclusions.....	87
3.7	Tables and Figures.....	89
4	CHAPTER: IMPACT OF PLATFORM AND STATISTICAL FILTERING ON TRANSCRIPTIONAL BENCHMARK DOSES AND MULTIPLE APPROACHES FOR SELECTION OF POINT OF DEPARTURE	94
4.1	Preamble.....	94
4.2	Abstract	95
4.3	Introduction.....	96
4.4	Materials and Methods.....	98
4.4.1	Animals and Exposures	98
4.4.2	RNA Extraction.....	99

4.4.3	RNA-Sequencing	99
4.4.4	Microarray	100
4.4.5	PCR Array	100
4.4.6	BMD Modeling	101
4.5	Results	103
4.5.1	Inter-technology Comparison of Differentially Expressed Genes	103
4.5.2	Effect of Statistical Filtering of Gene Expression Data Prior to BMD Modeling	103
4.5.3	Inter-Technology Comparison of BMD Values	105
4.5.4	Approaches for Deriving Transcriptional Point of Departure (PoD) Values	106
4.6	Discussion	107
4.6.1	Inter-Platform Comparison	107
4.6.2	Statistical Filtering	109
4.6.3	Point of Departure Selection	111
4.6.4	Conclusions	115
4.7	Tables and Figures	116
5	CHAPTER: PREPARATION OF ARCHIVAL FORMALIN-FIXED AND PARAFFIN-EMBEDDED MOUSE LIVER SAMPLES FOR USE WITH THE AGILENT GENE EXPRESSION MICROARRAY PLATFORM.....	127
5.1	Preamble	127
5.2	Abstract	128
5.3	Introduction.....	129
5.4	Materials and Methods.....	132
5.4.1	Animals and Tissues.....	132

5.4.2	RNA Extraction.....	133
5.4.3	cDNA Synthesis	135
5.4.4	qPCR	135
5.4.5	Microarray.....	136
5.4.6	Microarray Statistical Analysis	137
5.5	Results.....	138
5.6	Discussion	141
5.7	Tables and Figures.....	145
6	CHAPTER: MINING THE ARCHIVES: A CROSS-PLATFORM ANALYSIS OF GENE EXPRESSION PROFILES IN ARCHIVAL FORMALIN-FIXED PARAFFIN-EMBEDDED TISSUE	155
6.1	Preamble.....	155
6.2	Abstract	156
6.3	Introduction.....	157
6.4	Methods	158
6.4.1	Frozen and FFPE Tissue Samples	158
6.4.2	RNA Extraction and cDNA Synthesis	159
6.4.3	Microarray Analysis	160
6.4.4	Microarray Statistical Analysis	162
6.4.5	Microarray Bioinformatic Analysis.....	163
6.4.6	RNA-Sequencing Methods.....	164
6.4.7	RNA-Sequencing Bioinformatic Analysis	165
6.4.8	Online Data Access	167

6.5	Results	167
6.5.1	Gene Expression Profiles	167
6.5.2	Pathway Analysis	169
6.5.3	Meta-Analysis of Chemicals and Diseases.....	170
6.5.4	RNA-Sequencing Sample Profiles	171
6.5.5	Long-Term Archival Samples	173
6.6	Discussion	174
6.7	Tables and Figures	180
7	CHAPTER: FINAL DISCUSSION	187
7.1	Summary of Study Outcomes	187
7.2	General Conclusions and Contributions to Scientific Knowledge	192
7.2.1	Furan Case Study: Major Conclusions	192
7.2.2	Toxicogenomics: A Cost-Benefit Analysis	195
7.3	How Toxicogenomics Fits into the Current Toxicity Testing Paradigm	196
7.3.1	Tiered Testing	196
7.3.2	Guidelines for using Toxicogenomics Data in Formal Risk Assessment.....	200
7.4	Future Directions	203
7.4.1	FFPE Genomics	203
7.4.2	An Adverse Outcome Pathway for Cyp2E1 Substrates	205
7.5	Concluding Remarks	207
7.6	Figures	208
	REFERENCES	210

APPENDICES 237

**APPENDIX A GENE EXPRESSION ANALYSIS OF LIVERS FROM FEMALE B6C3F1 MICE
EXPOSED TO CARCINOGENIC AND NONCARCINOGENIC DOSES OF FURAN, WITH OR
WITHOUT 5 DAYS BROMODEOXYURIDINE (BRDU) TREATMENT.....238**

A.1	Abstract	239
A.2	Specifications	240
A.3	Publically Available Data.....	241
A.4	Experimental Design, Materials and Methods	241
A.5	Discussion	246
A.6	Tables and Figures	247

**APPENDIX B DIFFERENTIAL EXPRESSION OF LONG NON-CODING RNAS IN THE LIVERS
OF FEMALE B6C3F1 MICE EXPOSED TO THE CARCINOGEN FURAN.251**

B.1	Abstract	252
B.2	Introduction	253
B.3	Materials and Methods	255
B.4	Results	258
B.5	Discussion	260
B.6	Tables and Figures	267

APPENDIX C SUPPLEMENTAL MATERIAL.....278

C.1	Chapter 2	278
C.2	Chapter 3	306
C.3	Chapter 4	310
C.4	Chapter 5	314
C.5	Chapter 6	330

List of Tables

Table		Page
1.1	The results of the short-term genotoxicity testing of furan.	26
1.2	Incidence of furan-dependent liver cancer in female B6C3F1 mice.	27
2.1	BMD modeling of previously published apical endpoint data for furan exposures in B6C3F1 mice.	64
2.2	Genes that are differentially expressed in response to 8 mkd furan and relevant to the furan mode of action.	65
2.3	Margin of exposure values for furan-induced molecular pathways and cancer endpoints.	66
3.1	Gene expression values of probes differentially expressed by BrdU in control and furan-treated mice.	89
3.2	Number of differentially expressed probes (Agilent probe IDs) or genes (unique GenBank accession numbers) for each furan dose group, \pm BrdU treatment.	90
4.1	Gene and cancer dose response.	116
4.2	Statistics for box-and-whisker plots (Fig. 4.3).	117
5.1	RNA extraction from FFPE tissues.	145
5.2	Select qPCR and microarray gene expression values for phenobarbital-responsive genes.	146
5.3	Top 10 enriched molecular pathways.	149
6.1	Number of differentially expressed genes (by unique gene symbol) in each microarray and RNA-seq experiment using different filtering thresholds.	180
6.2	Chemical Signature Analysis.	181
6.3	Disease Signature Analysis.	182
B1	Differentially Expressed lncRNAs at 4 and 8 mkd furan.	267

C1	List of relevant correlated disease as determined by a metanalysis of published gene expression studies using the NextBio Human Disease Atlas.	287
C2	Full list of correlated compounds as determined by a metanalysis of published gene expression studies using NextBio.	290
C3	Full list of enriched GO terms.	303
C4	Enriched IPA Signaling Pathways for the high dose groups (\pm BrdU).	306
C5	qPCR primers.	314
C6	QC metrics of the one-colour arrays.	316
C7	QC metrics of the two-colour arrays	316
C8	RNA quality and yield.	330

List of Figures

Figure		Page
1.1	Three approaches for analyzing toxicogenomics data.	28
1.2	Experimental design of this thesis.	29
2.1	Dose response furan of furan-induced effects.	70
2.2	Cluster analysis comparing furan-dependent changes in gene expression data with publically available datasets.	71
2.3	Furan mode of action.	72
2.4	Adaptive and adverse outcomes in response to non-carcinogenic and carcinogenic doses of furan.	73
3.1	Hierarchical clustering of global gene expression profiles from each dose group (\pm BrdU treatment).	91
3.2	Hepatic ToxFunction enrichment analysis.	92
4.1	Gene expression levels are correlated between genomics platforms.	118
4.2	The intra-dose, inter-platform overlap of DEGs and pathways.	119
4.3	Filtering gene expression data significantly changes mean values for gene and pathway BMD(L)s.	120
4.4	Filtering gene expression data changes the distribution of pathway BMD-mean values.	121
4.5	Filtering gene expression data changes the distribution of pathway BMDL-mean values.	122
4.6	BMD values are not well correlated between genomics platforms.	123
4.7	BMD-means of furan MoA pathways are consistent across platforms.	124
4.8	Four approaches to deriving transcriptomic PoD values.	125
5.1	Qualitative descriptions of the four differentially expressed probe lists generated by each microarray protocol.	150

5.2	Comparison of the DEGs identified by each microarray protocol.	151
5.3	Literature validation of each protocol.	152
5.4	Schematic of experimental steps taken for analysis of gene expression in archival, slide-mounted FFPE tissue samples.	153
6.1	Differential gene expression.	183
6.2	Pathway analysis.	184
6.3	Examination of RNA-seq data quality metrics.	185
7.1	A summary of quantitative, predictive, and mechanistic toxicogenomics.	208
7.2	Adverse Outcome Pathway for Cyp2E1 Activation Leading to Liver Cancer.	209
A1	A representative MA plot.	247
A2	Background estimates of each array for reference and sample mean fluorescent signal of the 3xSLv1 negative control probe.	248
A3	Hierarchical clustering of all probes based on normalized signal intensity ratios.	249
A4	Summary of steps taken to generate, normalize and analyze two-colour Agilent gene expression microarray data.	250
B1	Chromosome distribution of differentially expressed lncRNAs based on microarray data in liver of furan-exposed mice.	272
B2	Chromosome localization of differentially expressed lncRNAs and nearby protein-coding genes based on microarray data in liver of furan-exposed mice.	273
B3	qPCR of lincRNA Chr. 9 78107225–78118850, <i>Gsta1</i> , and <i>Dppa5a</i>	274
B4	qPCR of lincRNA-p21 Chr. 9 78107225–78118850 and <i>Cdkn1a</i>	275
B5	The induction of a transcriptome response in furan-exposed mice induces the apparent bidirectional transcription of <i>Cdkn1a</i> and lincRNA-p21.	276
B6	qPCR of <i>Cdkn1a</i> and lincRNA-p21.	277

C1	Model counts and gene BMD/BMDL plots for each platform.	310
C2	Inter-platform comparisons of BMDLs for genes, pathway-means, and pathway medians for ANOVA filtered data.	311
C3	Inter-platform comparisons of BMDs and BMDLs for genes, pathway-means, and pathway medians for FDR filtered data.	312
C4	Inter-platform comparisons of BMDs and BMDLs for genes, pathway-means, and pathway medians for unfiltered data.	313
C5	Pathway analysis for ribo-depletion RNA-seq, polyA-enrichment RNA-seq, 1-colour microarrays, and 2-colour microarrays using all DEGs (i.e., unmapped gene list).	331

List of Appendices

Appendix A	Gene expression analysis of livers from female B6C3F1 mice exposed to carcinogenic and noncarcinogenic doses of furan, with or without 5 days Bromodeoxyuridine (BrdU) treatment.....	238
A.1	Abstract.....	239
A.2	Specifications.....	240
A.3	Publically Available Data.....	241
A.4	Experimental Design, Materials, and Methods.....	241
A.5	Discussion.....	246
A.6	Tables and Figures.....	247
Appendix B	Differential Expression of Long Non-Coding RNAs in the Livers of Female B6C3F1 Mice Exposed to the Carcinogen Furan.....	251
B.1	Abstract.....	252
B.2	Introduction.....	253
B.3	Materials and Methods.....	255
B.4	Results.....	258
B.5	Discussion.....	260
B.6	Tables and Figures.....	267
Appendix C	Supplemental Material.....	278
C.1	Chapter 2.....	278
C.2	Chapter 3.....	306
C.3	Chapter 4.....	310
C.4	Chapter 5.....	314
C.5	Chapter 6.....	330

List of Abbreviations

3Rs	Reduction, Refinement, and Replacement
AIC	Akaike's Information Criterion
AOP	Adverse Outcome Pathway
AP-1	Activating Protein 1
BDA	Cis-2-Butene-1,4-Dial
BEC	Biliary Epithelial Cells
BMD	Benchmark Dose
BMDL	Benchmark Dose Lower Confidence Limit
BMR	Benchmark Response
BMDS	Benchmark Dose Software
BrdU	Bromodeoxyuridine
CEPA	Canadian Environmental Protection Act
Cyp P450	Cytochrome P450
DAMP	Damage Associated Molecular Pattern
DEG	Differentially Expressed Gene
EURL- ECVAM	European Union Reference Laboratory for Alternatives to Animal Testing
ESRAB	Existing Substances Risk Assessment Bureau
ERK	Extracellular-Regulated MAPK
EPA	Environmental Protection Agency
FDR	False Discovery Rate
FFPE	Formalin-Fixed, Paraffin-Embedded
FRO	Frozen
GEO	Gene Expression Omnibus
GSH	Glutathione
HCA	Hepatocellular Adenoma
HCC	Hepatocellular Carcinoma
IARC	International Agency for Research on Cancer
ICATM	International Cooperation on Alternative Testing Methods

ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IHC	Immunohistochemistry
IPA	Ingenuity Pathway Analysis
IRIS	Integrated Risk Information System
IVIVE	<i>In vitro</i> to <i>in vivo</i> extrapolation
JaCVAM	Japanese Center for the Validation of Alternative Methods
JNK	c-Jun N-terminal Kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LincRNA	Long Intergenic Non-coding RNA
LncRNA	Long Non-coding RNA
LOWESS	Locally Weighted Scatterplot Smoothing
MAANOVA	Microarray Analysis of Variance
MAPK	Mitogen Activated Protein Kinase
MIE	Molecular Initiating Event
mkd	Milligram per Kilogram Bodyweight per Day
MoA	Mode of Action
MoE	Margin of Exposure
mRNA	Messenger RNA
miRNA	microRNA
NCBI	National Center for Biotechnology Information
NICEATM	The National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods
NF-kB	Nuclear Factor kappa B
NTP	National Toxicology Program
OECD	Organization for Economic Cooperation and Development
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PoD	Point of Departure
RNA-Seq	RNA Sequencing
ROS	Reactive Oxygen Species

RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SAPK	Stress-Activated MAPK
SEURAT-1	Safety Evaluation Ultimately Replacing Animal Testing 1
SOD	Superoxide Dismutase
qPCR	Quantitative Polymerase Chain Reaction
TNF	Tumour Necrosis Factor

Statement of Contributions

Chapter 2 Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan

Published in: *Toxicology and Applied Pharmacology* 274(1): 63-77. 1 January 2014.

Authors: A. Francina Jackson ^{a, b}, Andrew Williams ^a, Leslie Recio ^c, Michael D. Waters ^c, Iain B. Lambert ^b, Carole L. Yauk ^a

Affiliations:

^a Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^b Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

^c ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

DOI: 10.1016/j.taap.2013.10.019

Overall Data Interpretation and Manuscript Writing.....	Francina Webster
Experimental Design.....	Francina Webster Andrew Williams Leslie Recio Michael Waters Carole Yauk
Manuscript Preparation.....	Francina Webster Iain Lambert Carole Yauk
Animal Handling, Exposure, and Tissue Collection.....	Leslie Recio ILS staff
RNA Extraction.....	Leslie Recio ILS staff
Microarray.....	Francina Webster
Statistical Analysis.....	Andrew Williams
Hierarchical Clustering and PCA.....	Andrew Williams

Chapter 3 Bromodeoxyuridine (BrdU) treatment to measure hepatocellular proliferation does not mask furan-induced gene expression changes in mouse liver

Published in: *Toxicology* 323: 26-31. 2 September 2014.

Authors: A. Francina Jackson ^{a,b}, Andrew Williams ^a, Leslie Recio ^c, Carole L. Yauk ^a

Affiliations:

^a Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^b Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

^c ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

DOI: 10.1016/j.tox.2014.06.002

Overall Data Interpretation and Manuscript Writing.....	Francina Webster
Experimental Design.....	Francina Webster Andrew Williams Leslie Recio Carole Yauk
Manuscript Preparation.....	Francina Webster Carole Yauk
Animal Handling, Exposure, and Tissue Collection.....	Leslie Recio ILS staff
RNA Extraction.....	Leslie Recio ILS staff
Microarray.....	Francina Webster
Statistical Analysis.....	Andrew Williams

Chapter 4 Impact of Genomics Platform and Statistical Filtering on Transcriptional Benchmark Doses (BMD) and Multiple Approaches for Selection of Chemical Point of Departure (PoD)

Published in: *PLoS ONE* 10(8): e0136764. August 2015.

Authors: A. Francina Jackson ^{a,b}, Nikolai Chepelev ^a, Remi Gagne ^a, Byron Kuo ^a, Leslie Recio ^c, Andrew Williams ^a, Carole L. Yauk ^a

Affiliations:

^a Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^b Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

^c ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

DOI: 10.1371/journal.pone.0136764

Overall Data Interpretation and Manuscript Writing.....	Francina Webster
Experimental Design.....	Francina Webster Andrew Williams Remi Gagne Nikolai Chepelev Carole Yauk
Manuscript Preparation.....	Francina Webster Nikolai Chepelev Carole Yauk
Animal Handling, Exposure, and Tissue Collection.....	Leslie Recio ILS staff
RNA Extraction.....	Leslie Recio ILS staff
Microarray and qPCR array experiments.....	Francina Webster
Statistical Analysis.....	Andrew Williams
Box plots.....	Andrew Williams
Benchmark Dose Express.....	Byron Kuo
PolyA-enrichment RNA-seq.....	Remi Gagne Genome Quebec

Chapter 5 Preparation of archival FFPE mouse liver samples for use with the Agilent gene expression microarray platform

Published in: *Journal of Pharmacological and Toxicological Methods* 68(2): 260-8. September–October 2013.

Authors: A. Francina Jackson ^{a,b}, Andrew Williams ^a, Ivy Moffat ^a, Suzanne L. Phillips ^c, Leslie Recio ^c, Michael D. Waters ^c, Iain B. Lambert ^b, Carole L. Yauk ^a

Affiliations:

^a Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^b Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

^c ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

DOI: 10.1016/j.vascn.2013.02.008

Overall Data Interpretation and Manuscript Writing.....	Francina Webster
Experimental Design.....	Francina Webster Ivy Moffat Carole Yauk
Manuscript Preparation.....	Francina Webster Iain Lambert Carole Yauk
FRO and Freshly-Embedded FFPE Sample Preparation.....	Lorraine Casavant Mike Wade
Archival FFPE Sample Identification and Retrieval.....	Leslie Recio, Michael Waters, and ILS staff
RNA Extraction.....	Francina Webster
Microarray Experiments.....	Francina Webster
Statistical Analysis.....	Andrew Williams
Primer Design.....	Suzanne Phillips

Chapter 6 Mining the archives: a cross-platform analysis of gene expression profiles in archival formalin-fixed paraffin-embedded (FFPE) tissue.

Published in: *Toxicological Sciences*. ePub: September 10, 2015.

Authors: Anna Francina Webster ^{a,b}, Paul Zumbo ^c, Jennifer Fostel ^d, Jorge Gandara ^c, Susan D. Hester ^e, Leslie Recio ^f, Andrew Williams ^a, Charles E. Wood ^e, Carole L. Yauk ^a, Christopher E. Mason ^{c, g, h}

Affiliations:

^a Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^b Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa K1S 5B6, Canada

^c Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY 10065, USA

^d NIEHS, Research Triangle Park, NC 27709, USA

^e US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

^f ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

^g The Feil Family Brain and Mind Research Institute (BMRI), 413 East 69th St, New York, NY 10021, USA

^h The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, 1305 York Ave., New York, NY 10065, USA

DOI: 10.1093/toxsci/kfv195

Overall Data Interpretation and Manuscript Writing.....	Francina Webster
Experimental Design.....	HESI FFPE Genomics Working Group Christopher Mason Andrew Williams Francina Webster Carole Yauk Paul Zumbo
Manuscript Preparation.....	Francina Webster Paul Zumbo Charles Wood Susan Hester Christopher Mason Carole Yauk
Animal Handling, Exposure, and Tissue Collection.....	Leslie Recio ILS Staff

Identified and Retrieved Archival FFPE Tissues from the NTP.....	Jennifer Fostel
RNA Extraction.....	Leslie Recio (FRO) Francina Webster (FFPE)
Microarray Experiments.....	Francina Webster
Statistical Analysis of Microarray Experiments.....	Andrew Williams
PolyA-enrichment RNA-seq.....	Remi Gagne Genome Quebec
Ribo-depletion RNA-seq.....	Paul Zumbo Jorge Gandara Christopher Mason
Statistical Analysis of RNA-seq Data.....	Paul Zumbo Christopher Mason

Appendix A Gene expression analysis of livers from female B6C3F1 mice exposed to carcinogenic and non-carcinogenic doses of furan, with or without bromodeoxyuridine (BrdU) treatment

Published in: *Genomics Data* 2:117-122. December 2014.

Authors: A. Francina Jackson ^{a,b}, Andrew Williams ^a, Leslie Recio ^c, Carole L. Yauk ^a

Affiliations:

^a Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^b Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

^c ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

DOI: 10.1016/j.gdata.2014.05.013

Manuscript Writing.....	Francina Webster Andrew Williams
Experimental Design.....	Andrew Williams
Manuscript Preparation.....	Francina Webster Andrew Williams Carole Yauk
Animal Handling, Exposure, and Tissue Collection.....	Leslie Recio ILS staff
RNA Extraction.....	Leslie Recio ILS staff
Microarray.....	Francina Webster
Statistical Analysis.....	Andrew Williams

Appendix B Differential Expression of Long Noncoding RNAs in the Livers of Female B6C3F1 Mice Exposed to the Carcinogen Furan

Published in: *Toxicological Sciences* 135(2):369-72. October 2013.

Authors: Leslie Recio ^a, Suzanne L. Phillips ^a, Tim Maynor ^a, Michael D. Waters ^a, A. Francina Jackson ^{b,c}, Carole L. Yauk ^b

Affiliations:

^a ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA

^b Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada

^c Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

DOI: 10.1093/toxsci/kft153

Overall Data Interpretation and Manuscript Writing.....	Leslie Recio Suzanne Phillips
Experimental Design.....	Francina Webster Andrew Williams Leslie Recio Michael Waters Carole Yauk
Manuscript Preparation.....	Leslie Recio Suzanne Phillips Francina Webster Michael Waters Carole Yauk
Animal Handling, Exposure, and Tissue Collection.....	Leslie Recio Tim Maynor ILS staff
RNA Extraction.....	Leslie Recio Tim Maynor ILS staff
Microarray.....	Francina Webster
Statistical Analysis.....	Andrew Williams

CHAPTER 1

1 CHAPTER: INTRODUCTION

1.1 Chemical Testing and Toxicogenomics

A major goal in human health risk assessment is the identification and management of chemicals that may cause cancer in human populations. The current gold-standard for assessing chemical carcinogenicity is the two-year rodent cancer bioassay (OECD 2009), which is animal, time, and resource intensive. Since its inception by the National Toxicology Program (NTP; <http://ntp.niehs.nih.gov/>) in the 1970's, only ~1500 chemicals have been assessed. Each chemical evaluation costs on average \$2-4 million and uses 800-1000 rodents (Waters *et al.* 2010). The Canadian Environmental Protection Act (CEPA, 1999) requires that chemicals on Canada's domestic substances list (DSL) be assessed for potential human health risks. There are over 23,000 chemicals on the DSL. These chemicals were divided up into phases based on existing exposure and hazard information. The regulators have completed phase 1 and 2 assessment, and upcoming priorities include completing evaluations for approximately 1700 of these chemicals (Phase 3) by 2020. Chemicals are prioritized based on risk of exposure, risk of persistence or bioaccumulation, and estimated toxicity to humans and non-human species (CEPA section 73, 1999). However, one of the challenges associated with this goal is that the majority of these chemicals are data poor (e.g., have no conventional toxicological testing data associated with them). Other agencies face similar challenges, with the US Environmental Protection Agency (US EPA) and Europe's Registration, Evaluation, Authorization and Restriction (EU-REACH) listing 75,000 and 140,000 chemicals, respectively, on their lists of those requiring evaluation. It is clear that a more efficient system to assess the potentially harmful substances that are present in the environment is

necessary in order to address the massive backlog of as-of-yet untested compounds, as well as the large number of novel compounds that enter the market and environment each year.

The use of toxicogenomics—the measurement of chemically dependent changes in global gene expression—for chemical risk assessment was first proposed over 15 years ago (Nuwaysir *et al.* 1999) because of its potential to produce toxicologically relevant data more quickly, using fewer animals, and at a lower cost than the two-year cancer bioassay. While the two-year cancer bioassay produces a detailed inventory of chemical-dependent lesions, toxicogenomics is the analysis of chemical-dependent changes to global gene expression. Moreover, toxicogenomics provides comprehensive mechanistic data that are not obtained using standard tests. The toxicogenomics methodology assumes that prior to the occurrence of visible phenotypic changes, such as tumours, important changes on a molecular scale must occur. Further, it assumes that similar cancers share some common underlying biology at an early stage in the disease and that this biology is distinct from a generalized cellular toxic- or stress-response mechanism (Auerbach *et al.* 2010). There are three main toxicogenomics approaches: quantitative, predictive, and mechanistic (Fig. 1.1). The dose response of transcriptomic perturbations (i.e., quantitative toxicogenomics) can be modeled to quantify the benchmark dose (BMD), or the dose at which important biological processes depart from background levels (Yang *et al.* 2007). Predictive molecular signatures are derived from controlled experiments on well-characterized chemicals that are established to operate through specific modes of action (MoAs), or cause specific phenotypic effects. Meta-analyses can also be applied to explore the correlation between the gene expression profiles derived from a new test

chemical to existing repositories of expression profiles from well-characterized chemicals, to predict the chemical's hazards based on similarities (Waters *et al.* 2010). Finally, mechanistic data are important for understanding and exploring the specific molecular changes that lead to a chemical's toxicity, which can be used to establish human health relevance (Holsapple *et al.* 2006). Mechanistic assessment involves the application of bioinformatics tools to identify the molecular pathways and functions that are responding to the treatment. All three approaches are usually applied, to varying degrees, in toxicogenomics experiments.

Toxicogenomics data has been shown to be predictive of genotoxicity and cancer outcomes (Powell *et al.* 2006, Thybaud *et al.* 2007, Thomas *et al.* 2007, Ellinger-Ziegelbauer *et al.* 2008, Auerbach *et al.* 2010, Waters *et al.* 2010, Labib *et al.* 2012, Bourdon *et al.* 2013, Bhat *et al.* 2013, Thomas *et al.* 2013a, Romer *et al.* 2014, Jackson *et al.* 2014, Moffat *et al.* 2015). Within the field of cancer toxicology, predictive and mechanistic toxicogenomics are especially important for identifying non-genotoxic and indirectly genotoxic carcinogens. A battery of short-term genotoxicity assays has been developed that is highly effective in the identification of genotoxic carcinogens, which account for >70% of carcinogens (Waters *et al.* 2010). However, non-genotoxic and indirectly genotoxic carcinogens give negative or equivocal results in these tests, which is problematic because the results from these assays are used to prioritize compounds for further carcinogenicity testing. Thus, it has been proposed that short-term toxicogenomic analyses will be useful for identifying a chemical's ability to perturb molecular pathways that are associated with non-genotoxic MoAs in cancer in order to prioritize chemical testing in the two-year cancer bioassay. Eventually, with the development of robust

predictive signatures, the hope is that, together with other short-term testing strategies, toxicogenomics will replace the two-year cancer bioassay (Thomas *et al.* 2013b). This is an appealing prospect as this methodology promises to greatly increase efficiencies, and reduce resources and rodents used. Given the regulatory agenda, it is imperative that a more timely strategy for the identification of rodent (and probable human) carcinogens is developed.

1.2 Reducing Animal Use in Toxicity Testing

Since toxicity testing cannot be conducted in humans, animals are used as surrogates to measure *in vivo* toxicity. However, the use of animal models comes with its own set of concerns. Toward the reduction of animal testing, the 3Rs—reduction, refinement, and replacement—were originally proposed over half a century ago by Russell and Burch (1959). *Reduction* refers to reducing the number of animals used in toxicity testing, *refinement* refers to the refining of protocols to reduce suffering of experimental animals, and *replacement* refers to the development of *in vitro* and *in silico* assays that can replace *in vivo* testing. While the concept of the 3Rs is widely accepted, animal testing remains a massive part of chemical toxicity testing (Burden *et al.* 2015b). The benefits of and barriers to adopting alternative testing strategies (defined here as strategies that are mindful of the 3Rs, and, more broadly, alternatives to the current testing paradigm that improve upon efficiency of data generation and accuracy) are discussed below.

There are a number of compelling reasons for the adoption of alternative testing strategies. The most obvious is the ethical problem of enlisting 100s-1000s of animals in

assays whose endpoints are typically expected to be highly advanced disease states, or death. Public knowledge of these studies and the level of suffering experienced by these animals have led to increased pressure on regulatory agencies to legislate limits on animal testing. For example, the European Union has passed legislation that bans the use of animal testing on cosmetic products sold in the European Union (Regulation (EC) No 1223/2009). Animal testing is also a problem from a financial point-of-view because the costs associated with housing and caring for experimental animals are enormous. Budgeting constraints of businesses or governmental agencies that carry out animal testing limit the amount of testing that can be accomplished, and, therefore, the number of chemicals that can be realistically evaluated within a reasonable timeframe. Finally, from a biological perspective, important questions regarding the predictivity of rodent disease outcomes to human disease outcomes have been raised. Importantly, inbred populations of experimental animals do not adequately represent the genetic or environmental variation present across human populations. Further, there are important inter-species differences in biology that call into question the human relevance of results obtained in rodents. For example, phenobarbital, which is an anti-seizure medication that is routinely prescribed to treat epilepsy in human populations, causes tumours in rodents through a mechanism that is not considered relevant in humans (Holsapple *et al.* 2006). It has been suggested that human relevancy can be improved by performing *in vitro* assays using human cells, or by measuring global gene expression in (small groups of) exposed animals and establishing human relevance using chemical MoA considerations. Overall, there is a strong suite of ethical, regulatory, financial, and biological reasons to reduce, refine, and replace animal testing.

Despite the appeal of moving away from animal testing, there are many barriers to the elimination of animal use in chemical toxicity testing. First, the results produced in animal-based assays are embedded in today's standard approaches in risk assessment; therefore, a transition away from animal testing would require a massive paradigm change and will take a significant amount of time. Second, animal-based assays have a historical precedent for use, and even in instances where animal tests aren't explicitly required they are often perceived as less risky than non-standard approaches. Third, it is not always clear how to interpret *in vitro* data and extrapolate human health significance. Indeed, *in vitro* assays have their own inherent set of limitations, including legitimate experimental hurdles with respect to *in vitro* to *in vivo* extrapolation (IVIVE). Fourth, data derived from new technologies, particularly genomics technologies, are complex and can be difficult to interpret. And finally, it takes time to establish trust in new assays, which is vital to expediting the use of novel forms of data in a regulatory context. There are number of national and international consortia whose purpose is to build this trust through the validation of alternative methods (including: ICCVAM ⁹, ICATM ¹⁰, JaCVAM ¹¹, EURL-ECVAM ¹², and NICEATM ¹³). However, these validation schemes can be slow and expensive, and do not necessarily lead to assay implementation following their completion. Taken together, barriers to transitioning away from animal testing are typically bureaucratic or technical.

⁹ Interagency Coordinating Committee on the Validation of Alternative Methods

¹⁰ International Cooperation on Alternative Testing Methods

¹¹ Japanese Center for the Validation of Alternative Methods

¹² European Union Reference Laboratory for Alternatives to Animal Testing

¹³ The National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods

Programs like the US EPA's ToxCast¹⁴ have shown some success in linking *in vitro* data to *in vivo* outcomes by screening large chemical libraries using a very large collection of *in vitro* assays. ToxCast uses a suite of over 700 *in vitro* assays, which to date have been used to screen over 2000 chemicals. Using this immense dataset as a training set, it was shown that machine learning could be used to identify putative hepatotoxicants (Liu *et al.* 2015). Computer models like these are quickly becoming important tools in IVIVE and animal-free chemical risk assessment. Another study, which focused on ToxCast primary human cell line assays, screened 641 environmental chemicals and 135 failed pharmaceuticals. Chemical targets and molecular pathways that were indicative of chemical effect were identified and used to group chemicals by MoA (Kleinstreuer *et al.* 2014). However, ToxCast chemical screening is typically used for chemical prioritization for targeted testing (Dix *et al.* 2007, Judson *et al.* 2010, Judson *et al.* 2010, Judson *et al.* 2010), as opposed to being used directly for regulatory decision-making. And, in addition to the more than 700 *in vitro* assays, ToxCast also uses some model organism assays, for example zebrafish are used to test chemical effect on development (Padilla *et al.* 2012). Therefore, even in primarily *in vitro* testing programs, the complete elimination of animal models is not always possible (which has been attributed to a lack of confidence in and ability to fully interpret *in vitro* results, and the inability of *in vitro* systems to fully recapitulate *in vivo* conditions).

The limitations of *in vitro* assays are largely due to their not being able to truly recapitulate *in vivo* conditions. In particular, the cell lines used in *in vitro* assays are typically immortalized, which means that they are often derived from a tumour and are

¹⁴ ToxCast (<http://www.epa.gov/ncct/toxcast/>)

therefore genetically abnormal and might not respond to a toxicant in the same way that a normal cell would. In addition, immortalized cells are not always metabolically competent, which means that metabolic activation of compounds must be achieved prior to addition of the compound to the cells using S9 (a liver homogenate containing metabolic enzymes). Further, they are in monoculture, which means that the interactions between cell types that would be seen *in vivo* (for example: between hepatocytes and Kupffer cells in the liver) are not present; therefore, contributions to toxicity of systems like the inflammatory system will not be detected *in vitro*. Finally, *in vitro* systems cannot predict the physiological distribution of the metabolite(s) following biotransformation, which means that the affected tissues remain unknown. Issues surrounding IVIVE are being actively debated and were recently the topic of a special issue of the journal 'Toxicology' (Volume 332, June 2015), which is summarized by Yoon *et al.* (2015). Importantly, toxicogenomics data have significant potential to bridge the gap between the mechanistic data produced *in vitro* and the adverse outcomes produced *in vivo*.

While it does employ animals, *in vivo* toxicogenomics improves upon the two-year rodent bioassay because it queries a larger number of endpoints using a smaller number of animals that are not taken to a full-blown disease state. However, a large barrier to the use of toxicogenomics data is the complexity and magnitude of the data that are produced, whose interpretation requires advanced expertise in bioinformatics. Therefore, a certain amount of flexibility from the risk assessment community with respect to acquiring (or hiring) new skill sets is fundamental to the inclusion of these types of data in chemical assessments. Further, an important aspect of the adoption of

toxicogenomics (or any new assay) is to not have a “like-for-like” expectation with respect to data generation; for example, an ‘omics assay will never provide site-specific tumour data, and a cancer bioassay cannot reveal molecular MoA. However, *in vivo* toxicogenomics can provide data that bridge the gap between tier one high-throughput *in vitro* assays and tier three chronic *in vivo* carcinogenicity studies (tiered testing proposed by Thomas et al. (2013b) is discussed in greater detail in 7.3).

The 3Rs provide different ways to approach the reduction of animal use in toxicity testing. Programs like ToxCast and Tox21¹⁵ (in the US), and SEURAT-1¹⁶ (in Europe) aim to *replace* animal testing with high-throughput *in vitro* screens and computational tools. *In vitro* toxicogenomics is another strategy to *replace* the use of animals; whereas, *in vivo* toxicogenomics aims to *reduce* the use of animals and *refine* methods such that suffering is reduced because animals are necropsied at sub-chronic time-points. Toxicogenomic analysis of archival tissues (discussed below in 1.3) has the potential to massively *reduce* the number of animals required for production of predictive molecular signatures of carcinogenesis. Ultimately, the strategies that support the 3Rs are fundamental to the ethical treatment of animals and the timely assessment of chemical contaminants.

1.3 Formalin-Fixed, Paraffin-Embedded (FFPE) Archival Tissue

Tissue repositories have immense untapped potential in toxicogenomic research. Applications of archival tissue resources range from biomarker discovery to large-scale studies of genome-disease interactions. Curated sample banks house diverse tissue types

¹⁵ www.ntp.niehs.nih.gov/go/tox21

¹⁶ Safety Evaluation Ultimately Replacing Animal Testing 1 (www.seurat-1.eu)

from different models or study populations often linked with detailed clinical or pathologic outcomes. In many cases tissue archives contain unique samples from animal bioassays, clinical trials, or epidemiologic studies that may be impractical or impossible to repeat. However, even in cases where the animal study could be repeated, the use of archival samples is an experimental strategy that supports the 3Rs.

Despite their promise, the use of archival samples for transcriptomic profiling has been relatively limited to date. The majority of archival tissues are stored in FFPE blocks, which preserve tissue architecture for histopathological analysis and allow for tissue storage at room temperature. However, formalin treatment degrades RNA (through cross-linking, fragmentation, and adenine monomethylation), which significantly impairs molecular analyses (Farragher *et al.* 2008, Klopfleisch *et al.* 2011, Bass *et al.* 2014). These RNA effects can result in inconsistent genomic data and present important technical and analytical challenges when working with FFPE samples. Nevertheless, studies have demonstrated the value of using microarrays to analyze FFPE samples for understanding the molecular basis of a variety of cancers (Sadi *et al.* 2011, Jacobson *et al.* 2011, Budczies *et al.* 2011, Linton *et al.* 2012). As technologies continue to improve, so too does the prospect of FFPE genomics.

Genomic analyses of FFPE tissues were first conducted using biopsied tumour samples. Like toxicity studies, FFPE preservation of biopsied samples is the standard protocol in most hospitals (because FFPE allows for histological examination of the tissues). More recently these approaches have been used for toxicogenomics studies (Merrick *et al.* 2012, Jackson *et al.* 2013, Auerbach *et al.* 2014). Demonstration of the

preservation of important, chemically-dependent, disease-prognostic molecular signatures in FFPE samples will be fundamental to their use in predictive toxicogenomics.

1.4 The Liver

The liver is a major target organ in toxicological studies because it is a first-pass organ following oral exposures, the primary site of xenobiotic metabolism, and is a major target for poisoning and toxicity. Understanding normal liver function is important for comprehending the toxicological response of the liver following exposure to harmful chemicals.

1.4.1 Liver Physiology

The liver is organized into functional units called lobules. Lobules are primarily composed of hepatocytes that are organized around the liver's blood supplies. Zone 1 of the lobule is the area that is closest to the hepatic artery (HA, which provides oxygenated blood) and the portal vein (PV, which carries nutrient/chemical rich blood from the gastrointestinal tract to the liver) and zone 3 is the furthest from the blood supply. Blood from the PV and HA mixes, travels through the lobule in sinusoidal capillaries, and eventually drains into the central vein. Bile travels in the opposite direction (from zone 3-1) using channels called canaliculi, which converge on the Canals of Hering that drain into the bile duct.

The three main roles of the liver are: biosynthesis, metabolism (detoxification), and bile secretion. Biosynthesis (of serum proteins, bile, and cholesterol) and metabolism (of endogenous chemicals, and xenobiotics) are carried out by hepatocytes, which form

the largest cell population in the liver (~60%). Importantly, zone 3 hepatocytes have high levels of cytochrome P450 enzymes (Cyp P450), which are a large family of monooxygenases involved in chemical detoxification. Following mono-oxidation by P450, conjugation occurs, which facilitates chemical excretion in the bile or urine. Liver function, architecture, and homeostasis are reviewed by Stanger (2015).

The reactive metabolites produced by P450s are electrophiles that are, paradoxically, they are often more harmful to the cell than their parent compound. These metabolites can react with cellular nucleophiles, such as proteins and nucleic acids, leading to cellular damage that, in extreme cases, can result in cell death (e.g., furan; discussed in 1.5). The liver has the capacity to regenerate itself (discussed in 1.4.2); however, the concomitant increase in cellular proliferation can result in tumourigenesis (discussed in 1.4.3).

1.4.2 Liver Regeneration

Liver cells are normally quiescent, with only 1-2% turnover under normal circumstances. However, following surgical resection or chemical injury, liver cells rapidly divide to replace damaged tissue. The liver has two modes of regeneration: (1) following surgical resection, the liver regenerates via cellular hypertrophy and division of existing cells; (2) following chemically dependent damage, regeneration is thought to occur via proliferation of a population of facultative stem cells that is located at the Canals of Hering (in zone 1 where canaliculi join and drain into the main bile duct). Facultative stem cells are functional, differentiated cells that will dedifferentiate in response to tissue damage, thereby becoming a population of progenitor cells that can

redifferentiate to replace multiple lost cell types. The bile duct is lined with biliary epithelial cells (BECs) and it is thought that, upon toxicant-induced injury, they dedifferentiate into ‘oval’ cells, which can then redifferentiate into hepatocytes or BECs in order to regenerate damaged tissue; this process is known as ‘ductal expansion’ (Yanger & Stanger 2011, Stanger 2015). Ductal expansion is an important process to consider when studying hepatocarcinogens. A common mode of hepatocarcinogenesis following chemically dependent liver damage is dysregulated regenerative proliferation leading to the development of hepatocellular adenomas (HCA) and hepatocellular carcinomas (HCC). The cell-of-origin of these cancers is typically assumed to be a hepatocyte; however, it is important to also consider that oval cells might represent a reservoir of cancer progenitor cells since they are multipotent and primed for division (Oishi *et al.* 2014).

1.4.3 Liver Cancer

Worldwide, liver cancer is the fifth/seventh most common and the second/sixth most deadly cancer in men and women, respectively (Jemal *et al.* 2011). Risk factors for the development of liver cancer include viral hepatitis, alcoholic liver disease, nonalcoholic steatohepatitis (fatty liver disease), and chemical exposure (including cigarette smoke, and food contaminants). In particular, exposure to aflatoxin B1 (a metabolite of some fungi of the *Aspergillus* genus) has been strongly implicated in the development of HCC. Human exposure to aflatoxin B1 typically occurs via contaminated maize and groundnuts in areas with tropical or sub-tropical climates (sub-Saharan Africa, Eastern Asia, and South America). Aflatoxin B1 is metabolized in the liver to a reactive

epoxide that binds to DNA to produce guanine adducts (at position N7). These are highly mutagenic and carcinogenic adducts and, as a result, aflatoxin B1 is a major human health concern (Kensler *et al.* 2011, Kew 2013). More than 80% of liver cancers occur in conjunction with cirrhosis, fibrosis, oxidative stress, and/or chronic inflammation (Han 2012). During chemical exposure chronic oxidative stress is an important contributor to mutagenesis, which is necessary for malignant transformation. Further, during necrotic cell death the cell ruptures and releases its contents into the extracellular space. Released cellular components include damage-associated molecular patterns (DAMPs) that trigger an inflammatory response, which further creates a pro-tumourigenic environment. In fact, both genomic instability and tumour-promoting inflammation were recently recognized among four hallmarks of cancer (together with deregulating cellular energetics and avoiding immune destruction) (Hanahan & Weinberg 2011) that are in addition to the original six hallmarks of cancer (Hanahan & Weinberg 2000). Taken together, the liver's high contact with xenobiotics and its regenerative potential predispose it to carcinogenesis.

The original six hallmarks of cancer are: sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis, and resisting cell death (Hanahan & Weinberg 2000). Cancer development is a multistep process during which mutations are acquired and signaling is altered in order to achieve these hallmarks. Sustaining proliferative signaling is especially relevant here because chronic toxicant exposure typically leads to cytotoxicity, which activates regenerative proliferation in the liver. When liver regeneration occurs on a background of genomic instability, increased rates of cellular

proliferation increase the likelihood that driver mutations will become fixed leading to cellular transformation and clonal expansion of a malignant cell. Not surprisingly, a large number of mutations have been found in HCC tumours (Han 2012). Changes in cell physiology can be detected by measuring changes in gene expression. Recently, c-Jun (part of the dimeric AP-1 transcription factor) was identified as an important initiator of HCC in mice (Min *et al.* 2012). Further, chronic activation of the NRF2 oxidative stress response pathway has also been reported in HCC. It is thought that malignant cells hijack this normally protective pathway to gain a proliferative advantage (Ma & He 2012).

1.5 Test Article: Furan

Furan (CAS No. 110-00-9) is a liver toxicant and rodent hepatocarcinogen; it is classified as possibly carcinogenic to humans (group 2B) by the International Agency for Research on Cancer (IARC 1995). It is used in the synthesis of nylon, lacquers, insecticides, pharmaceuticals, and stabilizers. Furan was first reported in foods over 30 years ago (Maga & Katz 1979). It is formed during heat-treatment of food, probably through thermal decomposition of carbohydrates, and is found in coffee, and canned and jarred foods (Hasnip *et al.* 2006, Crews *et al.* 2007, Crews *et al.* 2009, Moro *et al.* 2012, Pye & Crews 2014). It is also produced during combustion and is therefore found in engine exhaust, wood smoke and tobacco smoke (IARC 1995). Levels of furan in foods vary considerably. Furan levels tend to be highest in canned foods that contain meat (up to 175 µg/kg) and in coffee (up to 100 µg/kg). Of particular concern are high levels of furan in baby food (10-100 µg/kg) because store bought baby food is often a very large proportion (up to 100%) of the infant's diet. However, on a population level, brewed

coffee is the largest contributor of furan exposure. Because coffee is one of the most widely used beverages around the world, its health effects have been extensively studied. Coffee is a complex mixture of many substances, including antioxidants, which have actually been shown to have positive health effects. Indeed, coffee has been reported to reduce risk of liver pathologies in high risk subjects (Morisco *et al.* 2014, Setiawan *et al.* 2015, Khalaf *et al.* 2015). Typical dietary exposure ranges to furan in the USA are 0.3-1 $\mu\text{g}/\text{kg}\text{-bodyweight}/\text{day}$. A margin of exposure (MoE) is the ratio of the toxicity effect level to the exposure dose. An MoE ratio of 10,000 would indicate that toxicity effect level is 10,000-times higher than the expected exposure level, which indicates very low risk of toxicity; whereas, a lower MoE, for example 100, would indicate a very high risk of toxicity. A typical MoE threshold for a well characterized chemical (like furan) is 1000. An MoE calculated for furan-dependent HCC produced values of 750 and 4300 for infants and adults, respectively, which suggests that furan is an important dietary risk factor, especially in infants (Carthew *et al.* 2010a). Taken together, furan is a rodent hepatocarcinogen and food contaminant that is a potential human health risk.

The National Toxicology Program's (NTP) two-year cancer bioassay showed that furan induces HCC and HCA in a dose-dependent manner in B6C3F1 mice. F344 rats developed cancer at lower doses than mice and, in addition to HCC and HCA, the rats developed cholangiocarcinoma and mononuclear cell leukemia (NTP 1993). Due to equivocal results in the standard battery of genotoxicity tests (Table 1.1), it seemed unlikely that furan was producing cancer via a directly genotoxic MoA. Additionally, the NTP's 13-week study showed dose-dependent increases of histopathological markers for cytotoxicity, necrosis and cellular proliferation in B6C3F1 mice and F344 rats (both

genders). In their two-year study, there was evidence of chronic inflammation, hepatic fibrosis, hyperplasia, degeneration, and necrosis in the liver. Therefore, the MoA proposed for furan was chronic cytotoxicity and inflammation followed by dysregulated regenerative proliferation, which is thought to be the most common MoA for spontaneous HCC (Nakagawa & Maeda 2012), and is supported by follow-up studies on furan (Fransson-Steen *et al.* 1997, Moser *et al.* 2009).

Extremely high rates of spontaneous liver cancer in control male mice (26/50 and 7/50 in males and females, respectively) in the NTP cancer bioassay led Moser *et al.* (2009) to conduct their follow-up cancer bioassay using female mice. In addition, Moser *et al.* used a lower dose range (0, 0.5, 1, 2, 4 or 8 milligrams per kilogram bodyweight per day, mkd, furan) than was used in the NTP cancer bioassay (0, 8, 15 mkd), which included non-carcinogenic doses (see Table 1.2 for cancer incidences from both studies). Mice were exposed to furan for three weeks or two years. Consistent with furan's proposed MoA, significant increases in hepatic cytotoxicity (beginning at 1 mkd; measured by serum alanine aminotransferase, ALT) and hepatocyte proliferation (at 8 mkd; measured by BrdU incorporation) were observed after three weeks. Tumourigenesis was observed at 4 and 8 mkd after two years.

The ability of furan to induce sufficient cytotoxicity to promote regenerative proliferation is related to the chemical's metabolism. Experiments in F344 rats using radio-labeled furan have demonstrated that 80% of ingested furan is eliminated in the urine or expired within 24 hours; the remaining [14C] was associated with liver proteins (Burka *et al.* 1991). Furan does not react with DNA *in vivo* (Churchwell *et al.* 2015). While Burka *et al.* detected ten metabolites of furan in the urine, the most important

metabolite of furan is cis-2-butene-1,4-dial (BDA). The furan heterocycle is opened by oxidization to form BDA by cytochrome P450 2E1 (CYP2E1), which is a monooxygenase (Kedderis *et al.* 1993). The role of CYP2E1 in chemical toxicity is well documented and is reviewed by Gonzalez (2005). *In vitro* experiments have demonstrated that metabolic activation to BDA is necessary for furan-induced cytotoxicity (Kellert *et al.* 2008), and inhibition of CYP2E1 is sufficient to prevent cytotoxicity in female B6C3F1 mice and in mouse, rat or human microsomes (Fransson-Steen *et al.* 1997, Gates *et al.* 2012). The latter indicates that CYP2E1 is solely responsible for the xenobiotic metabolism of furan. Further, the mechanism for furan oxidation to BDA by CYP2E1 is similar in mouse, rat, and human hepatocytes, which suggests that the furan-dependent adverse effects seen in rodents are applicable to humans (Gates *et al.* 2012, Gates *et al.* 2014).

The CYP2E1 catalytic cycle is prone to producing reactive oxygen species (ROS) due to dissociation of superoxide radical from complex I (Gonzalez 2005, Lu & Cederbaum 2008). CYP2E1 is unique among P450 enzymes because it is substrate stabilized; therefore, chronic exposures will have the effect of chronically increasing the oxidative status of the cell. Indeed, CYP2E1-over-expressing HepG2-E47 cells have elevated GSH, NRF2 (mRNA and protein), and NRF2 target gene expression, compared to standard HepG2 cells that do not express CYP2E1, which is demonstrative of the elevated oxidative status of these cells (Cederbaum 2009). GSH plays important roles both as an antioxidant molecule and a conjugant in xenobiotic metabolism. BDA has been shown to bind cellular glutathione (GSH) and reduce it to 30% of control levels in F344 rat hepatocytes, which reduces the amount of GSH available for the reduction of

ROS, and is sufficient to produce cytolethality (Carfagna *et al.* 1993). In addition to the resultant cytotoxicity, evidence of oxidative stress-induced genomic damage in response to furan includes 8-oxo-dG adducts (Hickling *et al.* 2010) and oxidized purine and pyrimidine bases (Ding *et al.* 2012) in rat liver. Therefore, it seems clear that cellular damage due to oxidative stress is an important part of furan's MoA.

Another Cyp2E1 substrate, acetaminophen, is the most common cause of drug-induced liver failure in humans. Acetaminophen is metabolized to N-acetyl-p-benzoquinone imine (NAPQI), which binds and depletes GSH. GSH is completely depleted in instances of overdose and excess NAPQI is free to adduct to cellular proteins. In particular, adducts formed with mitochondrial proteins produce mitochondrial dysfunction leading to increased oxidative stress and decreased ATP production. Together, these culminate in overwhelming cytotoxicity resulting in organ failure (Jaeschke *et al.* 2014).

Genomics data provide valuable insight into toxicant MoA, building on apical data and enabling more informed assessment of the relevance of the molecular perturbations observed in the rodent studies to human risk. Gene expression analysis of furan exposed rodent livers would provide valuable insight into the molecular underpinnings of the furan MoA. In addition, furan is extremely well characterized using standard toxicity tools. Thus, furan provides an exceptional case study to examine how toxicogenomics can be used for hazard identification, to inform MoA, to explore human relevance and to evaluate the potential use of dose-response data from toxicogenomics experiments in quantitative human health risk assessment.

1.6 Thesis Overview

1.6.1 Rationale

Chemical contaminants in the environment or in foods represent potential human health risks and therefore must be tested for their ability to produce adverse health effects (here we focus on carcinogens). Currently, putative carcinogens are tested in the two-year rodent cancer bioassay to determine if they cause cancer and at what dose. Unfortunately, performing the cancer bioassay on the huge backlog of priority contaminants is not feasible. Further, on a societal level, there is less tolerance for using the huge number of experimental animals that would be required. Ultimately, whether the bottom-line is logistical, financial, or ethical, it is widely accepted that alternative testing strategies are urgently needed. The goal of this thesis is to explore whether toxicogenomics might contribute to the solution for this problem. Using furan as a model compound, we approached this question from three angles, asking if toxicogenomics could provide accurate information addressing: quantitative prediction of doses for apical effects (Chapter 2,4), prediction of potential disease and MoA (Chapter 2), and mechanistic understanding of response (Chapter 2). Although toxicogenomics studies already use only a fraction of the experimental animals required for standard cancer testing, we wished to explore additional ways to further reduce animal testing. In Chapter 3 we compared gene expression patterns across two cohorts of mice (\pm bromodeoxyuridine, BrdU) to determine if one cohort of animals (those treated with BrdU to assess hepatocellular proliferation, a standard apical test) would have been sufficient. This addresses the ability of toxicogenomics to be integrated into standard tests as well. In Chapter 4, we address the challenge of consistency and reproducibility of the evolving

genomics technologies, to establish confidence in the use of different toxicogenomics approaches and establish best practices to move the field forward. We also explored approaches that could be used in the selection of point of departure (PoD) from genomics data. In Chapter 5 and 6 we asked whether these studies could be conducted using archival tissue samples, thereby eliminating the need for experimental animals for signature or biomarker development that is typically conducted on chemicals with established MoAs. Specific objectives of each chapter are described below and the experimental design of this thesis is presented in Fig. 1.2.

1.6.2 Chapter Objectives

Chapter 2

Hypothesis: Furan-dependent transcriptional changes following a 21-day exposure will provide predictive and mechanistic information regarding furan's MoA and will be prognostic of the dose at which furan-dependent liver cancer occurs in mice.

Objectives:

- To perform global gene expression analysis of furan-exposed and control mouse liver tissues using 2-colour Agilent 8 x 60K SurePrint Gene Expression microarrays.
- To use quantitative toxicogenomics to model the dose-response of furan-dependent transcriptional changes, predictive toxicogenomics to investigate the potential hazards posed by furan exposure in the liver, and mechanistic toxicogenomics to understand furan's MoA in the liver.

Importance: This case study provides a phenotypically anchored example of how toxicogenomics data can be used to produce novel predictive, mechanistic, and quantitative data that can be used to support chemical risk assessment.

Chapter 3

Hypothesis: BrdU treatment does not mask furan-dependent changes in gene expression.

Objectives:

- To compare global gene expression profiles of the furan exposed (or control) liver tissues to a second cohort of animals from the same study that were co-exposed to BrdU, using 2-colour Agilent 8 x 60K SurePrint Gene Expression microarrays.
- To determine if treatment with BrdU significantly changes toxicant-dependent gene expression profiles.

Importance: If BrdU does not produce significant changes in gene expression, then the number of experimental animals required for this type of study can be halved in future studies.

Chapter 4

Objectives:

- To compare BMD values for furan that are modeled using data from three genomics platforms: 2-colour Agilent 8 x 60K SurePrint Gene Expression microarrays, PolyA-

enrichment RNA-sequencing (Illumina HiSeq2000), and qPCR arrays (SA BioSciences).

- To understand the effects of statistical filtering of genomics data on BMDExpress output values.
- To compare four approaches for choosing a transcriptional PoD.

Importance: These data provide important recommendations for design and interpretation of quantitative toxicogenomics data in studies applying BMD analysis using different genomics platforms.

Chapter 5

Objective: To develop a protocol for RNA extraction and microarray analysis of FFPE samples. Specifically, to compare the quality of RNA produced from six commercially available FFPE RNA extraction kits and the quality of data from four one- and two-colour microarray protocols.

Chapter 6

Hypotheses:

1. Furan-dependent changes in gene expression can be detected in FFPE liver samples.
2. Ribo-depletion RNA-seq will outperform polyA-enrichment RNA-seq, and 1-colour microarrays will outperform 2-colour microarrays for global gene expression profiling of FFPE tissue samples.
3. Increased time in formalin will negatively impact gene expression data.
4. Increased age of tissue block will negatively impact gene expression data.

Objectives:

- To determine the best method for toxicogenomic analysis of FFPE tissue samples by comparing gene expression profiles of paired frozen (FRO) and FFPE furan-exposed mouse liver tissue using two genomics platforms (RNA-seq and microarray) and four protocols.
- To understand the time-in-formalin effect by comparing gene expression profiles of samples that were fixed in formalin for 18 hours or three weeks.
- To understand the time-in-block effect by comparing quality control metrics of 8-, 19-, and 26-year-old tissue blocks (assayed by ribo-depletion RNA-seq).

Importance: Successful demonstration that FFPE samples are acceptable for use in toxicogenic experiments will allow for the rapid production of phenotypically-anchored predictive signatures using archival tissue samples. This will significantly reduce the number of experimental animals that would otherwise be required to produce these signatures.

1.7 Tables and Figures

Table 1.1 The results of the short-term genotoxicity testing of furan.

Assay	Dose Range (mkd)	Result	Reference
Ames Reverse Mutation Assay (\pm S9)		Negative	(NTP 1993)
L5178Y/TK+/- Mouse Lymphoma Forward Mutation assay (-S9)		Positive	(NTP 1993)
Sister Chromatid Exchange Assay			
In vitro sister chromatid exchange in CHO cells test (\pm S9)		Positive	(NTP 1993)
<i>In vivo</i> sister chromatid exchange in B6C3F1 mouse bone marrow cells		Negative	(NTP 1993)
In vitro Micronucleus (dsDNA breaks, aneuploidy)			
In vitro micronucleus (human lymphocytes) (\pm S9)	0-100 mM	Negative	(Durling <i>et al.</i> 2007)
Modified in vitro micronucleus assay in B6C3F1 mice (Splenocytes)	2-15	Positive	(Leopardi <i>et al.</i> 2010)
In vivo Micronucleus (dsDNA breaks, aneuploidy)			
<i>In vivo</i> micronucleus in Big Blue rats	2-30	Negative	(McDaniel <i>et al.</i> 2012)
<i>In vivo</i> micronucleus (Balb/C and CBA mice)	225-300	Negative	(Durling <i>et al.</i> 2007)
<i>In vivo</i> micronucleus assay (F344 rat)	2-16	Negative	(Ding <i>et al.</i> 2012)
Comet Assay (dsDNA breaks)			
Comet assay in Big Blue rats (liver)	2-30	Positive >16 mkd	(McDaniel <i>et al.</i> 2012)
Comet Assay in B6C3F1 mice (Splenocytes)	2-15	Negative	(Leopardi <i>et al.</i> 2010)
<i>In vivo</i> comet assay (F344 rats; liver)	2-16	Positive (16 mkd)	(Ding <i>et al.</i> 2012)
<i>In vivo</i> comet assay (F344 rats; bone marrow)	2-16	Positive (>8 mkd)	(Ding <i>et al.</i> 2012)
Fpg-sensitive DNA damage (Oxidative DNA damage)	2-16	Positive (>12 mkd)	(Ding <i>et al.</i> 2012)
EndoIII-sensitive DNA damage (Oxidative DNA damage)	2-16	Positive (>4 mkd)	(Ding <i>et al.</i> 2012)
Other			
Gamma-H2AX assay in B6C3F1 mice (Splenocytes)	2-15	Positive >8 mkd	(Leopardi <i>et al.</i> 2010)
Pig-a, hprt, and cII mutation assays in Big Blue rats	2-30	Negative	(McDaniel <i>et al.</i> 2012)

CHO=Chinese hamster ovary; Fpg= formamidopyrimidine DNA glycosylase; EndoIII= endonuclease III

Table 1.2 Incidence of furan-dependent liver cancer in female B6C3F1 mice

	Dose of furan (mkd)						
	0	0.5	1	2	4	8	15
Study: NTP, 1993							
HCA	5/50 (10%)					31/50 (62%)	48/50 (96%)
HCC	2/50 (4%)					7/50 (14%)	27/50 (54%)
HCC or HCA	7/50 (14%)					34/50 (68%)	50/50 (100%)
Study: Moser et al. 2009							
HCA	3/36 (8%)	4/72 (6%)	4/53 (8%)	4/41 (10%)	11/36 (31%)	25/39 (64%)	
HCC	0/36 (0%)	4/72 (6%)	2/53 (4%)	1/41 (2%)	2/36 (6%)	11/39 (28%)	
HCC or HCA	3/36 (8%)	8/72 (11%)	6/53 (11%)	5/41 (12%)	12/36 (33%)	29/39 (74%)	

Bold text=significant over control (p<0.05)

Greyed out=doses not used in study

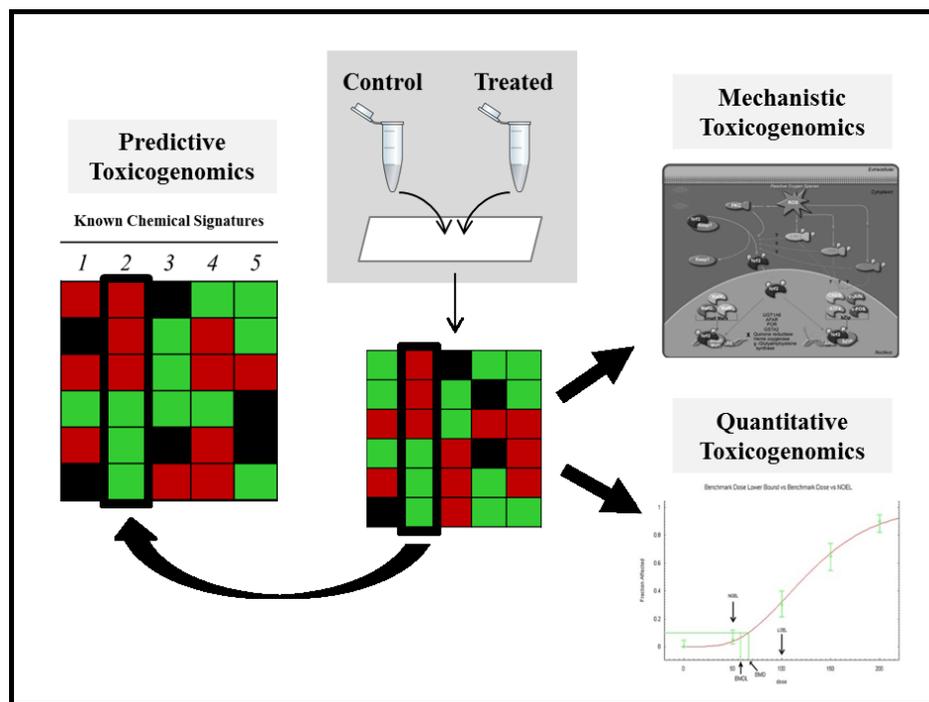


Figure 1.1 Three approaches for analyzing toxicogenomics data. In predictive toxicogenomics novel gene expression signatures are compared to known chemical signatures in order to predict the chemical's potential hazard or MoA. In mechanistic toxicogenomics, bioinformatics software is used to identify perturbed pathways and processes in order to understand the compound's molecular MoA. In quantitative toxicogenomics, the dose-response of gene expression data are modeled in order to pinpoint doses at which important biological processes depart from background levels.

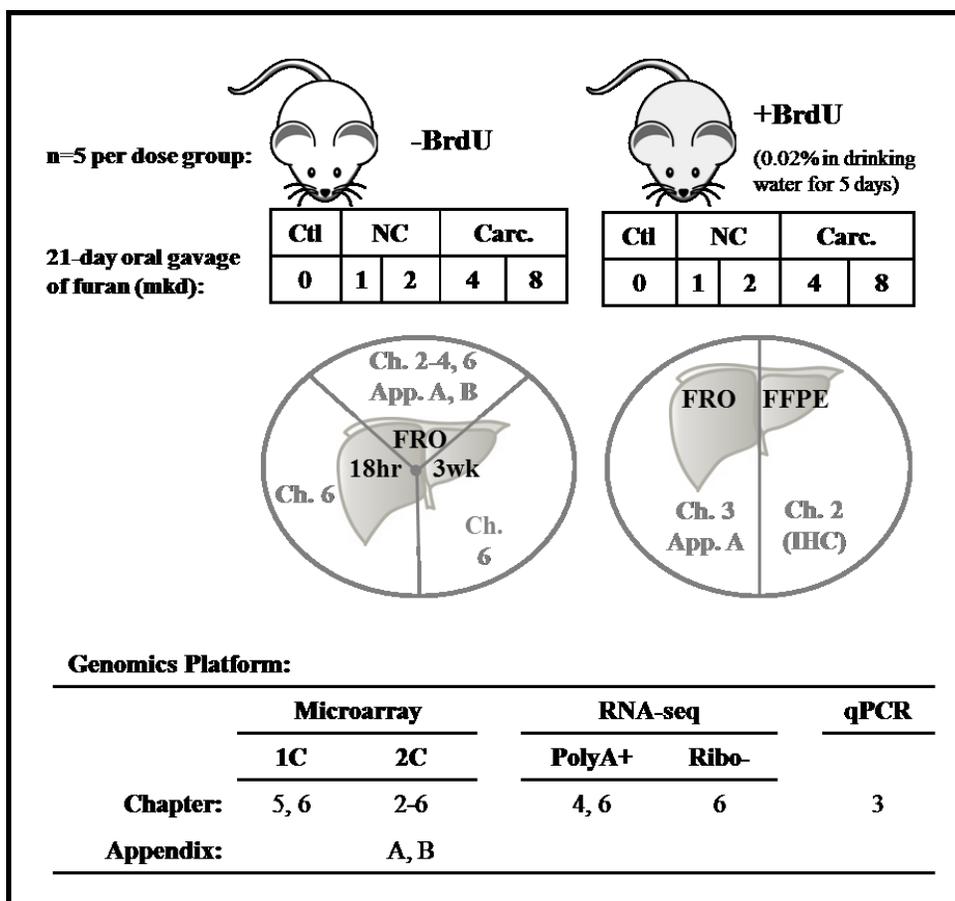


Figure 1.2 Experimental design of this thesis. The top panel shows that two cohorts of mice were used (\pm BrdU) and they were exposed for 21 days to non-carcinogenic or carcinogenic doses of furan (or vehicle-only control). Extracted livers were divided and pieces were preserved by freezing or in formalin (shown in the circles). There were two time-points (18 hours and 3 weeks) of formalin fixation in the $-$ BrdU group. All samples were used for genomics analysis, with the exception of the $+BrdU$ FFPE samples, which were used to immunohistochemically quantify cellular proliferation. A variety of genomics platforms and protocols were used (shown in the table at the bottom). Acronyms: 1C=one-colour microarrays; 2C=two-colour microarrays; App=appendix; BrdU=bromodeoxyuridine; Carc.=carcinogenic; Ch=chapter; Ctl=control; FRO=frozen; FFPE=formalin-fixed, paraffin-embedded; IHC=immunohistochemistry; mkd=milligrams per kilogram bodyweight per day; NC=non-carcinogenic; PolyA+=PolyA-enrichment RNA-sequencing; qPCR=quantitative polymerase chain reaction; Ribo-=ribosomal RNA depletion RNA-sequencing.

CHAPTER 2

2 CHAPTER: CASE STUDY ON THE UTILITY OF HEPATIC GLOBAL GENE EXPRESSION PROFILING IN THE RISK ASSESSMENT OF THE CARCINOGEN FURAN.

Modified from: A.F. Jackson, Williams, A., Recio, R., Waters, M.D., Lambert, I.B.,
Yauk, C. 2014. *Toxicology and Applied Pharmacology*, 274(1):63-77.

2.1 Preamble

The purpose of this case study was to perform quantitative, predictive, and mechanistic toxicogenomics for the hepatocarcinogenic food contaminant, furan. Furan is an ideal candidate for this type of study since its toxic and carcinogenic effects to the liver have been well studied. Therefore, molecular changes can be phenotypically anchored, which is an important aspect of this proof-of-concept study. In this study we tested the hypothesis that prior to the occurrence of furan-dependent tumour formation, prognostic changes to gene expression would occur. We performed global gene expression analysis of liver tissue from mice that were sub-chronically exposed to non-carcinogenic and carcinogenic doses of furan.

2.2 Abstract

Furan is a chemical hepatocarcinogen in mice and rats. Its previously postulated cancer MoA is chronic cytotoxicity followed by sustained regenerative proliferation; however, its molecular basis is unknown. To this end, we conducted toxicogenomic analysis of B3C6F1 mouse livers following three week exposures to non-carcinogenic (0, 1, 2 mkd) or carcinogenic (4 and 8 mkd) doses of furan. We saw enrichment for pathways responsible for cytotoxicity: stress-activated protein kinases (SAPK) and death receptor (DR5 and TNF-alpha) signaling, and proliferation: extracellular signal regulated kinases (ERK) and TNF-alpha. We also noted the involvement of NF-kappaB and c-Jun in response to furan, which are genes that are known to be required for liver regeneration. Furan metabolism by CYP2E1 produces BDA, which is required for ensuing cytotoxicity and oxidative stress. NRF2 is a master regulator of gene expression during oxidative stress and we suggest that chronic NRF2 activity and chronic inflammation may represent critical transition events between the adaptive (regeneration) and adverse (cancer) outcomes. Another objective of this study was to demonstrate the applicability of toxicogenomics data in quantitative risk assessment. We modeled benchmark doses for our transcriptional data and previously published cancer data, and observed consistency between the two. Margin of exposure values for both transcriptional and cancer endpoints were also similar. In conclusion, using furan as a case study we have demonstrated the value of toxicogenomics data in elucidating dose-dependent MoA transitions and in quantitative risk assessment.

2.3 Introduction

Established assays for the identification of non-genotoxic carcinogens are a current gap in the standard battery of short-term carcinogenicity tests. Toxicogenomics is expected to provide an effective tool for the identification of diverse MoAs and may fill an important testing gap for non-genotoxic carcinogens (Waters *et al.* 2010). MoA information gleaned from toxicogenomic studies of carcinogens has the added benefit of providing mechanistic information, thereby facilitating inter-species comparisons for inferring human risk. In this study we characterize global gene expression profiles in liver tissue taken from female B6C3F1 mice that were sub-chronically exposed to non-carcinogenic (0, 1, 2 mkd) and carcinogenic (4 and 8 mkd) doses of furan. We perform extensive bioinformatics analyses that are anchored in previously published apical (phenotypic) endpoint data (NTP 1993, Moser *et al.* 2009, Gill *et al.* 2011) in order to elucidate the molecular mechanism by which furan causes liver cancer. More broadly, our goal was to use furan as a case study to demonstrate the value of using sub-chronic exposures and genomic tools to inform risk assessment of non-genotoxic carcinogens.

2.4 Materials and Methods

2.4.1 Chemical

Furan (CAS No. 110-00-9) (> 99% pure) (Sigma-Aldrich Chemical Co., Milwaukee, WI) was mixed with Mazola corn oil to the appropriate concentration. Doses were prepared separately on a volume-to-weight ratio (v:w), were de-aerated with inert gas, and were stored in 8 mL brown glass vials (sealed with plastic closures and modified

silicon septa) in the refrigerator for up to 14 days (based on previous reports of furan stability (NTP 1993)).

2.4.2 Animals

5-6 week old female specific pathogen free B6C3F1 mice were purchased from Charles River Breeding Laboratories (Portage, ME) and were allowed to acclimatize for at least 7 days prior to the start of the study. Feed (NIH-07; Zeigler Brothers, Inc., Gardners, PA) and tap water were available ad libitum up until the time of necropsy. Mice were housed five per cage in polycarbonate cages in a specific pathogen free (SPF) and Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) accredited facility. All procedures were conducted in compliance with the Animal Welfare Act Regulations (9CFR1–4). Mice were handled and treated according to the guidelines provided in the National Institute of Health (NIH) Guide for the Care and Use of Laboratory Animals (ILAR, 1996; <http://dels.nas.edu/ilar/>).

Female mice were dosed with furan in corn oil at 0, 1, 2, 4, or 8 mkd by oral gavage for three weeks (n=10 per dose). We chose to use female mice because they have a lower spontaneous tumour rate than males (Haseman et al., 1998). N=4, 5, 3, 4, 5 per dose group were treated with 0.02% BrdU (Sigma Chemical Co., St. Louis, MO, USA) in drinking water for 5 days just prior to sacrifice. Upon necropsy, there remained n=5 mice in each non-BrdU group. Some mice were lost due to early (pre-BrdU treatment) mis-dosing or esophageal puncture. Four hours after their final dosing, mice were anesthetized by CO₂ inhalation prior to euthanasia by exsanguination achieved by cutting the caudal vena cava after blood collection. One animal per group was killed and this

continued until all mice had been sacrificed; this occurred over a period of 100 minutes (beginning at 1 pm). Body weights were recorded. Serum enzymes were not analyzed since these endpoints have been previously examined and reported (NTP 1993, Moser *et al.* 2009, Gill *et al.* 2011). The left, median, right posterior and right anterior lobes of the liver were cut into 3 sections. One section from each lobe was cut into 0.25-0.5 cm³ pieces. These were snap-frozen in liquid nitrogen then stored at or below -70°C until processed for RNA extraction. The remaining two sections of liver were fixed for 18 hours or three weeks in neutral phosphate-buffered formalin, then in 100% ethanol for 72 hours and stored at room temperature.

2.4.3 Immunohistochemistry for BrdU Nuclear Staining

Hepatocytes undergoing DNA synthesis were identified by BrdU nuclear staining using immunohistochemistry. Briefly, liver sections were stained for BrdU in an automatic staining machine (IntelliPATH, Biocare, Concord, CA). A small section of duodenum was evaluated on each slide as a positive control to ensure adequate BrdU delivery and staining. Deparaffinized sections were incubated with hydrochloric acid for 20 minutes at 37°C for antigen retrieval, neutralized in borate buffer, and immersed in trypsin for 3 minutes at 37°C for antigen retrieval. The sections were exposed to 10% normal rabbit serum for 20 minutes at room temperature to inhibit nonspecific binding. The sections were then incubated for 1 hour at room temperature with rabbit monoclonal anti-BrdU antibodies (Accurate Chemical and Scientific Corporation, Westbury, NY) diluted 1:1000. Sections were incubated with hydrogen peroxide for 15 minutes followed by incubation with avidin and biotin (Vector, Burlingame, CA) at room temperature for

15 minutes each. Afterwards the sections were incubated with rabbit anti-rat IgG (Vector, Burlingame, CA) at a 1:500 dilution for 30 minutes at room temperature to capture primary antibody. Sections were then incubated with Vector ABC label for 30 minutes at room temperature. The antigen-antibody reaction was visualized with betazoid diaminobenzidine (Biocare, Concord, CA) for 5 minutes. The sections were counterstained with CAT hematoxylin (Biocare, Concord, CA) for 2 minutes.

A liver section from each mouse was microscopically evaluated for hepatocyte cell proliferation. Random fields in the liver were scored without knowledge of animal identity for BrdU incorporation by light microscopy. At least 1000 cells per slide were counted. The percent labeling index was calculated by the total number of immunoreactive BrdU labeled hepatic nuclei divided by the total number of nuclei counted times 100. Representative images from each dose group were captured on an Olympus BX41 microscope and Olympus DP25 camera.

A generalized linear model was applied to the data with the binomial error distribution using the R software (R-Core-Development-Team 2012). The `glht` function in the `multcomp` library (Hothorn *et al.* 2008) for testing general linear hypotheses was used to test for differences between exposed and controls. The p-values from the Wald Test were adjusted using the Dunnett multiple comparison procedure (Dunnett 1955). Estimates were back transformed to obtain fold change estimates and the delta method was applied to obtain the estimate for the standard errors.

2.4.4 RNA Extraction

RNA was extracted from ~100 mg frozen liver tissue using the RNeasy Midi RNA Extraction kit (Qiagen, Mississauga, ON, Canada). An Omni tissue homogenizer with a disposable 7 mm Omni generator probe was used (Omni #34750, Omni International, Marietta, GA). RNA was quantified using a NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA) and qualified using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Mississauga, ON, Canada) and stored at -80°C.

2.4.5 Microarray

RNA was extracted from each of the non-BrdU mice for all furan doses (0, 1, 2, 4, 8 mkd; n=5 per dose group). Sample RNA (200 ng) was used together with a mouse universal reference RNA (Stratagene by Agilent Technologies Inc., Mississauga, ON, Canada) to synthesize, amplify and label cRNA using the Low Input Quick Amp Labeling Kit (Agilent Technologies Inc., Mississauga, ON, Canada). Labeled cRNA was purified using the RNease Mini Kit (Qiagen, Mississauga, ON, Canada). Amplification and labeling efficiency of cRNA were quantified using a NanoDrop spectrophotometer. Hybridization mixes were prepared using the Hi-RPM Gene Expression Hybridization Kit (Agilent Technologies Inc., Mississauga, ON, Canada). 300 ng of Cy3-labeled reference RNA and 300 ng Cy5-labeled sample cRNA were hybridized on SurePrint G3 Mouse GE 8x60K microarrays (Agilent Technologies Inc., Mississauga, ON, Canada) at 65°C for 17 hours at 10 rpm. Slides were washed according to the manufacturer's specifications with Gene Expression Wash Buffers 1 and 2 (Agilent Technologies Inc.,

Mississauga, ON, Canada), scanned using an Agilent G2505B scanner at 5 μm resolution. Data were extracted using Agilent Feature Extraction Software, version 11. The complete dataset is available through the Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo/) accession number GSE48644.

A block design, treating the slide as the blocking effect was employed to determine global differential gene expression. Median signal intensities were normalized using locally weighted scatterplot smoothing (LOWESS) (Bolstad *et al.* 2003) in R (R-Core-Development-Team 2012). Probes with technical replicates were then averaged. Differential gene expression was determined using the microarray analysis of variance (MAANOVA) library (Wu *et al.* 2003). The F_s statistic (Cui *et al.* 2005), a shrinkage estimator, was used to determine gene-specific treatment effects, and the associated p-values were estimated using the permutation method (30,000 permutations with residual shuffling). The p-values were adjusted for multiple comparisons using the false discovery rate (FDR) approach (Benjamini & Hochberg 1995). Fold changes were estimated using least square means of each pairwise comparison. Genes having an FDR-adjusted $P \leq 0.05$ and a fold change $\geq \pm 1.5$ were deemed differentially expressed. Upon removal of outliers (arrays with high background), the final sample sizes used for gene expression analysis were $n=5, 4, 5, 4, 5$ for 0, 1, 2, 4, 8 mkd furan dose groups, respectively.

2.4.6 Bioinformatics

2.4.6.1 Ingenuity Pathway Analysis (IPA)

IPA (www.ingenuity.com/products/ipa) was used to identify molecular pathways that were affected by furan treatment and to predict activated upstream regulators. The

expression and significance cut-offs applied to the data were fold change $\geq \pm 1.5$ and FDR-adjusted $P \leq 0.05$. Enrichment of a canonical pathway is determined in IPA based on the number of differentially expressed genes in the data set that also appear in the pathway. The significance threshold for canonical pathways is $P \leq 0.05$, which is calculated using a right-tailed Fisher's exact test by IPA. Significance of predicted upstream regulators was calculated by IPA and quantified using a Z-score (we considered $Z > 1.9$ as activated and $Z < -1.9$ as inhibited), which takes into account the number of downstream genes that are differentially expressed in the gene list.

2.4.6.2 Disease Prediction and Chemical Profile Comparison

Gene expression changes induced by 8 mkd furan were mined against public genomic data repositories for chemical- and disease-induced changes in gene expression using NextBio (www.nextbio.com). Pairwise gene signature correlations and rank-based enrichment statistics are employed in the NextBio score calculations, where the chemical or disease with the highest similarity will be given a score of 100 and the rest are normalized accordingly (Kupersmidt *et al.* 2010).

2.4.6.3 PCA in R

Principle component analysis (PCA) was conducted on the relative signal intensities for the significant probes based on the correlation matrix in R using the `prcomp()` function (Mardia *et al.* 1979, Becker *et al.* 1998, Venables & Ripley 1999). The first three components were then plotted using the `scatterplot3d()` function in the

scatterplot3d library (Ligges & Mächler 2003). A full explanation of the PCA is available in Appendix C, section C1.

2.4.6.4 Cluster Analysis

Data was obtained from GEO and European Bioinformatics Institute (EBI; www.ebi.ac.uk/arrayexpress/). All data sets that used the Affymetrix platform (E-MEXP-82; GSE13149; GSE18858; GSE20427; GSE26538) were normalized using Robust Multi-array Average, RMA (Irizarry *et al.* 2003), using the ReadAffy() function in the affy library in R (Gautier *et al.* 2004). Two-colour platform (Furan; BaP) background subtracted signal intensities were normalized using the LOWESS approach and background subtracted signal intensities for GSE35934 and GSE4874 (with a dye adjustment) were normalized using cyclic LOWESS (Bolstad *et al.* 2003).

Normalized probe intensities with common Gene Symbols were then collapsed using the median normalized signal intensity. Based on Gene Symbol, all normalized data were merged together yielding 3190 common Gene Symbols. A dendrogram using these data was produced using 1-pearson correlation as the distance metric with average linkage.

2.4.7 Benchmark Dose (BMD) Calculations

Apical Endpoint Data: BMD and BMDL (lower confidence limit of BMD) values were modeled for previously reported apical endpoint measurements (NTP 1993, Gill *et al.* 2011) using the United States Environmental Protection Agency's (US EPA) Benchmark Dose Software (BMDS) version 2.3.1 (www.epa.gov/ncea/bmds/) (Davis *et*

al. 2011). Data were pre-screened for homogeneity of variance. Continuous data were run against five models (Exponential, Hill, Linear, Polynomial and Power) and dichotomous data were run against nine models (Gamma Multi, Logistic, LogLogistic, LogProbit, Multistage, Multistage Cancer, Probit, Weibull and Quantal-Linear). The best model for each data set was chosen based on visual inspection of curve fitting, lowest Akaike's Information Criterion (AIC, which measures the relative goodness of fit) value, goodness of fit $P > 0.05$, and scaled residuals between +2.0 and -2.0. The AIC was used as the deciding factor between equally well fitting models. These selection criteria were applied when values generated from each model were within 3-fold of each other, otherwise the lowest value was used.

Gene Expression Data: BMDEExpress (Yang *et al.* 2007) was applied to calculate BMD (mean and median) and BMDL (mean and median) values for genes, IPA pathways and Gene Ontology (GO) terms. Datasets were pre-filtered with FDR $P \leq 0.05$ and fold change $\geq \pm 1.5$ in at least one dose before importing into BMDEExpress. Five models were compared (Hill, Power, Linear, Polynomial 2° and Polynomial 3°) and the best was chosen based on: (1) a nested Chi-square test, with cut-off of 0.05; (2) the lowest AIC; (3) maximum iterations of 250; (4) confidence level of 0.95; and (5) benchmark response (BMR) of 1.349 (number of standard deviation defining BMD), which corresponds to an excess risk of 10%. The power model had a power restriction of ≥ 1 . Selection of Linear and Polynomial 2° was based on choosing a model that describes the data with the least complexity. A Hill model was excluded if the “k” parameter of the model was less than 1/3 of the lowest positive dose as per (Black, et al. 2012). The datasets were mapped to Ingenuity Core Pathways using the Defined Category Analyses feature. The pathway

dataset was downloaded from Ingenuity Pathway Analysis (December 13, 2012) and array annotation dataset, MM_WG_GPL10333, was downloaded from NCBI Gene Expression Omnibus (October 18, 2011 version).

2.4.8 Margin of Exposure

The MoE is a metric used to infer public health risk. Typically a MoE of 10,000 or greater indicates that a compound is not a public health concern (EFSA 2005, Benford *et al.* 2010), however 1000 is the benchmark used for furan since it is a relatively well characterized chemical and therefore has a lower uncertainty factor associated with it (IRIS 2012). In this study MoEs were calculated by dividing the relevant BMDL ($\mu\text{g}/\text{kg}$ bw) by the actual environmental exposure ($\mu\text{g}/\text{kg}\text{-bw}/\text{day}$). Average (50th percentile) and high (90th percentile) environmental exposure estimates for USA or Europe, and for infants or combined child/adult populations were taken from Carthew *et al.* (2010b). Apical MoEs were calculated for HCC and HCA, with tumour BMDL values calculated using cancer rates published by Moser *et al.* (2009) and NTP (1993). Transcriptional MoEs were calculated for the NRF2 Oxidative Stress Response IPA pathway and for a ‘General Transcriptional Response,’ the latter was calculated using the average BMDL for all enriched pathways with at least five molecules and $P \leq 0.05$. We use the BMDL value obtained from the best model (as described above) for MoE calculations.

2.5 Results

2.5.1 Cellular Proliferation (BrdU), Cytotoxicity, and Cancer

Treatment-related effects of 0, 1, 2, 4, or 8 mkd furan for three weeks on mice (including measurement of serum enzymes and histopathological markers of cytotoxicity and inflammation) have been previously reported (NTP 1993, Moser *et al.* 2009, Gill *et al.* 2011) and therefore were not repeated in this study. Body weight was not affected by furan exposure. BrdU incorporation was observed to increase with increasing furan dose. Labeling indices for BrdU of 2.3, 2.4, 2.9, 5.2, and 7.6% at 0, 1, 2, 4 and 8 mkd furan, respectively, were observed, with a significant increase in BrdU incorporation at 8 mkd ($P = 0.00413$) (Fig. 2.1A). Similar results for BrdU incorporation were previously reported in (Moser *et al.* 2009).

Previous studies have reported increases in cytotoxicity and proliferative histopathological markers and cancer rates in response to furan (Moser *et al.* 2009, Gill *et al.* 2011). We modeled dose responses for these apical endpoints and our own BrdU incorporation data using BMDS (Table 2.1) and observed that BMD(L) variability between studies was low for cancer endpoints, but was quite large for cytotoxicity and proliferation markers.

2.5.2 Gene Expression Analysis

Following furan treatment, the number of differentially expressed genes increased in a dose-dependent manner with 2, 17, 27 and 339 genes (2, 18, 43 and 441 probes) in the 1, 2, 4, or 8 mkd dose groups, respectively (Fig. 2.1B). While fold changes ranging between +20.65 and -6.60 were observed, the majority of significantly changed probes

had fold changes of ± 1.5 -2.0. There was a complete overlap between the differentially expressed genes in the 4 and 8 mkd dose groups, and only one shared gene between the 2 and 4 mkd dose groups thus indicating a clean transition in the nature of the transcriptional response. A PCA analysis revealed a dose-related clustering of hepatic expression profiles (Fig. 2.1C). We saw up-regulation of a number of genes consistent with our proposed MoA for furan including genes involved in oxidative stress response, xenobiotic metabolism, inflammation, cell cycle arrest/cell death and cell survival and growth (Table 2.2). Using BMDExpress, we calculated median BMD(L) values for individual probes. We observed that many BMD(L) values for genes thought to be important for the furan MoA were close to the 2-4 mkd transitional dose range, and this trend persisted regardless of the magnitude of the fold change of the probes used in the modelling (Table 2.2).

2.5.3 Cluster Analysis

Cluster analysis was performed using gene lists from 1, 2, 4 and 8 mkd furan dose groups and gene expression datasets from GEO from spontaneous hepatocellular carcinoma, liver regeneration following partial hepatectomy, liver inflammation, and chemical exposure to hepatocarcinogens, non-hepatocarcinogens and Cyp2E1 ligands (acetaminophen, APAP, and carbon tetrachloride, CT) (Fig. 2.2). Furan dose groups clustered closely together, with a split between the lower dose groups and the 8 mkd dose group. The 8 mkd furan dose group clustered most closely with a HCC dataset, followed by liver regeneration 38 and 48 hours post-partial hepatectomy (PH). The 1, 2 and 4 mkd furan dose groups clustered with a dataset representing NF-kappaB-mediated liver

inflammation. Together, these form a cluster that splits from the Cyp2E, all of which split from a large cluster composed primarily of hepatocarcinogens.

2.5.4 Pathways and Upstream Regulators of Transcription Analyses

Pathway analysis was employed to provide information regarding the furan molecular MoA. The number of enriched pathways increased in a dose-dependent manner. There were no enriched pathways at the lowest dose (1 mkd) and pathway enrichment at the next two doses (2 and 4 mkd) was always based on three or fewer genes, therefore we focused on pathways enriched at the highest dose (8 mkd). Pathways relevant to the furan MoA included NRF2 Oxidative Stress Response, Glutathione Metabolism, Xenobiotic Metabolism, Death Receptor Signaling, TNF-alpha/ASK1/JNK1 Apoptosis Signaling (within the 14-3-3 Signaling Pathway), ERK/MAPK Signaling and Breast Cancer Regulation by Stathmin1 (Fig. 2.3). In addition, enrichment of the p53 Signaling pathway indicated that the DNA damage response could be involved. Other pathways, including the Xenobiotic Metabolism pathway and the Aryl Hydrocarbon Receptor (AhR) pathway, were also enriched in response to furan. While the Xenobiotic Metabolism pathway comprises CAR, PXR, AhR, PPAR and NRF2 signaling, NRF2 was the only one among these that was appreciably activated. Closer inspection of the AhR pathway showed that enrichment of this pathway was almost exclusively the result of an increase in the expression of the estrogen receptor (*Esr1*), and not the AhR. These findings are consistent with previous studies showing that furan does not undergo receptor-mediated metabolism. BMD modeling for individual pathways demonstrated that BMDL values fell around the 2-4 mkd transitional dose range (Fig. 2.3). For all IPA

pathways (that could be modeled and that had at least five genes with significant changes in expression) the average pathway BMDL mean was 2.2 mkd (range 1.5-4.4 mkd).

Upstream regulators of transcription were predicted using IPA software. There were no predicted upstream regulators for the 1 or 2 mkd dose groups. TNF-alpha was the only predicted upstream regulator in the 4 mkd dose group. There were many predicted upstream regulators for the 8 mkd dose group; in particular, activation of ROS, hydrogen peroxide, Cyp2E1 ligands (CT, APAP, EtOH), cytokines (including TNF-alpha), growth factors, mitogen-activated protein kinases (MAPKs, including SAPKs and ERKs), transcription factors (including: NRF2/NFE2L2 and NF-kappaB1), and hormones (including estrogen and thyroid hormone) were predicted (Fig. 2.3).

2.5.5 Disease and Chemical Prediction

Meta-analyses comparing the changes in gene expression observed in the 8 mkd dose group to other gene expression studies was conducted using the NextBio Human Disease Atlas and revealed similarities to the following: Injury to Liver (97.4), Liver Regeneration (91.7), Hepatic Fibrosis (82.25), Hepatocellular Dysplasia (80.9), Liver Cancer (73.0), Inflammatory Disease of the Liver (68.5), and Cirrhosis of the Liver (52.07). Full list is available in Appendix C, Table C1.

The NextBio Chemical Atlas was used to identify similarities were observed with chloroform (47.42), CT (43.02), hydrogen peroxide (40.75), and APAP (39.12). Interestingly, 8 mkd furan was found to produce gene expression changes that were most similar to the insecticide malathion (100). Full list is available in Appendix C, Table C2.

2.5.6 GO Terms

There were many, redundant GO terms enriched for that were relevant to furan's primary MoA. Interestingly, mean BMDL values approached the expected 2-4 mkd range, including: kinase activity (GO:0016301; BMDL 1.9 mkd), regulation of transcription, DNA dependent regulation of transcription (GO:0006355; BMDL 2.1 mkd), regulation of protein modification process (GO:0031399; BMDL 1.9 mkd), regulation of apoptosis (GO:0042981; BMDL 1.9 mkd), negative regulation of apoptosis (GO:0043066; BMDL 2.2 mkd), and regulation of gene expression (GO:0010468; BMDL 2.0 mkd). One genotoxicity-relevant GO term, response to DNA damage stimulus (GO:0006974; BMDL 2.1 mkd), was also enriched for. Full list is available in Appendix C, Table C3.

2.5.7 Margin of Exposure

Two MoEs each were calculated for apical (HCC and HCA) and transcriptional (NRF2 Oxidative Stress Response pathway and General Transcriptional Response) endpoints. Estimates of risk calculated for apical or transcriptional endpoints were generally quite consistent (Table 2.3).

2.6 Discussion

Global transcriptomic analyses on liver tissue collected from mice exposed sub-chronically to non-carcinogenic (0, 1, 2 mkd) and carcinogenic (4 and 8 mkd) doses of furan were carried out to provide qualitative and quantitative mechanistic insight into how furan causes liver cancer. The most common MoA for the development of liver

cancer is chronic cytotoxicity and inflammation leading to regenerative proliferation (Nakagawa & Maeda 2012). This MOA is also hypothesized to underlie carcinogenicity induced by other chemicals including the group 2B non-genotoxic carcinogens, including carbon tetrachloride (Manibusan *et al.* 2007) and chloroform (Larson *et al.* 1994, Templin *et al.* 1996). Furan is believed to cause cancer by this MoA (Fransson-Steen *et al.* 1997, Moser *et al.* 2009); however, it is not clear how furan stimulates this process on a molecular level. Our case study supports the hypothesis that furan's primary MoA is cytotoxicity followed by cellular proliferation and regeneration. The analysis also supported potential genotoxicity occurring at the high dose via indirect mechanisms (i.e.: ROS generation). Our discussion focuses primarily on the non-genotoxic MoA; an important outcome of this case study is the demonstration that sub-chronic *in vivo* toxicogenomics studies can be used to discern MoA information for non-genotoxic carcinogens, and that these data can be used to inform quantitative human health risk assessment. This is an important finding because non-genotoxic carcinogens aren't detected by short-term mutagenicity tests, and the length and cost of the two-year rodent bioassay is prohibitive.

In this study we observed an increase in cellular proliferation after three weeks of furan exposure at both of the carcinogenic doses examined (Fig. 2.1A), which confirmed the induction of the furan MoA and was consistent with previously published results (Moser *et al.* 2009). Although no measurements for cytotoxicity were taken in our study, this *in vivo* response to furan is well-characterized elsewhere (NTP 1993, Moser *et al.* 2009, Gill *et al.* 2011), as is the *in vitro* cytotoxicity of BDA (Kellert *et al.* 2008). We also noted that all of the differentially expressed genes at 4 mkd were contained within

the 8 mkd gene list and were distinct from those observed at 2 mkd, indicating that a shift from an adaptive/stress response to an adverse/cancer response had presumably occurred (Fig. 2.1B). The molecular MoA of the adverse response to furan was best characterized by relying on the carcinogenic 8 mkd dose group, since it had the most populated list of differentially expressed genes, it clustered with liver regeneration and liver cancer datasets, and clustered separately from the lower dose groups (Fig. 2.1C; Fig. 2.2). This separate clustering suggests that the 8 mkd livers had progressed further toward the HCC and HCA disease molecular phenotype during the three week exposure than the (also carcinogenic) 4 mkd dose group. Another possibility is that the MoA at 4 mkd is different from that at 8 mkd; however, there is limited evidence from our data or others' to support this hypothesis. We propose a molecular mechanism for furan that accounts for furan-induced increases in liver cytotoxicity, inflammation and proliferation, leading to tumours, and is anchored in the furan, Cyp2E1, liver regeneration, and spontaneous or chemically induced HCC literature. Briefly, the molecular initiating event is furan metabolism by CYP2E1. The resultant production of BDA and ROS bring about changes in the cell that trigger cytotoxicity and oxidative stress, which lead to inflammation. Regenerative proliferation follows liver injury; however, when paired with chronic elevation of NRF2 oxidative stress response and chronic inflammation, a transition point is created that tips the balances toward sustained cellular proliferation. Tumourigenicity occurs when the dose of furan results in a level of BDA and ROS production that exceeds the rate at which the cell can neutralize them, thereby requiring chronic elevation of these (normally protective) cellular pathways. The result is sustained and dysregulated cellular proliferation that ultimately leads to cancer (Figs. 2.3, 2.4).

In the larger context of spontaneous and chemically-induced liver cancer, many potential molecules and signaling pathways are hypothesized to be responsible for carcinogenic transformation during chronic liver injury. Prominent candidates, that are also implicated in the present study, include tumour necrosis factor alpha (TNF-alpha which is a cytokine), c-Jun NH2-terminal kinase 1 (JNK1, which is a SAPK), and nuclear factor kappa B (NF-kappaB) and c-Jun proto-oncogene/activator protein 1 (c-Jun/AP-1), which are transcription factors. Indeed, previous studies of liver regeneration have also identified TNF-alpha, JNK1, c-Jun and NF-kappaB as required for liver regeneration. In this study we obtained support for their involvement, as well as many other gene products and pathways including death receptor 5 (DR5), NRF2 oxidative stress response, extracellular signal regulated kinase (ERK) signaling, and c-Myc and Statmin1 oncogenes in the development of HCC and HCA in sub-chronically exposed mouse liver. All of these are thoroughly explored below and are summarized in Figure 2.3.

2.6.1 Furan MoA: Oxidative Stress

Furan metabolism to BDA by CYP2E1 has been reported to produce ROS and deplete GSH (Carfagna *et al.* 1993, Gonzalez 2005, Lu & Cederbaum 2008). We thus examined our transcriptomic data for evidence of induction of the oxidative stress response. First, we noted that ROS and hydrogen peroxide were predicted upstream regulators. Second, ROS producing CYP2E1 ligands (APAP, CT, and EtOH) were also predicted to be activated, which indicates that the downstream effects of their metabolism are likely comparable to those of furan. In addition, meta-analysis of gene expression data shows similarity in furan-induced changes in genes expression to those induced by

hydrogen peroxide, carbon tetrachloride, and acetaminophen (Fig. 2.2). Third, Enrichment of the Glutathione Metabolism pathway and increased expression of relevant genes including glutamate-cysteine ligase, catalytic subunit (*Gclc*) and glutathione-s-transferases (*Gsta1* was the most highly up-regulated gene), all indicated that ROS production, and GSH synthesis and depletion are occurring. Finally, the most important indication of oxidative stress was the enrichment of the NRF2 Oxidative Stress Response pathway and prediction of NRF2 activation. ROS can directly modify protein activity by oxidizing disulfhydryl groups on redox-sensitive proteins to disulfides. A canonical example of this is the oxidation of Kelch-like ECH protein 1 (KEAP1), which can only bind to and block NRF2 in its reduced form. Reduced KEAP1 holds NRF2 in the cytoplasm with the Cullin-3 (CUL3) ubiquitin ligase, which targets NRF2 for proteosomal degradation. When KEAP1 is oxidized, NRF2 enters the nucleus and alters gene expression of anti-oxidant and metabolic genes. We see up-regulation of a number of downstream genes committed to redox homeostasis and xenobiotic metabolism (Table 2.2). Therefore our toxicogenomics data clearly support that hepatic, ROS-associated toxicity occurs in the mouse model following exposure to furan.

Clearing ROS from the cell is a top priority as, in addition to modifying protein function by redox, ROS can damage cellular macromolecules (such as DNA). This damage, when severe, leads to cellular cell death. Two cellular systems for neutralizing ROS are thioredoxin and peroxiredoxin, which are involved in redox repair of oxidized cellular proteins and detoxification of hydrogen peroxide, respectively. NF-kappaB and AP-1 are both redox-sensitive transcription factors that require reduction by thioredoxin for DNA binding (Arnér & Holmgren 2000). Up-regulation of the NRF2-target-genes

thioredoxin reductase (*Trxnd1*) and sulfiredoxin (*Srxn1*) in our furan-exposed samples suggest that both of these redox pathways are active (since TRXND1 and SRXN1 reduce oxidized thioredoxin and peroxiredoxin, respectively). NRF2 control over maintenance of cellular redox state is normally cytoprotective; however, chronic elevation of NRF2-mediated gene expression has been shown to be detrimental and even to confer a survival advantage to pre-cancer and cancer cells (Hayes & McMahon 2006, Lau *et al.* 2008). We propose that this dual nature of NRF2 activation might be a tipping point between normal cytotoxicity/regeneration cycles and dysregulated ones. Moreover, chronic exposure to a ROS-producing agent, such as furan, could tip the balance by causing chronic KEAP1-inactivation/NRF2-activation and sustained maintenance of downstream effects (Fig. 2.4).

2.6.2 Furan MoA: Cytotoxicity

Cytotoxicity and liver injury following chemical exposure occur first in furan's proposed MOA. Our transcriptomic data indicates that cellular damage is occurring in our furan-exposed liver tissues. Meta-analysis of gene expression produced similar changes as the following disease states: Injury to Liver, Hepatic Fibrosis and Cirrhosis of the Liver. In addition we see enrichment of GO terms for Apoptosis and Regulation of Programmed Cell Death. Enrichment of the Arachidonic Acid Metabolism pathway suggests that Cyp2E1-catalyzed lipid peroxidation of polyunsaturated fatty acids likely contributes to cytotoxicity through apoptosis (Chen & Cederbaum 1998). Upstream regulator prediction is a powerful tool that makes predictions regarding molecule activation or inhibition based on downstream gene expression. TNF-alpha was the only

predicted upstream regulator in response to 4 mkd furan and was among many predicted for the 8 mkd dose group. As such, we believe that TNF-alpha must be among the earliest cytokines to respond to furan exposure. TNF-alpha is a cytokine that signals through its receptor, TNF receptor 1 (TNFR1). During EtOH metabolism by CYP2E1, oxidative stress promotes cell death by converting the (normally mitogenic) TNF-alpha signal to a cytotoxic signal and inducing mitochondrial permeability transition (MPT) (Pastorino & Hoek 2000, Lu & Cederbaum 2010). MPT occurs when the MPT pores in the mitochondrial membranes open, leading to membrane depolarization and uncoupling that causes ATP depletion and eventually necrosis. MPT pore opening can also result in mitochondrial swelling, leading to mitochondrial rupture, cytochrome c release, caspase activation and apoptosis. MPT is triggered by a variety of stimuli (e.g., various xenobiotics, Bcl-2-associated X protein (BAX) over-expression, disulfide formation, increased mitochondrial Ca^{2+} , GSH depletion, TNF-alpha signaling, and ROS) and the outcome (necrosis versus apoptosis) is dependent on the biological context but is usually a combination of the two (Jaeschke *et al.* 2002). Further evidence to support the important role of the TNF-alpha superfamily (SF) in the cellular response to furan is the enrichment of the Death Receptor (DR) pathway based on up-regulation of Death Receptor 5 (*DR5*, aka *Tnfrsf10b*; Table 2.2), which signals for cell death through TNFR-associated death domain (TRADD), Fas-associated death domain (FADD), and the caspase cascade.

SAPK phosphorylation cascades can also signal for MPT (Ashkenazi 2002, Yuan & Kaplowitz 2009, Chu 2013). During APAP metabolism by CYP2E1, ROS production and GSH depletion lead to the activation of apoptosis signaling kinase (ASK1). The

ensuing ASK1, MAP2K4, JNK1 phosphorylation cascade leads to activation of BAX, thereby causing hepatocyte cell death by MPT (Seki *et al.* 2012). Interestingly, when the cell regains redox homeostasis, reduced thioredoxin can block ASK1 (Saitoh *et al.* 1998), and thereby inhibit this cell death pathway (Fig. 2.3). Enrichment of 14-3-3 Signaling (which has a ASK1, JNK1 component), prediction of activation of the MAPK group and MAP2K4 (Fig. 2.3), enrichment of GO terms for protein kinase activity and regulation of protein phosphorylation, and up-regulation of *Bax* and *Dusp8* (a JNK-specific phosphatase) (Table 2.2, Fig. 2.3) all indicate that cytotoxicity from increased MPT (as per Seki *et al.* 2012) is relevant to furan. Another JNK1 substrate, c-Jun, is a monomer of the heterodimeric transcription factor AP-1. AP-1 has been described as a ‘double-edged sword’ in tumourigenesis because, depending on dimer composition and abundance, and the oxidation state of the cell, it has been shown to have either oncogenic or tumour suppressor capabilities (Eferl & Wagner 2003, Hess *et al.* 2004). C-Jun’s contribution is typically understood to be pro-proliferative; however, during periods of CYP2E1 activity and oxidative stress, c-Jun and JNK1 are also known to be involved in hepatocyte cell death (Liu *et al.* 2002, Singh & Czaja 2007, Singh *et al.* 2009, Amir *et al.* 2012). Indeed, loss of c-Jun has been shown to prevent the CYP2E1-induced TNF-alpha switch from proliferative to apoptotic signaling (Liu *et al.* 2002). The exact mechanisms by which c-Jun functions during periods of oxidative stress compared to normal cellular conditions are under investigation (Karin & Shaulian 2001, Amir *et al.* 2012). The combination of CYP2E1 activation, ROS production and up-regulation of *c-Jun* and *Dusp8* suggest that the oxidative stress- and JNK1/c-Jun-mediated cell death mechanism (as per Amir *et al.* 2012) is likely to also be relevant to furan.

2.6.3 Furan MoA: Inflammation

Inflammation accompanies cell death and is triggered by pro-inflammatory cytokines, released by Kupffer cells (liver-specific macrophages), and molecular signals released from dying cells (Nakagawa & Maeda 2012). Although acute inflammation is a normal part of the innate immune response, chronic inflammation is thought to link the cytotoxicity/regeneration MoA with HCC. This link is thought to be the result of an increase in pro-proliferative cytokines, pro-angiogenic factors and transcription factors with pro-proliferative target genes, particularly NF-kappaB (Berasain *et al.* 2009, Karin 2009, DiDonato *et al.* 2012). TNF-alpha can signal to activate NF-kappaB, which is important for initiation of both inflammation and hepatocyte proliferation (Karin 2006, Nakagawa & Maeda 2012). NF-kappaB is a dimeric transcription factor that is typically retained in the cytoplasm by its inhibitor (IkappaB); however, up-regulation of IkappaB kinase (IkappaBK, which has three subunits: alpha, beta, and gamma) allows NF-kappaB to enter the nucleus and modify gene expression. NF-kappaB activation in Kupffer cells leads to production of pro-inflammatory cytokines, growth and survival signals, and angiogenic factors (Karin 2006). Kupffer cell-specific NF-kappaB inactivation (by IkappaBK gamma subunit knock-out) has been shown to reduce HCC tumour load, presumably by reducing production of inflammatory signals (Maeda *et al.* 2005, Berasain *et al.* 2009). NF-kappaB activation in hepatocytes leads to up-regulation of cell-cycle and anti-apoptotic genes (Karin 2006) and its inhibition prevents tumour development and triggers apoptosis of initiated hepatocytes (Pikarsky *et al.* 2004). In the presence of ROS, NF-kappaB and JNK1 are known to antagonize one-another (Papa *et al.* 2004, Zhang & Chen 2004, Nakano *et al.* 2006). In addition, NF-kappaB itself is a redox sensitive

protein (Yuan & Kaplowitz 2009). While the details of this molecular cross-talk are still being investigated, it seems reasonable to assume that if NF-kappaB activity were to get the ‘upper-hand’ it could tip the balance away from JNK1-mediated cytotoxicity and toward NF-kappaB-mediated inflammation and proliferation.

We expect that NF-kappaB is transcriptionally active in response to furan since we see up-regulation of IkappaBK gamma subunit (Table 2.2) and the NF-kappaB monomer, NF-kappaB1, is predicted to be activated (Fig. 2.3). In addition, our 1, 2 and 4 mkd dose groups each clustered most closely with an NF-kappaB-mediated liver inflammation dataset (GSE20427; Fig. 2.2) and our 8 mkd dose group is correlated with the ‘Inflammatory Disease of the Liver and Inflammatory Disorder’ phenotypes (by NextBio). These data are consistent with the notion that that NF-kappaB-mediated inflammation in pre-cancer cells is an important part of setting the stage for malignant transformation in furan exposed liver (but may not be sufficient for transformation on its own since 1 and 2 mkd doses furan do not cause HCC). We observed up-regulation of the pro-inflammatory chemokine C-X-C motif ligand (*Cxcl10*; Table 2.2), and prediction of activation of other cytokines (IL-1a, IL-1b, IL-6, IL-8, TNFSF11, TNF-alpha, CXCL12, CCL5, IFN-gamma) and of Triggered Receptor Expressed on Myeloid Cells 1 (TREM1) (Fig. 2.3). Kupffer cell expression of TREM1 has been shown to be required for HCC development; further, loss of TREM1 results in down-regulation of transcription of pro-inflammatory and pro-proliferative cytokines (IL-6, IL-beta1, TNF-alpha, CCL2 and CXCL10) and loss of activation of ERK1/2, JNK and NF-kappaB, ultimately preventing malignant transformation (Wu *et al.* 2012). Our data support the idea that inflammation precedes and facilitates carcinogenic transformation in chemically induced HCC. These

findings are consistent with previous studies that have demonstrated the chronic inflammation phenotype in furan-exposed mice and rats (NTP 1993, NTP 1993, NTP 1993, Moser *et al.* 2009).

2.6.4 Furan MoA: Regenerative Proliferation

Regenerative proliferation is the penultimate stage to tumour formation in the furan MoA. Molecular mechanisms that underlie hepatocyte proliferation have been studied using liver development, liver regeneration and spontaneous HCC models. NF-kappaB and c-Jun are both required for liver development. Loss of NF-kappaB (by knock-out of the beta or gamma IkappaBK subunits) is embryonic lethal because of extensive hepatocyte apoptosis and liver degeneration (Li *et al.* 1999, Rudolph *et al.* 2000). C-Jun-null mice are also embryonic lethal as a result of impaired liver and heart development (Eferl *et al.* 1999, Jochum *et al.* 2001, Behrens *et al.* 2002). Liver regeneration has been extensively studied in animals following 2/3 hepatectomy. Both NF-kappaB and c-Jun are not normally expressed/activated in adult liver, but are up-regulated following PH in a TNF-alpha/TNFR1-dependent manner, and are required for hepatic regeneration (Alcorn *et al.* 1990, Hilberg *et al.* 1993, Cressman *et al.* 1994, Diehl *et al.* 1994, FitzGerald *et al.* 1995, Yamada *et al.* 1997, Chaisson *et al.* 2002). Rats lacking NF-kappaB display impaired liver regeneration due to hepatocyte apoptosis (Schrum *et al.* 2000) and mice lacking AP-1 also show impaired liver regeneration, often leading to death (Behrens *et al.* 2002). Similarly, in HCC TNF-alpha signals activation of both c-Jun/AP-1 and NF-kappaB, which up-regulate expression of pro-survival and pro-proliferation genes. Increased NF-kappaB expression in HCC cells is important for the

transition from liver injury and inflammation to cancer (Karin 2009, Luedde & Schwabe 2011) and for HCC cell growth (Wang *et al.* 2003, Pikarsky *et al.* 2004). Recently, up-regulation of c-Jun gene expression in the early stages of liver cancer has been shown to be required for HCC initiation and development (Min *et al.* 2012). Both c-Jun and NF-kappaB have been proposed as molecular targets for liver tumour prevention (Young *et al.* 2003).

Meta-analysis of our transcriptomic data demonstrates similarity in gene expression changes produced by 8 mkd furan to those generated during liver regeneration, hepatocellular dysplasia, and liver cancer. Our 8 mkd dose group clustered most closely with the spontaneous HCC and liver regeneration datasets (Fig. 2.2). This is remarkable since our exposure was sub-chronic (therefore cells had presumably not yet undergone malignant transformation). We observed up-regulation of growth factors (*Pdgf* and *Gdf9/15*; Table 2.2) and prediction of activation of others (VEGF, HGF, IGF1, EGF, FGF, BMP2, BMP6; Fig. 2.3) in response to furan. Growth factors are secreted proteins that bind their tyrosine receptor kinase and relay signals through the cell via ERKs. The MAPK phosphorylation cascade follows the order of RAS (GTPase), RAF (MAP3K), MEK1/2 (MAP2K), and ERK1/2 (MAPK), all of which are predicted upstream activators as is the MAPK/ERK Signaling pathway. ERKs were first identified on the basis of being required for the G1/S transition (Meloche & Pouyssegur 2007) and are known to activate at least 160 substrates (Min *et al.* 2011), which go on to alter target gene expression. One target gene is dual-specific phosphatase 6 (*Dusp6*), which is up-regulated in response to furan (Table 2.2) and feeds back to reset the pathway by dephosphorylating ERK1/2. During periods of chronic oxidative stress, the ERK1/2 pathway is protective against

hepatocyte death (Singh & Czaja 2007); this could be due to the fact that DUSPs are oxidation-inactivated (Patterson *et al.* 2009). We suggest that ERK signaling is responsible for promoting hepatic cellular proliferation during liver regeneration.

In our samples, ERK signaling might promote cellular proliferation via activation of protein substrates including: c-Jun/AP-1, c-Myc, and Stathmin1. c-Jun has been shown to promote HCC by preventing hepatocyte apoptosis through inhibition of p53 tumour suppressor (Eferl & Wagner 2003) and SIRT6 (Sirtuin-6; a stress-responsive deacetylase). For the latter, c-Jun blocks up-regulation of Sirt6, which results in increased levels of Survivin (aka: Birc5) thereby blocking apoptosis (Min *et al.* 2012). Survivin is marginally up-regulated by 8 mkd furan. Two other AP-1 monomers were up-regulated by furan: activating transcription factors 3 and 4 (*Atf3/4*; Table 2.2). ATF3 is up-regulated during the stress response (Hai *et al.* 1999) and has been implicated in oncogenesis when bound to c-Jun (Thompson *et al.* 2009). ATF4 is up-regulated in cancer and is known to regulate genes involved in oxidative stress response, differentiation, metastasis and angiogenesis (Ameri & Harris 2008). C-Myc oncogene is a transcription factor that is associated with many cancers (Dang 2012) and has been shown to contribute to HCC progression during chronic EtOH exposure and CYP2E1 activation (Tsuchishima *et al.* 2013) and to up-regulate NRF2 expression during tumourigenesis (Denicola *et al.* 2011). Finally, the Breast Cancer Regulation by Stathmin1 pathway is enriched in both the 4 and 8 mkd furan dose groups. Stathmin1 (*Stmn1*; also known as Oncoprotein18, Op18) is a microtubule remodeling protein that is not actually specific to breast tissue/breast cancer (*Stmn1* and various *Tubs* are up-regulated by furan; Table 2.2). Indeed, it has been shown to be involved in both liver

regeneration (Okazaki *et al.* 1993, Rowlands *et al.* 1995) and HCC. Up-regulation of Stathmin1 in HCC is correlated with poor prognosis and tumour reoccurrence (Yuan *et al.* 2006, Hsieh *et al.* 2010). Silencing of Stathmin1 in an HCC cell line (HCCLM3) suppressed proliferation, induced apoptosis and impaired metastasis (Gan *et al.* 2010). ERKs likely direct hepatic cellular proliferation via activation of other substrates as well; however we believe that *c-Jun*, *c-Myc*, and *Stmn1* likely play important roles during this process.

2.6.5 Furan MoA: Summary

Taken together, we propose that all of the above molecules and pathways (summarized in Figure 2.3) are important contributors to furan-induced regenerative proliferation leading to HCC. We also acknowledge that, in order for HCA and HCC to develop, genomic damage must occur. We propose that furan-induced genomic damage occurs indirectly as a by-product of its metabolism (particularly the oxidative stress and cytotoxicity aspects of the MOA). This hypothesis is based on our observed induction of oxidative stress responses in combination with the mixed genotoxicity findings (discussed in the introduction; occurring at high doses). In support of the occurrence of DNA damage at high doses, we saw enrichment of the p53 Signaling pathway in response to 8 mkd (but not 4 mkd). This pathway enrichment was due to the up-regulation of the following genes: *Cdkn1a*, *Cyclin G1*, *Bax*, *DR5*, and *Teap1* (p53 inducible nuclear protein 1), which are all downstream targets of p53 that are suggestive of DNA damage response.

2.6.6 Margin of Exposure

In addition to applying our transcriptomics data to derive MoA information for furan, another objective of this work was to demonstrate the practical application of toxicogenomics in quantitative risk assessment. We accomplished this by calculating BMD and MoE values for apical and transcriptional endpoints since demonstration of a strong correlation between transcriptomic and apical endpoints will facilitate the implementation of toxicogenomics in routine risk assessments (Thomas, et al. 2011). Strikingly, the majority of the transcriptional BMD modeling that was conducted placed most values near to the expected 2-4 mkd PoD range (Table 2.2, Fig. 2.3) as did modeling of HCC and HCA cancer endpoints reported by Moser et al. (2009) and the NTP (1993) (Table 2.1). It is worth noting that, while the transcriptional BMD(L) values were often modeled using genes that had fold changes of less than ± 2.0 , the most significantly enriched pathways and functions tended to also contain genes with larger fold changes. In addition, the magnitude of the fold change of the individual probes did not appear to impact the magnitude of the BMD(L) values (Table 2.2). Modeling of previously-studied cytotoxicity and proliferation apical endpoints showed more variability. BMD values calculated from apical data for male and female B6C3F1 mice published by Gill et al. (2011) were often below the 2 mkd cancer threshold and resembled their NOAEL value (0.12 mkd) for furan. Alternatively, BMD values calculated using apical data published in the NTP's (1993) rodent cancer bioassay were much higher than the 2-4 mkd range. The most sensitive cytotoxicity apical endpoint was hepatocyte apoptosis in the caudate lobe (BMDL: 0.1 mkd), and the most sensitive

proliferation apical endpoint was hepatocyte basophilia (BMDL: 0.3 mkd) (Table 2.1). Both of these underestimate the known cancer PoD. The BMDLs for liver cancer more closely approximated cancer PoD with the HCA and HCC BMDLs equal to 0.9 and 1.6 mkd, respectively (Table 2.1). Transcriptional (pathway) BMDLs were within the same range of the cancer BMDLs (1.5-4.4 mkd), with an average pathway BMDL of 2.2 mkd. The two pathways that we believe are very important to the furan MoA, NRF2 Oxidative Stress Response and ERK/MAPK Signaling, have very similar mean BMDLs of 2.3 and 2.6 mkd, respectively (Fig. 2.3). Strikingly, after only a three week exposure, both the individual and the all-pathway-average BMDL values perfectly reflect the 2-4 mkd cancer PoD range.

We used the cancer and pathway BMDL values to estimate human health risk using MoE values. Pathway MoE values were calculated using either the NRF2 Oxidative Stress Response Pathway BMDL, or the all-pathway-average BMDL. Pathway and HCA MoEs calculated for transcriptional or apical endpoints were generally equivalent (Table 2.3). While it is still early days and this work should be repeated using other chemicals, we feel that this concordance between individual and all-pathway-average BMDL (and therefore MoE) is an interesting and important observation and tentatively suggest that an all-pathway-average MoE might be a useful screening tool for initial estimation of chemical risk.

2.6.7 Conclusions

In summary, using the MoA and the risk assessment information gathered in this study we assembled a simplified adverse outcome pathway for furan (Fig. 2.4). We

believe that chronic exposure to >2 mkd furan hijacks the normal response to manageable levels of oxidative stress-induced damage thereby tipping the balances in favor of chronic regenerative proliferation. The molecular initiating event is Cyp2E1 ligand binding and furan metabolism to produce ROS. The key events that follow are: (1) cytotoxicity, (2) activation of the NRF2 oxidative stress response, (3) inflammation, and (4) regenerative proliferation, where (2) and (3) represent tipping points between the adaptive and the adverse outcomes. There is a large body of literature supporting the idea that chronic NRF2 activation is carcinogenic, and it is known that the PoD for furan is between 2-4 mkd. Therefore, we propose that the NRF2 oxidative stress response pathway BMDL (2.3 mkd) would denote the best transcriptional benchmark dose for the PoD between the adaptive and adverse responses to furan. Above this dose the cytoprotective and defensive roles of NRF2 are hijacked in premalignant and malignant cells to confer a survival advantage, and this is exacerbated by chronic inflammation and up-regulation of other oncogenes (including *c-Jun*, *c-Myc* and *Stmn1*). While our molecular MoA should be validated by additional experiments, we feel that it provides realistic first insight into how furan causes cancer. We believe that our toxicogenomic approach is illustrative of how gene expression data can provide MoA information for indirectly acting and non-genotoxic carcinogens that is reflective of apical cancer data. In addition, we believe that the consistency between apical and genomic data demonstrates the utility of this approach for the investigation of other non-genotoxic carcinogens that do not have apical or two-year rodent bioassay data associated with them.

2.7 Tables and Figures

Table 2.1 BMD modeling of previously published apical endpoint data for furan exposures in B6C3F1 mice. For cytotoxicity and proliferative markers Gill et al. (2011) exposed male and female mice to 0, 0.03, 0.12, 0.5, 2 and 8 mkd furan for 13 weeks and the NTP (1993) exposed female mice to 0, 2, 4, 8, 15, 30 and 60 mkd furan for 13 weeks. For cancer studies, Moser et al. (2009) exposed female mice 0, 0.5, 1, 2, 4 and 8 mkd furan for two years and the NTP (1993) exposed female mice to 0, 8 and 15 mkd furan for two years; cancer incidences in these studies are in brackets.

Apical endpoint	BMD (mkd)	BMDL (mkd)
<i>Cytotoxicity markers</i>		
<u>Study: Gill et al. 2011</u>		
Serum alanine transaminase (ALT)	1.8	1.4
Hepatocyte apoptosis (caudate lobe)	0.4	0.1
Hepatocyte apoptosis (other lobes)	1.4	0.3
Kupffer cell pigmentation and inflammation	1.2	0.3
<u>Study: NTP, 1993</u>		
Hepatocyte degeneration	19.9	14.7
Hepatocyte necrosis	23.5	14.4
Kupffer cell pigmentation	23.5	14.4
Cholangiofibrosis	27.2	17.6
<i>Proliferation markers</i>		
<u>This study</u>		
BrdU labeling index	1.1	0.7
<u>Study: Gill et al. 2011</u>		
Hepatocyte basophilia	0.6	0.3
Biliary tract hyperplasia	5.8	1.1
<u>Study: NTP, 1993</u>		
Biliary tract hyperplasia	24.6	14.5
Hepatocyte cytomegaly	19.9	14.7
<i>Liver Cancer</i>		
<u>Study: NTP, 1993</u>		
HCA (5/50, 31/50, 48/50)	2.8	0.9
HCC (2/50, 7/50, 27/50)	6.6	5.3
HCA or HCC (7/50, 34/50, 50/50)	6.9	4.2
<u>Study: Moser et al. 2009</u>		
HCA (3/36, 4/72, 4/53, 4/41, 11/36, 25/39)	2.3	1.3
HCC (0/36, 4/72, 2/53, 1/41, 2/36, 11/39)	5.1	4.2
HCA or HCC (3/36, 8/72, 6/53, 5/41, 12/36, 29/39)	2.6	1.6

Table 2.2 Genes that are differentially expressed in response to 8 mkd furan and relevant to the furan MoA. BMD/BMDL values for individual genes are listed (unless the software was unable to model the dose response of their gene expression, or the value exceeded 8 mkd).

Probe ID	Genbank Accession	Gene Symbol	Gene Name	Fold Change (8 mkd furan)	BMD (mkd)	BMDL (mkd)
<i>Oxidative stress response and xenobiotic metabolism</i>						
A_52_P664506	NM_018811	<i>Abhd2</i>	abhydrolase domain containing 2	1.56	1.4	
A_55_P2177233	NM_026179	<i>Abhd5</i>	abhydrolase domain containing 5	1.61	2.2	1.7
A_52_P273821	NM_026179	<i>Abhd5</i>	abhydrolase domain containing 5	1.68	1.4	0.7
A_55_P2163098	NM_134066	<i>Akr1c18</i>	aldo-keto reductase family 1, member C18	2.78	4.4	3.0
A_55_P2154416	NM_030558	<i>Car15</i>	carbonic anhydrase 15	2.10	2.1	1.6
A_51_P455647	NM_009801	<i>Car2</i>	carbonic anhydrase 2	2.12	3.1	2.3
A_66_P108152	NM_007620	<i>Cbr1</i>	carbonyl reductase 1	1.63	2.9	2.2
A_51_P481159	NM_173047	<i>Cbr3</i>	carbonyl reductase 3	2.53	3.1	2.3
A_55_P2124712	NM_145603	<i>Ces2</i>	carboxylesterase 2	3.08	2.7	2.0
A_52_P318361	NM_145603	<i>Ces2</i>	carboxylesterase 2	3.62	2.7	2.0
A_51_P179919	NM_172759	<i>Ces5</i>	carboxylesterase 5	1.91	7.4	4.5
A_55_P1954718	NM_007805	<i>Cyb561</i>	cytochrome b-561	1.88		
A_55_P2140107	NM_009995	<i>Cyp21a1</i>	cytochrome P450, family 21, subfamily a, polypeptide 1	1.52		
A_55_P2007326	NM_009997	<i>Cyp2a4</i>	cytochrome P450, family 2, subfamily a, polypeptide 4	1.51		
A_51_P137452	NM_013809	<i>Cyp2g1</i>	cytochrome P450, family 2, subfamily g, polypeptide 1	-1.80	7.9	4.6
A_51_P482051	NM_007820	<i>Cyp3a16</i>	cytochrome P450, family 3, subfamily a, polypeptide 16	1.60	1.0	
A_55_P2001780	NM_001105159	<i>Cyp3a41b</i>	cytochrome P450, family 3, subfamily a, polypeptide 41B	1.67		
A_52_P164161	NM_020010	<i>Cyp51</i>	cytochrome P450, family 51	1.62	1.0	0.2
A_55_P2002577	NM_010145	<i>Ephx1</i>	epoxide hydrolase 1, microsomal	1.70	2.8	2.1
A_55_P2002578	NM_010145	<i>Ephx1</i>	epoxide hydrolase 1, microsomal	1.94	3.2	2.0

A_51_P365019	NM_010295	<i>Gclc</i>	glutamate-cysteine ligase, catalytic subunit	1.61	5.4	4.0
A_55_P2032946	NM_008181	<i>Gsta1</i>	glutathione S-transferase, alpha 1 (Ya)	15.28	5.0	3.4
A_55_P2170454	NM_008182	<i>Gsta2</i>	glutathione S-transferase, alpha 2 (Yc2)	3.63	3.3	1.7
A_55_P2062190	NM_010358	<i>Gstm1</i>	glutathione S-transferase, mu 1	1.66	6.3	5.0
A_55_P1957038	NM_181796	<i>Gstp2</i>	glutathione S-transferase, pi 2	1.56	7.7	5.8
A_51_P263965	NM_010442	<i>Hmox1</i>	heme oxygenase (decycling) 1	2.09	7.6	4.2
A_55_P2029687	NM_010442	<i>Hmox1</i>	heme oxygenase (decycling) 1	1.66	6.1	3.8
A_51_P424338	NM_008706	<i>Nqo1</i>	NAD(P)H dehydrogenase, quinone 1	1.71	3.1	2.3
A_51_P161354	NM_144907	<i>Sesn2</i>	sestrin 2	1.80	2.6	2.0
A_51_P243755	NM_011388	<i>Slc10a2</i>	solute carrier family 10, member 2	3.61	1.1	
A_55_P1969506	NM_009198	<i>Slc17a1</i>	solute carrier family 17 (sodium phosphate), member 1	1.52		
A_52_P286520	NM_018824	<i>Slc23a2</i>	solute carrier family 23 (nucleobase transporters), member 2	-1.52	4.6	3.1
A_55_P1964752	NM_194333	<i>Slc23a3</i>	solute carrier family 23 (nucleobase transporters), member 3	2.27	2.6	1.9
A_51_P514405	NM_019741	<i>Slc2a5</i>	solute carrier family 2 (facilitated glucose transporter), member 5	-1.92	4.9	3.3
A_52_P193194	NM_001001321	<i>Slc35d2</i>	solute carrier family 35, member D2	1.66	7.8	4.7
A_55_P2033120	NM_029688	<i>Srxn1</i>	sulfiredoxin 1 homolog (S. cerevisiae)	1.73	4.7	3.0
A_51_P353895	NM_026935	<i>Sult1c2</i>	sulfotransferase family, cytosolic, 1C, member 2	-1.81	3.4	2.5
A_55_P2129449	NM_020565	<i>Sult3a1</i>	sulfotransferase family 3A, member 1	-2.55	4.2	3.3
A_51_P320614	NM_001042523	<i>Txnrd1</i>	thioredoxin reductase 1	1.52	1.8	0.7
A_55_P2006236	NM_009466	<i>Ugdh</i>	UDP-glucose dehydrogenase	1.82	7.4	4.2
A_55_P2057577	NM_145079	<i>Ugt1a6a</i>	UDP glucuronosyltransferase 1 family, polypeptide A6A	1.51	5.2	3.4
A_51_P163578	NM_172881	<i>Ugt2b35</i>	UDP glucuronosyltransferase 2 family, polypeptide B35	2.26	7.7	4.8
A_55_P1967350	NM_172881	<i>Ugt2b35</i>	UDP glucuronosyltransferase 2 family, polypeptide B35	2.69	7.7	5.1
Inflammation						
A_55_P2016462	NM_021274	<i>Cxcl10</i>	chemokine (C-X-C motif) ligand 10	2.6	1.7	1.3

Cell cycle arrest/Cell Death						
A_55_P2145804	NM_026531	<i>Aen</i>	apoptosis enhancing nuclease	2.46	1.0	0.8
A_55_P2141860	NM_026531	<i>Aen</i>	apoptosis enhancing nuclease	2.34	1.0	0.8
A_55_P2002849	NM_175178	<i>Aifm3</i>	apoptosis-inducing factor, mitochondrion-associated 3	-1.62	3.2	2.6
A_55_P2137406	NM_007527	<i>Bax</i>	BCL2-associated X protein	1.73	2.5	1.8
A_51_P117794	NM_007546	<i>Bik</i>	BCL2-interacting killer	-1.68	5.2	3.4
A_55_P2029106	NM_138313	<i>Bmf</i>	BCL2 modifying factor	-2.30	6.3	4.1
A_51_P363947	NM_007669	<i>Cdkn1a / p21</i>	cyclin-dependent kinase inhibitor 1A (P21)	4.64	2.2	1.3
A_55_P1969131	NM_178373	<i>Cidec</i>	cell death-inducing DFFA-like effector c	3.08	4.5	3.1
A_52_P311853	NM_030143	<i>Ddit4l</i>	DNA-damage-inducible transcript 4-like	2.45	4.3	2.3
A_51_P336385	NM_175551	<i>Dido1</i>	death inducer-obliterator 1	1.63	3.2	2.3
A_51_P189361	NM_027950	<i>Osgin1</i>	oxidative stress induced growth inhibitor 1	2.38	3.6	2.6
A_55_P2027836	NM_020275	<i>Tnfrsf10b/DR5</i>	tumour necrosis factor receptor superfamily, 10b / death receptor 5	1.92	2.3	1.8
A_55_P2027879	NM_172571	<i>Fbfl</i>	Fas (TNFRSF6) binding factor 1	2.28	3.7	2.6
Both (Cell cycle arrest/Cell death and Cell survival/Growth)						
A_52_P452689	NM_007498	<i>Atf3</i>	activating transcription factor 3	2.47	3.9	2.6
A_51_P502614	NM_026268	<i>Dusp6</i>	dual specificity phosphatase 6	2.59	2.6	2.0
A_51_P263246	NM_008748	<i>Dusp8</i>	dual specificity phosphatase 8	1.94	2.6	2.0
A_55_P2158990	NM_010591	<i>Jun/c-Jun</i>	Jun oncogene	4.16	2.7	1.7
Cell survival/Growth						
A_55_P1979728	NM_009716	<i>Atf4</i>	activating transcription factor 4	1.58	2.7	2.0
A_55_P1983773	NM_001012273	<i>Birc5/survivin</i>	baculoviral IAP repeat-containing 5/survivin	1.54 *	6.3	4.0
A_66_P111562	NM_007631	<i>Ccnd1</i>	cyclin D1	1.53		
A_52_P612803	NM_009831	<i>Ccng1</i>	cyclin G1	2.00	1.6	0.5
A_51_P372550	NM_026770	<i>Cgrefl</i>	cell growth regulator with EF hand domain 1	1.65	4.1	2.9
A_52_P296913	NM_031396	<i>Cnnm1</i>	cyclin M1	1.61	7.8	4.9
A_55_P1976127	NM_007900	<i>Ect2</i>	ect2 oncogene	1.51	5.9	3.8
A_52_P237077	NM_007956	<i>Esr1/ER1</i> □	estrogen receptor 1 (alpha)	1.52	1.1	

A_52_P235347	NM_020013	<i>Fgf21</i>	fibroblast growth factor 21	3.25	7.6	5.2
A_55_P1960735	NM_011819	<i>Gdf15</i>	growth differentiation factor 15	4.13	2.0	1.5
A_51_P220806	NM_008110	<i>Gdf9</i>	growth differentiation factor 9	3.14	4.7	3.0
A_51_P447545	NM_008341	<i>Igfbp1</i>	insulin-like growth factor binding protein 1	1.86	7.9	4.9
A_51_P260300	NM_010547	<i>Ikbkg/NEMO</i>	inhibitor of kappaB kinase gamma	1.80	4.0	2.8
A_55_P2050652	NM_001136067	<i>Ikbkg/NEMO</i>	inhibitor of kappaB kinase gamma	1.76	2.8	2.1
A_52_P108346	NM_010849	<i>Myc/c-Myc</i>	myelocytomatosis oncogene	1.84	4.5	3.1
A_55_P2007470	NM_008808	<i>Pdgfa</i>	platelet derived growth factor, alpha	1.89	1.0	0.7
A_55_P2125588	NM_008808	<i>Pdgfa</i>	platelet derived growth factor, alpha	1.88	1.8	1.4
A_55_P2092909	NM_019713	<i>Rassf1</i>	Ras association (RalGDS/AF-6) domain family member 1	1.66	1.9	1.5
A_51_P111164	NM_172612	<i>Rnd1</i>	Rho family GTPase 1	-1.59	5.7	3.6
A_51_P404377	NM_009708	<i>Rnd2</i>	Rho family GTPase 2	1.89	1.1	0.5
A_55_P2184009	NM_009708	<i>Rnd2</i>	Rho family GTPase 2	1.73	4.2	2.9
A_55_P2031045	NM_011489	<i>Stat5b</i>	signal transducer and activator of transcription 5B	-1.51	3.9	2.8
A_55_P2068663	NM_019641	<i>Stmn1</i>	stathmin 1	1.53	4.9	3.3
A_51_P131408	NM_013749	<i>Tnfrsf12a/TweakR</i>	tumour necrosis factor receptor superfamily, 12a / TNF related weak inducer of apoptosis receptor	2.54	3.7	2.4
A_55_P2116496	NM_011654	<i>Tuba1b</i>	tubulin, alpha 1B	1.65	3.9	2.7
A_55_P2108837	NM_009448	<i>Tuba1c</i>	tubulin, alpha 1C	2.40	1.4	0.5
A_55_P2064547	BC022182	<i>Tuba1c</i>	tubulin, alpha 1C	2.36	1.5	0.6
A_52_P676271	NM_009447	<i>Tuba4a</i>	tubulin, alpha 4A	1.71	6.4	4.0
A_51_P490023	NM_009450	<i>Tubb2a</i>	tubulin, beta 2A	4.30	2.7	2.0
A_55_P2077783	NR_003964	<i>Tubb2a-ps2</i>	tubulin, beta 2a, pseudogene 2	2.48	2.8	2.1
A_55_P2153292	NM_146116	<i>Tubb2c</i>	tubulin, beta 2C	2.86	1.4	1.1
A_55_P1959703	XM_001473123	<i>Tubb2c-ps1</i>	tubulin, beta 2c, pseudogene 1	2.71	2.0	1.5
A_51_P421140	NM_026473	<i>Tubb6</i>	tubulin, beta 6	1.71	2.0	1.5

* p = 0.0776

Table 2.3 Margin of exposure (MoE) values for furan-induced molecular pathways and cancer endpoints. A MoE threshold of 1000 has been previously used for furan (IRIS, 2012). BMDLs for HCC and HCA were modeled using previously published data from Moser *et al.* (2009) and NTP (1993) for female B6C3F1 mice. BMDL mean for pathway data were calculated using transcriptional data from this study in BMDExpress. Estimated human exposure levels were taken from Carthew *et al.* (2010).

<i>Study:</i>	NRF2 Oxidative Stress Response Pathway	All-Pathway Average *	HCC		HCA	
	<i>(this study)</i>	<i>(this study)</i>	<i>Moser</i>	<i>NTP</i>	<i>Moser</i>	<i>NTP</i>
BMDL (µg/kg-bw/day):	2250	2200	4218	5303	1567	923
Exposure estimates for child and adult (> 2 years)						
<u>USA</u>						
Average (50%): 0.3 µg/kg-bw/day	7500	7333	14060	17677	5223	3077
High level (90%): 0.6 µg/kg-bw/day	3750	3667	7030	8838	2612	1538
<u>Europe</u>						
Average (50%): 0.8 µg/kg-bw/day	2813	2750	5273	6629	1959	1154
High level (90%): 1.75 µg/kg-bw/day	1286	1257	2410	3030	895	527
Exposure estimates for infants (0-1 years)						
<u>USA</u>						
Average (50%): 0.4 µg/kg-bw/day	5625	5500	10545	13258	3918	2308
High level (90%): 1 µg/kg-bw/day	2250	2200	4218	5303	1567	923
<u>Europe</u>						
Average (50%): 0.3 µg/kg-bw/day	7500	7333	14060	17677	5223	3077
High level (90%): 1 µg/kg-bw/day	2250	2200	4218	5303	1567	923

* Calculated using the average BMDL for all enriched pathways (enriched by 8 mkd furan, with at least 5 differentially expressed genes in each pathway).

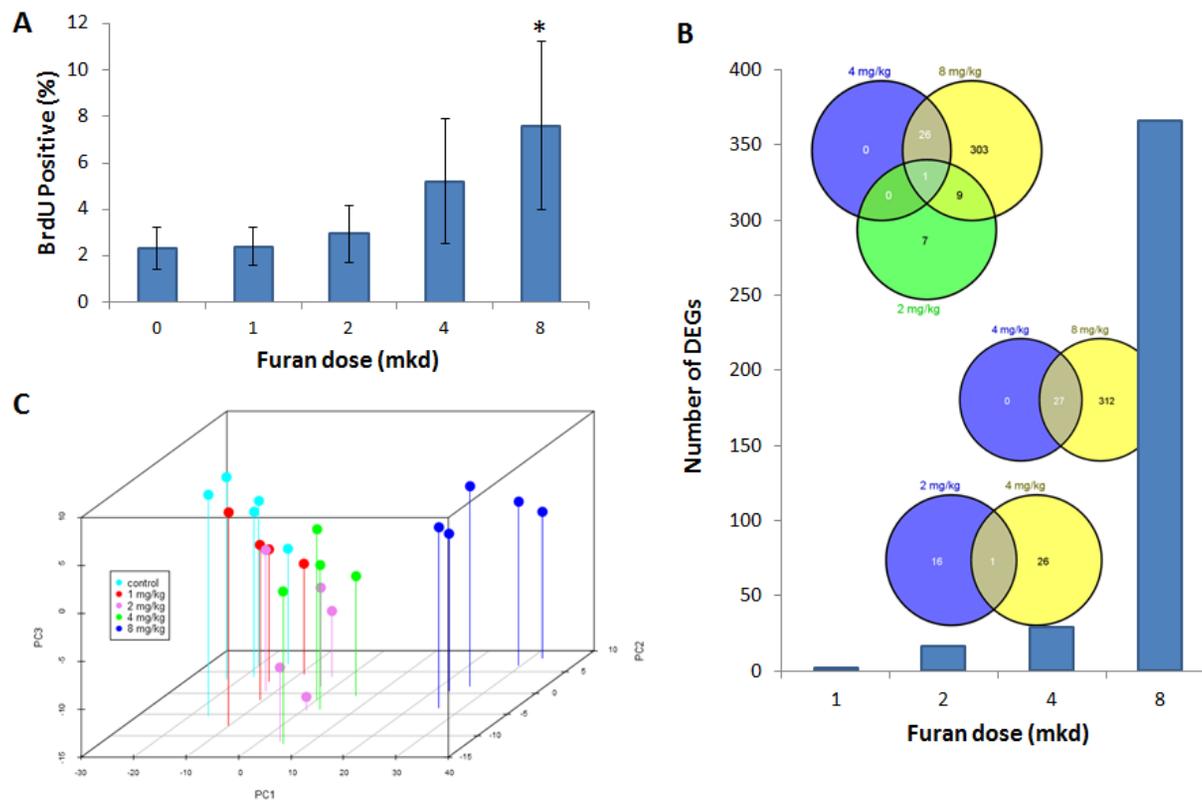


Figure 2.1 Dose response furan of furan-induced effects. Dose-dependent increase in liver cell proliferation upon exposure to furan based upon BrdU incorporation (\pm stdev; * $P = 0.00413$). (b) Number of furan-responsive differentially expressed genes in each dose group and Venn diagrams represent the overlap of individual furan-responsive genes in each dose group. (C) Principal component analysis (PCA) of furan dose groups, where circles represent individual mice and colours indicated dose groups (royal blue, 8 mkd; green, 4 mkd; pink, 2 mkd; red, 1 mkd; aqua blue, 0 mkd).

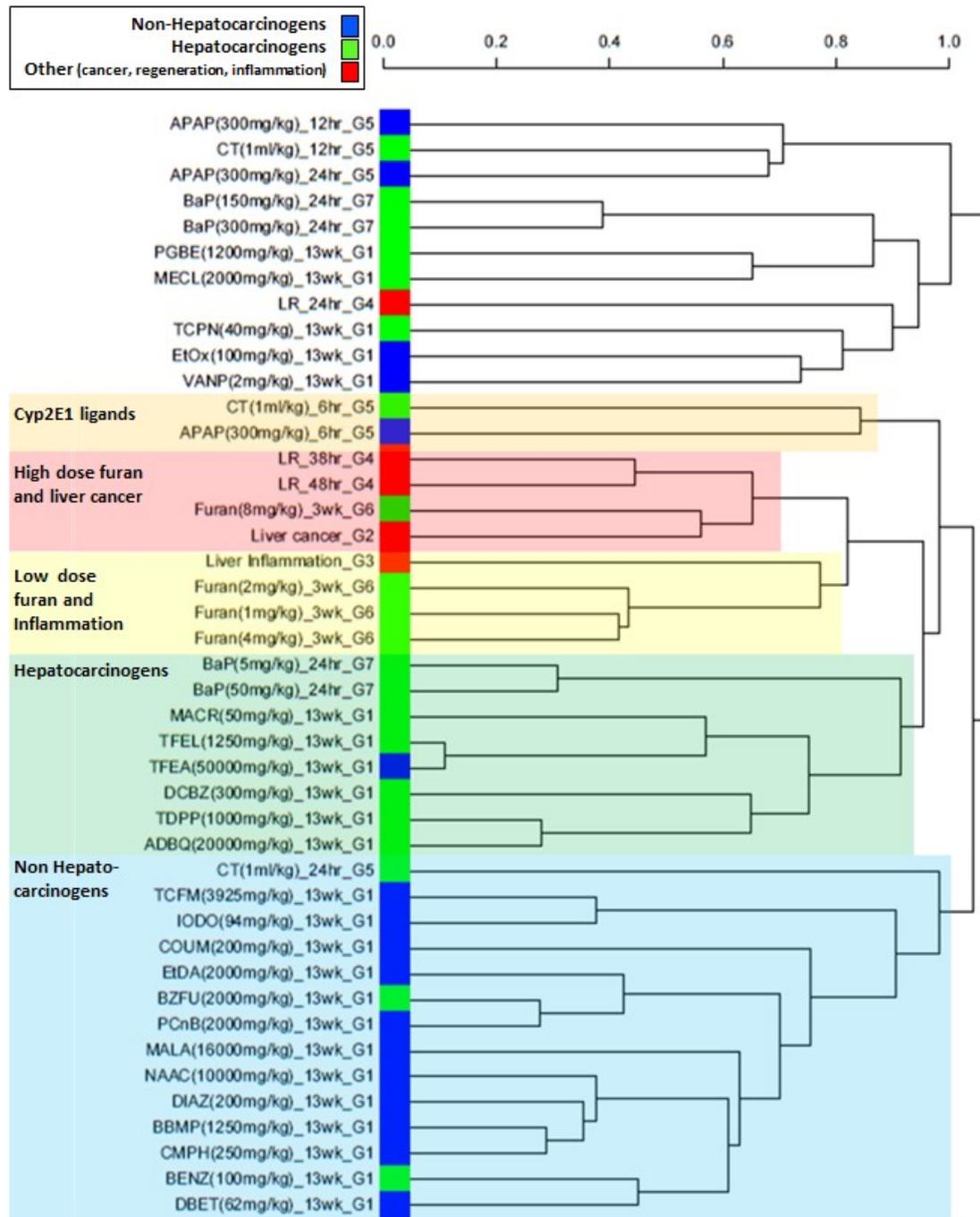


Figure 2.2 Cluster analysis comparing furan-dependent changes in gene expression data with publically available datasets. Publically available datasets are for liver cancer (HCC), liver regeneration (LR), Cyp2E1 ligand-, and hepatocarcinogen- and non-hepatocarcinogen-induced changes in gene expression in mouse liver. Datasets were downloaded from GEO, where G1 is GSE18858, G2 is GSE26538, G3 is GSE35934, G4 is GSE20427, G5 is GSE4874, G6 is this study (GSE48644), and G7 is from another study conducted by our laboratory (Yauk et al., unpublished data; manuscript in preparation).

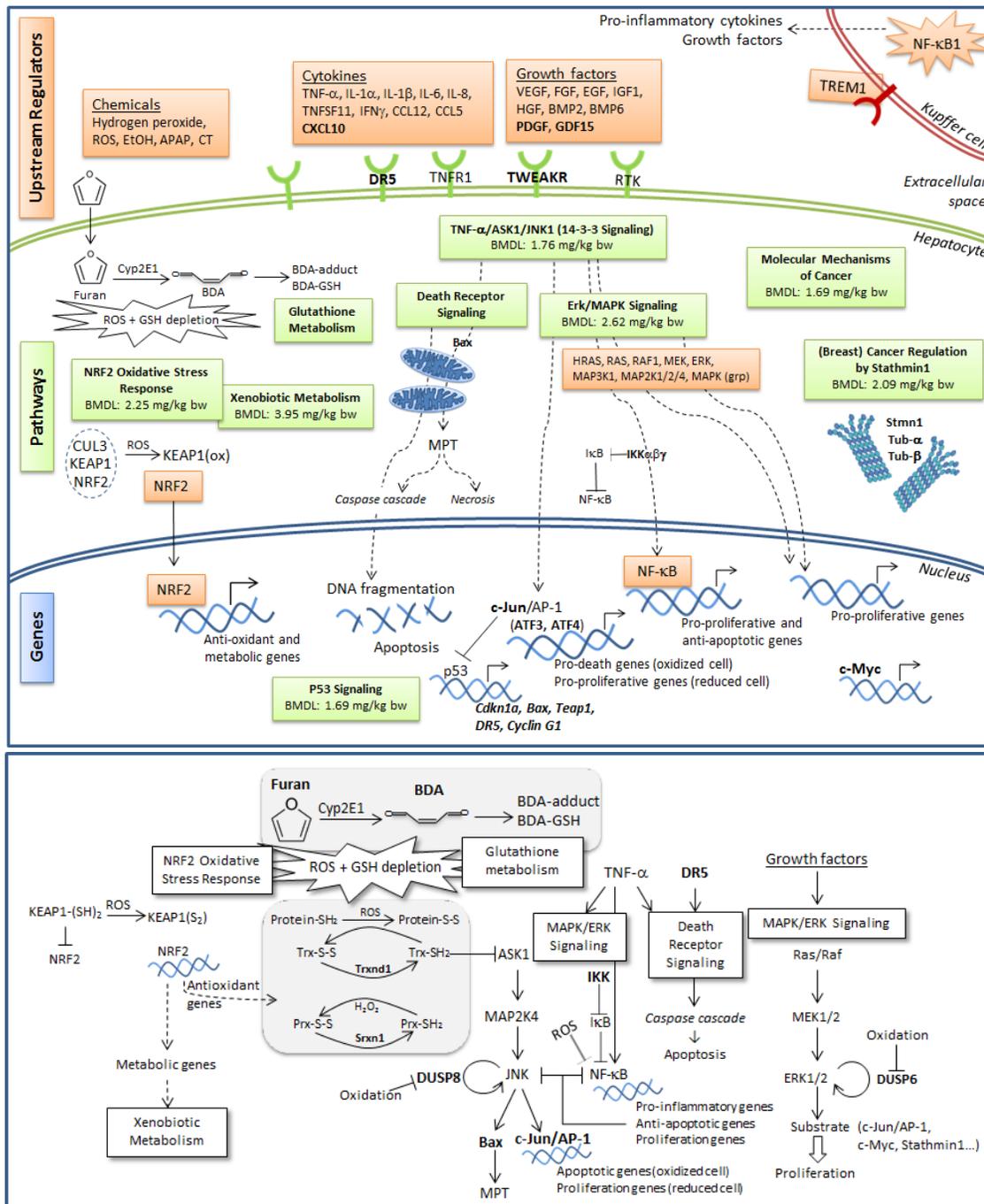


Figure 2.3 Furan MoA. (A) Summary of predicted upstream regulators (orange), enriched molecular pathways (green) with BMDL values indicated, and selected differentially expressed genes (bold text) for the 8 mkd dose group. (B) A more detailed image of the proposed redox switch and other molecular signals implicated in the furan MOA.

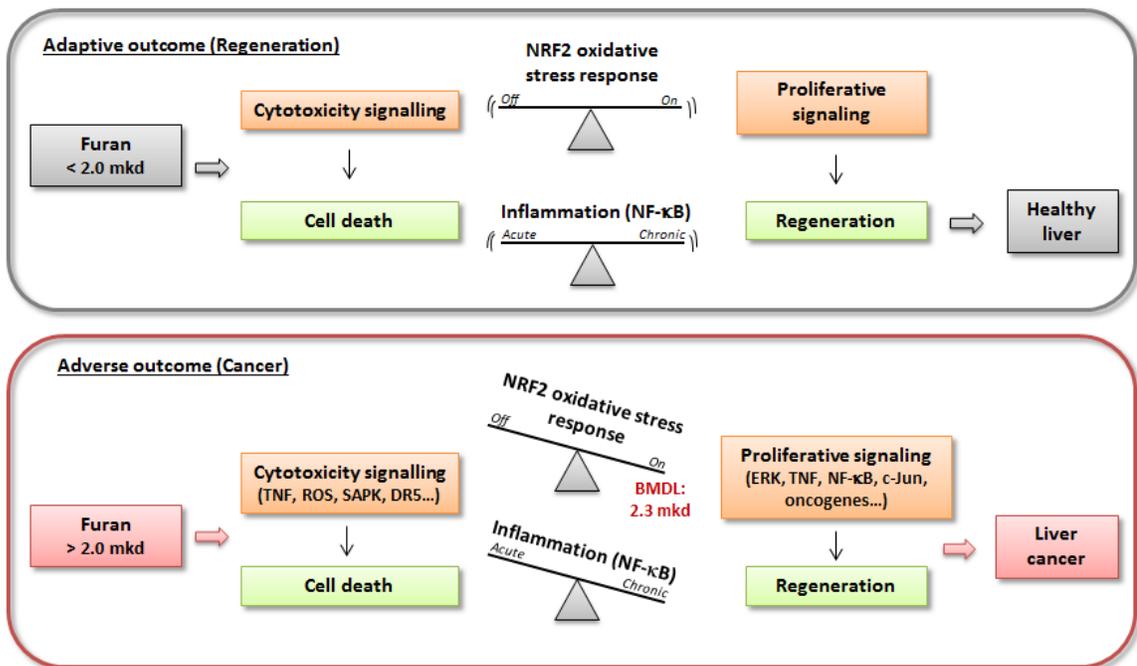


Figure 2.4 Adaptive and adverse outcomes in response to non-carcinogenic and carcinogenic doses of furan. We propose that NRF2 activation and inflammation could represent important tipping points in the MOA and that, when chronically activated, drive the cancer outcome.

CHAPTER 3

3 CHAPTER: BROMODEOXYURIDINE TREATMENT TO MEASURE HEPATOCELLULAR PROLIFERATION DOES NOT MASK CHEMICALLY INDUCED GENE EXPRESSION CHANGES IN MOUSE LIVER

Modified from: A.F. Jackson, Williams, A., Recio, L., Yauk, C.L. 2014. *Toxicology*, 323:26-31.

3.1 Preamble

An important theme of this thesis is the development of strategies that facilitate the reduction of the number of experimental animals that are required for toxicity testing. BrdU is a synthetic nucleoside that is an analog of thymidine. BrdU incorporation into DNA is a commonly used method for measuring cellular proliferation. Typically, studies that wish to measure both gene expression and BrdU incorporation use two cohorts of animals: one that is BrdU treated and one that is not. The former are used to assay for cellular proliferation and the latter are used to measure global gene expression changes. In Chapter 3 we asked whether or not treatment with BrdU masks or changes toxicant-dependent changes in gene expression. Importantly, if the answer to that question is ‘no’, then the number of experimental animals used in this type of study (such as the study described in Chapter 2) can be halved.

3.2 Abstract

BrdU is a synthetic nucleoside used to detect cellular proliferation. BrdU incorporates in the place of thymine but pairs with guanine, thereby increasing the risk of transition mutations in dividing cells. Given its mutagenicity, standard practice is to use a second cohort of animals for parallel toxicogenomics studies; however, the impact of BrdU on global gene expression is unknown. To test this we performed a case study to determine whether the molecular MoA of furan, a liver carcinogen, could be detected in BrdU-treated samples. We measure global hepatic gene expression in female B6C3F1 mice that were exposed for three weeks to 0, 1, 4, or 8 mg/kg furan in the presence (+BrdU) or absence (-BrdU) of BrdU. Exposure to 0.02% BrdU in drinking water for five days resulted in minimal gene expression changes. A comparison of +BrdU versus -BrdU control mice revealed only 11 significantly changed probes. The same comparison in the high dose group yielded only 3 probes. Differentially expressed gene lists generated for furan-treated versus control mice and were compared for the -BrdU and +BrdU groups. The high dose of furan had 452 shared probes and 27 and 90 unique probes for -BrdU and +BrdU groups, respectively. These differences did not impact hierarchical clustering, and they did not impair detection of the previously reported furan MoA. Taken together, we demonstrate that BrdU treatment does not mask important furan-induced transcriptional changes. We suggest that BrdU-treated mice can be used for toxicogenomic analysis, which would generally halve the number of rodents required for these studies. However, we also recommend that this type of case study should be repeated for other chemicals before the use of BrdU-treated animals in transcriptomics studies becomes common practice.

3.3 Introduction

Cellular proliferation is a standard endpoint in conventional toxicological testing. BrdU is a synthetic nucleoside that is used to detect cellular proliferation. BrdU is incorporated into replicating DNA in the place of thymidine, and BrdU positive cells are detected using immunohistochemistry (IHC). One advantage for using BrdU is that it is delivered to animals as a steady dose over 4-5 days (by osmotic pump or in drinking water) and therefore shows all of the cellular proliferation that has accumulated over that time period. Other IHC options for measuring cellular proliferation endpoints (such as Ki67 or PCNA) only provide a snapshot of the cellular proliferation that is occurring at the time of staining and are therefore much less sensitive. One perceived disadvantage for using BrdU is that, since it incorporates into DNA, it is generally assumed that it can cause mutations that will manifest as changes in gene expression. As a result, it has also been assumed that transcriptomics studies require two cohorts of animals, one for BrdU treatment and the other for gene expression analysis, which doubles the number of experimental animals required; for example: Jackson *et al.* (2014), Roat *et al.* (2014), and Zhang *et al.* (2014) . However, if BrdU treatment does not markedly alter global gene expression, then it might be possible to halve the number of experimental animals used because the +BrdU animals could also be used for gene expression analysis. Here we address this question by performing a case study in which we investigate whether BrdU treatment masks the molecular MoA of the liver carcinogen, furan.

Previous work on the cellular effects of BrdU has focused on the *in vitro* characterization of BrdU-induced T to C transition mutations (BrdU incorporates in the place of thymine but pairs with guanine) (Lasken & Goodman 1984, Goodman *et al.*

1985). More recent work has shown that, in addition to these predominant transition mutations, BrdU can also cause less common transversion, deletion and insertion mutations *in vitro* (Ma *et al.* 2008). Surprisingly, very little work has been done to test what effect BrdU has on gene expression. BrdU has been shown to increase ectopic gene expression when delivered near to the time of transfection in HeLa cells (Endoh *et al.* 2007) and has been reported to decrease expression of other genes including calmodulin (*Cmd1*) and myogenic differentiation 1 (*Myod1*) (Lin *et al.* 1989), integrin-alpha2 (*Itga2*) (Meleady & Clynes 2001), and tyrosinase (*Tyr*) (Rauth *et al.* 1990). Reduction in *Tyr* transcription was reported to not to be due to BrdU incorporation in either the upstream or coding regions of the tyrosinase gene (Rauth & Davidson 1993), suggesting an indirect effect. One explanation could be that BrdU is causing gene silencing by increasing gene methylation (Call & Thilly 1991). We could not find any published studies that examine the effects of BrdU treatment on global gene expression *in vivo*.

Previously, we used toxicogenomics to generate a molecular MoA for initiation of liver cancer by furan (Chapter 2). This study measured furan-induced changes to gene expression in mice exposed to non-carcinogenic and carcinogenic doses of furan. Each dose group was split into two cohorts: one cohort was used for gene expression analysis, and the other received 0.02% BrdU in drinking water for five days prior to necropsy and was used to measure cellular proliferation. Here, we compare gene expression profiles of these two cohorts of mice, +BrdU and -BrdU, from the control (vehicle only), lowest (1 mkd), middle (4 mkd) and highest (8 mkd) furan dose groups to determine whether important furan-induced changes in gene expression are still detectable following treatment with BrdU.

3.4 Materials and Methods

3.4.1 Animals

Animal care and exposures have been previously described (section 2.4.2). Livers were divided and pieces were either snap-frozen in liquid nitrogen and stored at -80°C, or FFPE-preserved for quantification of BrdU by IHC. BrdU positive cells were quantified by light microscopy (described in 2.4.3).

3.4.2 RNA Extraction and Microarray

Methods for RNA extraction and microarray analysis have been previously described (described in 2.4.4-2.4.5). The complete datasets are available through the GEO, accession numbers: GSE48644 (-BrdU) and GSE54078 (+BrdU).

A block design, treating the slide as the blocking effect, was employed to determine global differential gene expression (Kerr 2003). Arrays containing control, 1, and 8 mkd samples \pm BrdU were completed in one experiment. An additional array containing the 4 mkd furan +BrdU samples and corresponding +BrdU controls was run separately. Median signal intensities were normalized in R (R-Core-Development-Team 2012) using LOWESS (Yang *et al.* 2002). Probes with technical replicates were then averaged. The MAANOVA library (Wu *et al.* 2003) was used for differential gene expression determination. The Fs statistic (Cui *et al.* 2005), a shrinkage estimator, was used to determine gene-specific treatment effects, and the associated p-values were estimated using the permutation method (30,000 permutations with residual shuffling). Using the FDR approach, p-values were adjusted for multiple comparisons (Benjamini & Hochberg 1995). Fold changes were estimated using least square means of each pairwise

comparison. Genes having an FDR-adjusted $P \leq 0.05$ and a fold change ≥ 1.5 were deemed differentially expressed. Pair-wise comparisons between the dose groups to controls with and without BrdU were conducted as well as direct comparisons testing for differences between controls with and without BrdU, and similarly for the 1, 4, and 8 mkd dose groups.

3.4.3 Identification of Unique and Shared Probes

Unique and shared probes were identified according to Thomson et al. (2012). Briefly, probes were identified as unique to the high dose -BrdU group if they were significantly changed (FDR $P \leq 0.05$ and fold change $\geq \pm 1.5$) in the 8 mkd -BrdU versus control -BrdU gene list, and were not significantly changed (unadjusted $p \geq 0.05$ and fold change $\leq \pm 1.5$) in the 8 mkd +BrdU versus control +BrdU gene list. Similarly, probes were considered unique to the high dose +BrdU that were significantly changed (FDR $p \leq 0.05$ and a fold change $\geq \pm 1.5$) in the 8 mkd +BrdU versus control +BrdU gene list, and were not significantly changed (unadjusted $p \geq 0.05$ and fold change $\leq \pm 1.5$) in the 8 mkd furan -BrdU versus control -BrdU gene list. The remaining probes were identified as shared between the two lists.

3.4.4 Bioinformatic Analysis

Cluster analysis and data visualization via the heatmap was done using (R-Core-Development-Team 2012). The genes used for clustering were present in at least 4 of 5

samples with intensities above background for at least one treatment condition (i.e.: an FDR $P \leq 0.05$ and a fold change $\geq \pm 1.5$ for at least one of the comparisons made).

A Core Comparison Analysis using IPA software to identify enriched Signaling Pathways, Hepatotoxicity ToxFUNCTIONS, and Predicted Upstream Regulators (PURs) was conducted for the 8 mkd dose group (which was the only dose group with a sufficiently high number of differentially expressed genes). The significance threshold for IPA canonical pathways and ToxFUNCTIONS was $P \leq 0.05$ ($-\log P \geq 1.3$) where the p-value is calculated using a Fisher's Exact test by IPA; only pathways with at least 4 molecules were considered. PURs were identified in IPA based on the number of differentially expressed downstream genes in the data set and quantified by IPA using a z-score (activated: $Z \geq 2$ and inhibited $Z \leq -2$).

3.5 Results and Discussion

The purpose of this research was to assess whether treatment of rodents with BrdU would impair the ability to detect furan-induced changes in gene expression. BrdU is the preferred method for detection of cellular proliferation because it shows an accumulation of cellular proliferation over a period of days, which is more sensitive than other methods (Ki67 and PCNA) that only provide snapshots. BrdU is generally expected to have an impact on gene expression because it is an analog of thymidine that pairs with guanine, thereby increasing the potential for the occurrence of transition mutations throughout the genome. However, the effects of BrdU on global gene expression have not been robustly tested. If five days of 0.02% BrdU in drinking water has little effect on global gene expression, then the number of experimental animals used in omics studies

could be halved because the BrdU-treated animals could also be used for gene expression analysis. This case study is an important first step in determining if the multi-purposing of BrdU-treated animals is a feasible route towards a reduction in the number of experimental animals used in this type of study.

3.5.1 Effects of BrdU Treatment on Gene Expression (within dose groups)

There were 11 probes whose expression was significantly changed in control mice as a result of BrdU treatment (i.e., a comparison of control +BrdU versus control –BrdU mice). At 1 mkd furan, BrdU treatment affected the expression of eight probes. At 8 mkd furan, BrdU treatment affected the expression of three probes (Table 3.1). None of these probes were shared across the three dose groups (indicating no consistent effect of BrdU treatment) and only one probe (A_52_P137371, corresponding to *Hmgcr*) was in common between the 0 and 1 mkd dose groups. These data suggest that 5 days of treatment with BrdU in drinking water does not alter gene expression to a significant degree or in a consistent manner.

3.5.2 Detection of Furan-Induced Gene Expression Changes in the Presence of BrdU

Treatment with 1 mkd furan resulted in the up-regulation of two probes (A_55_P2148071 and A_51_P290207, corresponding to LOC673748 and *Insig1*) relative to the control mice (in the –BrdU mice). When the same comparison was made using the +BrdU mice there were no differentially expressed probes. These data suggest that neither 1 mkd furan or 1 mkd furan +BrdU have a significant impact on gene expression.

The middle dose of furan (4 mkd) caused the differential expression of 47 probes (corresponding to 31 unique Genbank accession numbers) relative to the control group (–BrdU). When the same comparison was made using the corresponding +BrdU mice, there were 84 altered probes (corresponding to 63 unique Genbank accession numbers). The high dose of furan (8 mkd) had major transcriptional effects with 442 differentially expressed probes (corresponding to 354 unique Genbank accession numbers) relative to the control group (–BrdU). When the same comparison was made using the corresponding +BrdU mice, there were 301 altered probes (corresponding to 244 unique Genbank accession numbers) (Table 3.2). In both 8 mkd furan gene lists *Gsta1*, followed by Gm10639 and *Cdkn1a*, were the most up-regulated genes, and *Tubb2a*, *c-Jun*, *Gdf15*, and LOC625953 were shared in the top 10 up-regulated genes. Up-regulation of other important MoA genes including: *Atf3/4*, *Bax*, *Birc5*, *Ces2*, *Dusp6*, *Dusp8*, *DR5* (*Tnfrsf10b*), *Gclc*, various *Gsts*, *Hmox1*, *Myc*, *Nqo1*, *Stathmin1*, *Srxn1*, *Tnfrsf12a*, and *Txnrd1*, was maintained in both lists.

Given the outcome of the direct comparisons (Table 3.1), we were surprised to see that the 4 mkd furan +BrdU gene list had 37 more probes than the corresponding –BrdU gene list; further, the 8 mkd furan +BrdU gene list had 141 fewer differentially expressed probes compared to the –BrdU gene list. One hypothesis is that furan and BrdU interact with each other inside the hepatocytes. We think that this is unlikely since BrdU is a thymidine analog and furan’s metabolite, BDA, has been shown to produce nucleoside adducts *in vitro* with adenine, cytosine and guanine, but not thymine (Byrns *et al.* 2006, Byrns *et al.* 2002). Further, BDA is generally understood to be reactive toward protein (not DNA) *in vivo* (Burka *et al.* 1991, Chen *et al.* 1997). Another possibility is

that the FDR p-value is too stringent a parameter to use when comparing between lists. To address this we relaxed our probe thresholds according to the method described by Thomson et al. (2012). We compared the differentially expressed probe lists generated by comparing 1mkd furan +BrdU versus control +BrdU, and 1mkd furan versus control. There was one shared probe and one probe that was unique to the –BrdU group. The same comparison made for the middle dose group (4 mkd furan) yielded 52 shared probes and 30 probes that were unique to each list. At the high dose group (8 mkd furan) we identified 90 probes that were unique to the –BrdU group, 27 that were unique to the +BrdU group, and 452 probes that were shared, which was more consistent with the direct comparison results.

The presence or absence of BrdU had minimal impact on higher-order analyses. Indeed, clustering of each dose group demonstrated that in the absence of influence on gene expression by a chemical exposure (in this case 0 or 1 mkd furan) BrdU directed clustering; however, this effect was subtle and was not robust enough to mask the gene expression profile produced following exposure to carcinogenic doses (Fig. 3.1).

3.5.3 Detection of Furan’s Molecular MoA

The primary goal of this study was to determine if important MoA information for furan-induced hepatocarcinogenicity is accurately represented in BrdU-treated samples. Furan is known to be metabolized by CYP2E1 in the liver. While non-metabolized furan is biologically inert, BDA is highly cytotoxic (Kellert *et al.* 2008). Furan metabolism by CYP2E1 also produces oxidative stress in the cell (Gonzalez 2005). Following injury the liver can regenerate itself; however, in the case of furan, dysregulated liver regeneration

is thought to result in excessive cellular proliferation leading to tumour formation. This cytotoxicity followed by regenerative proliferation MoA is well documented based on apical (Fransson-Steen *et al.* 1997, Moser *et al.* 2009, Hickling *et al.* 2010) and transcriptomic (Jackson *et al.* 2014) data. We observed that this MoA is well represented in the high dose +BrdU samples, which reached significance for a number of hepatotoxicity ToxFUNCTIONS that were very consistent with furan's MoA, including: liver necrosis/cell death, glutathione depletion, liver damage, liver proliferation and hyperproliferation, liver regeneration, and hepatocellular carcinoma (Fig. 3.2).

IPA ToxFUNCTIONS classify gene expression data into very high level ontologies. A greater mechanistic understanding of the data comes from pathways analysis. We conducted an enrichment analysis for IPA signaling pathways for both high dose gene lists (\pm BrdU). The +BrdU gene list enriched for 14 signaling pathways and the -BrdU gene list enriched for 27, which was not surprising since the -BrdU gene list had more DEGs (for the full list of molecular pathways refer to Appendix C, Table C4). This notwithstanding, we can confirm that important furan MoA pathways are present in the +BrdU analysis. The Nrf2-mediated Oxidative Stress Response is a top ranked pathway for both -BrdU ($p < 0.0001$; number of molecules, NoM: 15) and +BrdU ($p < 0.0001$; NoM: 10), which is crucial since activation of this pathway is a key event in the MOA. Aryl Hydrocarbon Receptor Signaling (-BrdU: $p < 0.0001$, NoM: 12; +BrdU $p < 0.0001$, NoM: 10) and Xenobiotic Metabolism Signaling (-BrdU: $p = 0.0002$, NoM: 13; +BrdU $p = 0.0006$, NoM: 10) were also highly ranked; however, this is more a reflection of the enrichment of the Nrf2-mediated Oxidative Stress Response since there is a large overlap of molecules and redundancy across of these pathways. Enrichment of cell death

pathways: 14-3-3-mediated Signaling (-BrdU: $p=0.0006$, NoM: 8; +BrdU $p=0.01$, NoM: 5) and Death Receptor Signaling (-BrdU: $p=0.01$, NoM: 4; +BrdU $p=0.03$, NoM: 3) was statistically significant in both, but the latter only had three molecules in the +BrdU group (normally pathways are only considered if they have at least 4 molecules). Interestingly, p53 Signaling was more significantly enriched in the +BrdU group (-BrdU: $p=0.004$, NoM: 6; +BrdU $p<0.0001$, NoM: 8); this difference is largely driven by the upregulation of Mdm2 in the +BrdU group, which could be an indication of BrdU-mediated DNA damage. Two cellular proliferation pathways from the MoA that did not reach significance in the +BrdU group were: (Breast) Cancer Regulation by Stathmin1 (-BrdU: $p=0.003$, NoM: 9; +BrdU $p=0.07$, NoM: 5) and ERK/MAPK Signaling (-BrdU: $p=0.01$, NoM: 8; +BrdU $p=0.06$, NoM: 5). It is noteworthy that *Stmn1* gene expression was still increased in the +BrdU group. Further, p38 MAPK Signaling, a similar pathway to ERK/MAPK Signaling, was enriched in the +BrdU group (-BrdU: $p=0.1$, NoM: 4; +BrdU $p=0.04$, NoM: 4) and MAPKs have predicted activity in both \pm BrdU groups (discussed below). Taken together we can say that, while there are differences in the lists of pathways produced by the -BrdU and +BrdU groups, very important pathways are well conserved and minor pathways continue to manifest in the data.

Many important signaling molecules (particularly signaling proteins) are regulated post-transcriptionally and therefore they do not appear as differentially regulated in the transcriptional data produced by microarrays. To address this and fill in the gaps in pathway analysis, activation or inhibition of upstream regulators can be predicted based on downstream gene expression. Top PURs include: oxidative stress (NRF2/NFE2L2), cell death and proliferation (TNF-alpha, ERK, MAP2K1/2; MAP3K1,

and various growth factors), and inflammation (TNFSF11 / RANKL, TNF superfamily 11 / Receptor activator of nuclear factor kappa-B ligand; and, various cytokines). In toxicogenomics studies, it is also useful to know if the downstream effects of your chemical are similar to the effects of other, well-studied chemicals. To this end, a chemical PUR analysis can be conducted. Top chemical PURs in this study included hydrogen peroxide and the reactive oxygen species group (indicative of oxidative stress), and carbon tetrachloride (indicative of Cyp2E1 activation). Taken together, these higher order transcriptomic analyses (Hepatotoxicity ToxFUNCTION, Signaling Pathway, and PUR) have demonstrated that furan's MoA is detectable in the +BrdU samples, which indicates that any transcriptional effects of BrdU are not strong enough to occlude the transcriptional effects of furan.

3.6 Conclusions

Our study demonstrates that when only BrdU-treated animals are available, gene expression analyses on these samples is very worthwhile. Indeed, we have demonstrated that toxicogenomic profiles from BrdU-treated samples can reveal important molecular information regarding toxicant MoA. The lack of any substantial change in global gene expression in mice exposed to 0 or 1 mkd furan demonstrates that 5 days of exposure to 0.02% BrdU in drinking water does not substantively alter baseline gene expression patterns in female B6C3F1 mice. The two carcinogenic doses of furan tested in this study, 4 and 8 mkd, did show differences between the +BrdU and -BrdU groups. However, these differences were not robust enough to interfere with hierarchical clustering. Further, BrdU did not mask important MoA information. Nevertheless, we

believe that our findings should be confirmed with several more chemicals across a variety of modes of carcinogenic action before the use of BrdU treated animals for gene expression profiling becomes common practice. However, we feel that this study is an important step towards reduction in the number of experimental animals used in conventional toxicity testing and we hope that it will encourage others to test if there is an effect of BrdU in their toxicogenomic data.

3.7 Tables and Figures

Table 3.1 Gene expression values of probes differentially expressed by BrdU in control and furan-treated mice.

Probe ID	Genbank Accession	Gene symbol	Fold Change
<u>Control (+BrdU) vs. Control (-BrdU)</u>			
A_52_P137371	NM_008255	<i>Hmgcr</i>	2.29
A_66_P134265	NM_001170572	<i>Gm1381</i>	2.28
A_55_P1975185	NM_009270	<i>Sqle</i>	2.13
A_52_P663526	NM_145497	<i>BC016495</i>	1.97
A_51_P463440	NM_130450	<i>Elovl6</i>	1.95
A_51_P209372	NM_025436	<i>Sc4mol</i>	1.69
A_51_P161946	NM_175332	<i>E130012A19Rik</i>	1.56
A_55_P2009217	NR_004415	<i>Rnu3b1</i>	-1.53
A_51_P209736	NM_153778	<i>Atoh8</i>	-1.66
A_55_P2009988	NM_175093	<i>Trib3</i>	-1.92
A_30_P01021780	(none)	lincRNA:chr19:6122754 1-61274669 forward strand	1.56
<u>1 mkd (+BrdU) vs. 1 mkd (--BrdU)</u>			
A_55_P2387665	AK033690	<i>9130221J18Rik</i>	3.09
A_52_P163021	NM_182959	<i>Slc17a8</i>	2.22
A_51_P473953	NM_001081295	<i>4631416L12Rik</i>	2.16
A_52_P547662	NM_008772	<i>P2ry1</i>	1.88
A_52_P627269	NM_198171	<i>BC015286</i>	1.88
A_52_P318361	NM_145603	<i>Ces2</i>	1.81
A_52_P653825	NM_029550	<i>Kegl</i>	1.80
A_52_P137371	NM_008255	<i>Hmgcr</i>	1.76
<u>8 mkd (+BrdU) vs.81 mkd (--BrdU)</u>			
A_52_P156775	NM_001099742	<i>Scgblc1</i>	1.88
A_55_P1985835	NM_010061	<i>Dnase1</i>	1.56
A_51_P411345	NM_177448	<i>Mogat2</i>	1.52

All probes are FDR $P \leq 0.05$

Table 3.2 Number of differentially expressed probes (Agilent probe IDs) or genes (unique GenBank accession numbers) for each furan dose group, without (top) or with (bottom) BrdU treatment. All probes/genes were fold change $\geq \pm 1.5$. Two p-value thresholds are shown: a stringent FDR $P \geq 0.05$ (in bold text), and a less stringent unadjusted $P \geq 0.05$ (in plain text).

Without BrdU (-BrdU)						
Furan (mkd)	<u>1 vs. 0</u>		<u>4 vs. 0</u>		<u>8 vs. 0</u>	
p-value (≤ 0.05)	<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>Unadj</i>
Number of Probes	2	85	47	82	442	564
GenBank Accession Numbers	2	60	31	59	354	456
With BrdU (+BrdU)						
Number of Probes	0	44	84	107	301	409
GenBank Accession Numbers	0	42	63	80	244	337

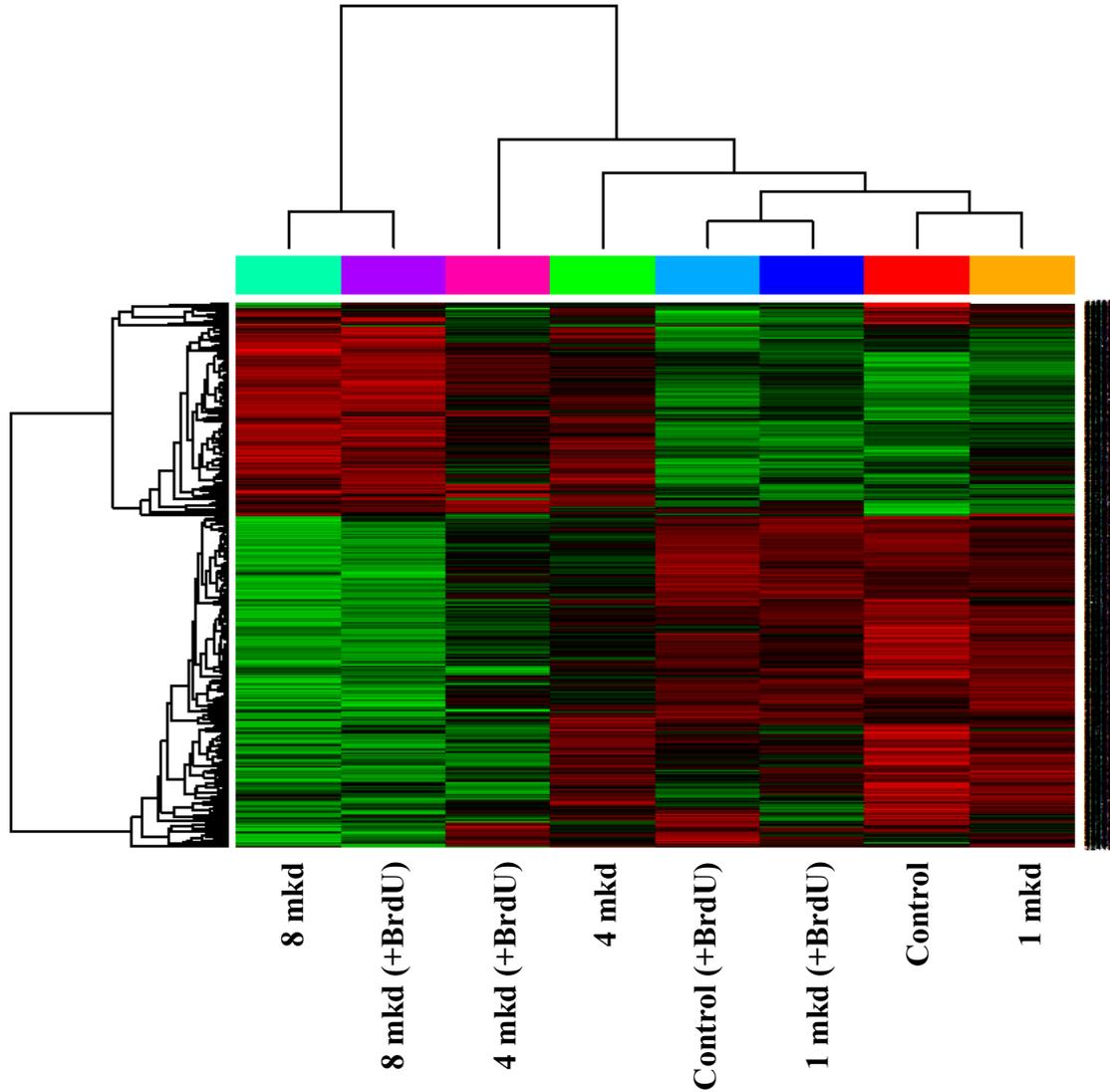


Figure 3.1 Hierarchical clustering of global gene expression profiles from each dose group (\pm BrdU treatment). Each column in the heatmap represents a different treatment group and each row a differentially expressed probe. Column colours identify dose groups where the light blue, royal blue, pink, and purple groups are the +BrdU groups, and red, orange, green, and aqua green are the -BrdU groups, for the control, 1, 4, or 8 mkd furan dose groups, respectively. Within the heatmap, red and green indicate high and low levels of expression relative to control, respectively. Genes were included if they reached significance (FDR $p \leq 0.05$, fold change ≥ 1.5) in at least one treatment group. Groups represent an average of all mice in that group.

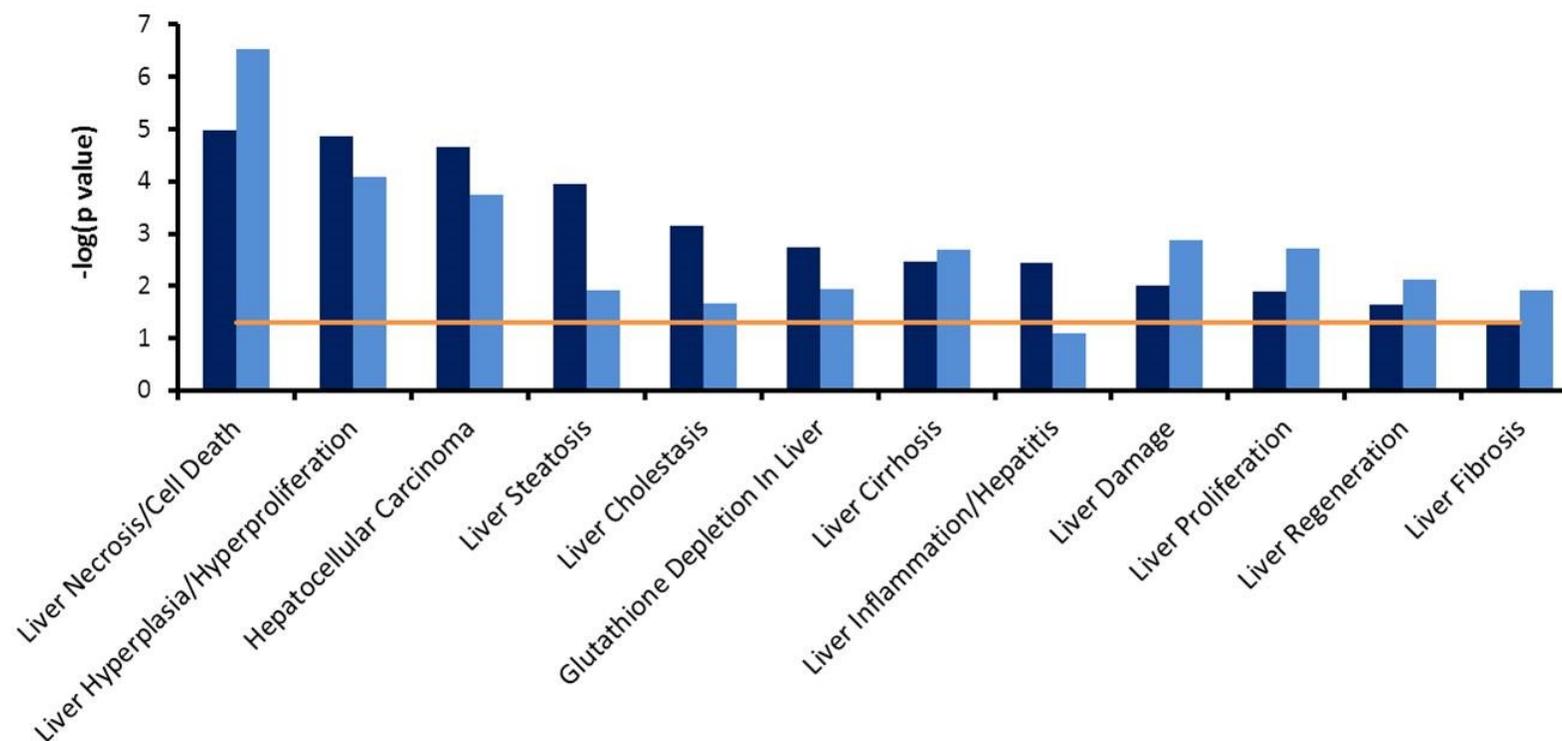


Figure 3.2 Hepatic ToxFUNCTION enrichment analyses. All IPA hepatotoxicity ToxFUNCTIONs that were enriched for in the 8 mkd furan dose group by the $-BrdU$ (dark blue) and $+BrdU$ (light blue) gene lists. The $P \leq 0.05$ significance threshold is indicated in orange ($-\log(p \text{ value}) = 1.3$); only ToxFUNCTIONs with at least 4 molecules were considered.

CHAPTER 4

4 CHAPTER: IMPACT OF PLATFORM AND STATISTICAL FILTERING ON TRANSCRIPTIONAL BENCHMARK DOSES AND MULTIPLE APPROACHES FOR SELECTION OF POINT OF DEPARTURE

Modified from: A. F. Webster, Chepelev, N., Gagne, R., Kuo, B., Recio, L., Williams, Yauk, C.L. *PLoS ONE* 10(8): e0136764. doi: 10.1371/journal.pone.0136764.

4.1 Preamble

Modeling the dose-response of a toxicant-dependent effect is an important aspect of quantitative chemical risk assessment. There are publicly available BMD analysis software tools that are commonly used for this purpose. These tools perform curve fitting of dose-response data in order to identify the specific dose (the BMD) at which the biological response differs from background levels. The BMD of an important, toxicant-dependent adverse outcome can be designated as the toxicant's PoD, which is a value that can be used for regulatory decision-making. Historically the BMD approach has been used to model apical (phenotypic) endpoints; however, with the advent of BMDExpress software, high-throughput BMD modeling of large transcriptomic datasets has become possible. In Chapter 4 we explore a number of aspects of transcriptional BMD modeling, using furan as an example. In particular we investigate the effects of genomics platform and statistical filtering of data on BMD values. We also explore a variety of strategies for PoD selection. Throughout the study, transcriptional BMD values are anchored to known cancer BMD values.

4.2 Abstract

Many regulatory agencies are exploring ways to integrate toxicogenomic data into their chemical risk assessments. The major challenge lies in determining how to distill the complex data produced by high-content, multi-dose gene expression studies into quantitative information. It has been proposed that BMD values derived from toxicogenomics data be used PoD values in chemical risk assessments. However, there is limited information regarding which genomics platforms are most suitable and how to select appropriate PoD values. In this study, we compared BMD values modeled from RNA sequencing-, microarray-, and qPCR-derived gene expression data from a single study, and explored multiple approaches for selecting a single PoD from these data. The strategies evaluated include several that do not require prior mechanistic knowledge of the compound for selection of the PoD, thus providing approaches for assessing data-poor chemicals. We used RNA extracted from the livers of female mice exposed to non-carcinogenic and carcinogenic doses of furan for three weeks. We show that furan's transcriptional BMD values are quite consistent across technologies and are predictive of the two-year cancer bioassay-based PoD. We also demonstrate that filtering data based on statistically significant changes in gene expression prior to BMD modeling creates more conservative BMD values. Taken together, this study demonstrates that high-content toxicogenomics studies produce robust data for BMD modelling that are minimally affected by inter-technology variability and highly predictive of cancer-based PoD doses.

4.3 Introduction

One important quantitative metric in risk assessment is the BMD, which is the dose at which there is a change in a biological response compared to background levels. BMD modeling was adopted by the US EPA to improve upon traditional NOAEL/LOAEL (no observed adverse effect level; lowest observed adverse effect level) approaches (Davis *et al.* 2011). Since toxicogenomics studies produce a very large amount of data, most of which can be BMD modeled, these studies generate a huge number of BMD values (representing BMDs for individual genes, molecular pathways, gene ontologies, and more) from which a single PoD must be identified. A PoD is typically chosen as the dose at which an important, disease-predicting biological response departs from background levels. A major challenge in using toxicogenomics data for this purpose lies in determining how to distill these complex datasets into manageable pieces of usable information.

Ultimately, if toxicogenomics data are to be included in chemical risk assessment, a consensus must be reached regarding which genomics platforms are appropriate, how the data should be modeled, and how to choose appropriate transcriptomic PoD values. DNA microarrays have been the predominant technology used in toxicogenomics studies over the past decade. However, microarrays are experimentally constrained by pre-defined probe sequences and limited dynamic ranges. RNA-sequencing (RNA-seq), which does not have these experimental limitations, is beginning to be used more routinely to quantify transcript abundance. RNA-seq also provides information on alternative splicing, novel transcripts, and, since it is possible to adjust sequencing depth, can detect low abundance transcripts with greater accuracy than microarrays (Wang *et al.*

2014). However, there are currently no established best practices for handling RNA-seq data (e.g., to align, filter, normalize, and identify differentially expressed RNAs), which represents a barrier to its use in applied and regulatory settings (as these require a degree of experimental reproducibility and transparency). In order to consider RNA-seq as an alternative to microarrays, it will be important to understand how these two high-content technologies compare with respect to the mechanistic insight and quantitative outputs they produce.

Two recent studies have evaluated the abilities of RNA-seq and microarrays in producing comparable and reliable quantitative data for human health risk assessment. To determine whether a shift from DNA microarrays to RNA-seq might influence transcriptional BMDs, Black *et al.* (2014) applied BMD modeling of genes and pathways for bromobenzene-dependent gene expression changes in rat liver measured using both RNA-seq (on a 5500xl Series SOLiD Next Generation Sequencer) and Affymetrix microarrays. These authors reported low to moderate concordance ($r= 0.2-0.5$) between the two technologies for pathway-based BMD values. The modest pathway-based BMD correlation was attributed to differences in the dynamic range of the two technologies and the different normalization methods used. Wang *et al.* (2014) compared toxicant-dependent changes in rat liver gene expression between the two platforms for 27 chemicals. They reported greater inter-platform concordance of gene expression for chemicals that elicit larger transcriptional effects (both in magnitude of fold-change and in number of differentially expressed genes, DEGs). They also reported that the low inter-platform correlation of DEG fold-change is attributable to the inferior ability of microarrays to detect low abundance transcripts. These two studies highlight some of the

difficulties encountered when comparing microarray and RNA-seq data, and demonstrate the need for additional research to address these inter-platform inconsistencies.

In the present study we compared the dose response of transcriptomic data obtained from samples using three genomics platforms: Agilent DNA microarrays, Illumina poly-A RNA-sequencing and custom RT2 Profiler PCR arrays. We used RNA extracted from the livers of female mice sub-chronically exposed to furan to make inter-platform comparisons of DEG fold-change, gene BMDs, and pathway mean- and median-BMDs. We compared these transcriptomic BMDs to the published BMDs for furan-induced cancers. We further compared BMD values derived from full gene lists to those derived from gene lists that had been filtered for differential gene expression. Finally, we explored two established and two novel approaches for PoD selection from transcriptional data: 1) the lowest BMD(L)-mean for a molecular pathway (Thomas *et al.* 2011c, Thomas *et al.* 2013c); 2) the lowest BMD(L)-mean for a molecular pathway that has been validated using a second gene expression platform; 3) the mode, mean, and median of all pathway BMD(L)-means; and, 4) the BMD of a key MoA-based signaling pathway (Clewell *et al.* 2014, Jackson *et al.* 2014). The purpose of this study is to understand how inter-platform differences affect gene expression data, particularly the dose-response, in order to increase user confidence in toxicogenomics data and to propose best practices for use in chemical risk assessment.

4.4 Materials and Methods

4.4.1 Animals and Exposures

Animal exposures are described above (section 2.4.3).

4.4.2 RNA Extraction

RNA extraction is described above (section 2.4.4).

4.4.3 RNA-Sequencing

Standard polyA-enrichment Illumina RNA-sequencing (RNA-seq) libraries were prepared following Illumina's TruSeq v2 Library Preparation protocols. Briefly, polyA-enriched RNA from four biological replicates per dose group was chemically fragmented, primed using random hexamers, and converted into cDNA libraries. Illumina adapters were ligated to the resultant libraries and libraries were paired-end sequenced on a HiSeq2000 with a sequencing depth of 40 million reads and a read length of 100 base pairs. The data were processed using Illumina's Real Time Analysis software, and converted into FASTQ files using CASAVA software (Illumina). The FASTQ files were aligned to the GRCm38 mouse genome with STAR (Dobin *et al.* 2013) using default parameters. Gene expression values were calculated for the Ensembl Gene Set (GRCm38v75). Feature counting was performed with Python using HTSeq-count (version 0.6.1) (Anders *et al.* 2015) with the m parameter set to "intersection-nonempty". DEG analysis was then performed using the limma pipeline (Smyth 2005) for the following set of contrasts: control versus 2 mkd furan, control versus 4 mkd furan, and control versus 8 mkd furan. Genes were considered 'present' if there were at least 0.5 counts per million (cpm) in at least 3 of 4 samples, in at least one dose group (and were called 'absent' if they did not pass this filter). TMM normalization was applied before BMDEpress analysis. The dataset is publically available through the GEO, accession number: GSE64371.

4.4.4 Microarray

The microarray experimental design and analysis are described above (section 2.4.5).

4.4.5 PCR Array

Custom RT2 Profiler PCR arrays were designed with 88 target genes, 5 housekeeping genes, and 3 internal controls (SA Biosciences by Qiagen, Frederick, MD, USA). The 88 target genes were chosen based on the results of microarray experiments. The majority of genes chosen were differentially expressed (up-regulated or down-regulated) by the highest dose of furan (8 mkd). The rest were chosen because they were also (or only) differentially expressed at lower doses (2 mkd: *Saa1*, *Esr1*, *Cyp51*, *Slc10a2*; 4 mkd: *Ccng1*, *Pdgfa*, *Rnd2*, *Slc23a3*, *Bax*, *Aen*, *Tubb2c*, *Cdkn1a*, *Tuba1c*, *Ces2*, *Gdf15*). 500 ng samples of RNA (n=5 samples per dose group) were reverse transcribed using RT2 First Strand kits (Qiagen). cDNA was combined with RT2 SYBR Green Mastermix (Qiagen) and was thermocycled according to: 1x (10 min at 95°C); 40x (15 seconds at 95°C, 1 minute at 60°C), on a CFX96 Thermocycler (BioRad Laboratories, Mississauga, ON, Canada). Fold change values were calculated using the deltaCt method using the RT2 Profiler PCR Array Data Analysis web tool, version 3.5 (<http://pcrdataanalysis.sabiosciences.com/pcr/arrayanalysis.php>). Normalized deltaCt values were used for dose response modeling in BMDExpress.

4.4.6 BMD Modeling

BMDEExpress version 1.4.1 (Yang *et al.* 2007) was used to perform BMD analysis on RNAseq, microarray, and PCR array datasets. Only genes for which there was a 'present' call in at least one dose-group (i.e., 4 out of 5 samples with 'present' call for that gene in at least one dose group for microarrays; or, 3 out of 4 samples with > 0.5 cpm for the gene in at least one dose group for RNA-seq) were modeled. Prior to modeling in BMDEExpress, these datasets were then pre-filtered for differential gene expression in three ways: (1) FDR $p < 0.05$ (in at least one dose); (2) ANOVA $p < 0.05$ (in at least one dose); (3) no statistical filtering. These three filtering conditions were not applicable to the qPCR data because all genes measured by qPCR were significantly changed in at least one dose. Constant variance was assumed for the RNA-seq data in order to analyze it using the models that are currently available in BMDEExpress. While the models in BMDEExpress were not originally intended for count-based datasets, it has been used previously (Black *et al.* 2014) and it was necessary to use the same software for the analysis of data from all platforms in order to make cross-platform comparisons.

Hill, Power, Linear and Polynomial (2° and 3°) models were used to model gene expression dose-response. For each gene, the best fitting model was selected based on: (1) a nested chi-square test (cut-off of 0.05) that was used to choose between linear and polynomial models, followed by (2) the lowest Akaike Information Criterion (AIC) value for the nested, Hill and power models, and (3) a curve goodness of fit $p > 0.1$. Genes with BMD values that were higher than the highest dose were excluded. Model parameters included: maximum iterations (the convergence criteria for the model) set to 250, confidence level (the statistical lower confidence limit applied to the model that is used to

determine the BMDL) set to 0.95, benchmark response (BMR, which is equivalent to the number of standard deviations defining the BMD) set at 1.349 (which corresponds the amount of change required to shift the mean response by 10% above background), and a restriction of the power parameter to ≥ 1 . The Hill model was restricted and flagged if the k parameter of the model was less than one third of the lowest positive dose, as per Black *et al.* (2012). In the case of a flagged Hill model, the next best model was selected only if it had a goodness of fit $p > 0.05$. In the case when no other model had $p > 0.05$, the Hill model was used and modified to 0.5 of the lowest BMD value. The resulting datasets were mapped to Ingenuity Pathway Analysis (IPA) canonical pathways, which were downloaded on April 24, 2014. Mean and median BMD(L)s were reported for pathways with at least three DEGs; however, PoDs were derived from pathways with a stricter threshold of at least 4 DEGs.

BMD software (BMDS250) was used to model the cancer dose response for HCA and HCC from Moser *et al.* (2009) and the dose response of the number of DEGs at each dose from each genomics platform. Methods for the former are described above (section 2.4.7). For the latter, the BMDS wizard v1.9 was used (www.epa.gov/ncea/bmds/). Data were treated as dichotomous and the number of DEGs in any one dose was modeled as a percentage of the total number of DEGs summed across all doses. The benchmark response (BMR) was set to 10% extra risk as recommended in the Benchmark Dose Technical Guidance document (US-EPA 2012). BMD(L) values were chosen based on the lowest AIC value of models with goodness-of-fit $P \geq 0.1$ and BMD values (across all models) that differed by less than 3-fold.

4.5 Results

4.5.1 Inter-technology Comparison of Differentially Expressed Genes

There was a dose-dependent increase in the number of DEGs on each platform. The number of DEGs obtained for each dose group on each platform, together with the BMD of their dose response, are summarized in Table 4.1. The magnitude and direction of fold changes for the DEGs were compared between technologies (Fig. 4.1). The magnitudes of fold changes produced by RNA-seq and qPCR were generally greater than those produced by microarrays, which is consistent with signal compression produced by two-colour microarrays. All comparisons yielded significant correlations (linear regression $p < 0.05$); however, some regressions were driven by single points. For each inter-technology comparison, the best correlations were produced at the highest dose. Consistent with previous reports (Black *et al.* 2014, Wang *et al.* 2014, Wang *et al.* 2014, Wang *et al.* 2014), the overlap of DEGs between technologies improved when the treatment effect was large (Fig 4.2). Overlap of enriched pathways at the high dose approached 50% (Fig. 4.2).

4.5.2 Effect of Statistical Filtering of Gene Expression Data Prior to BMD

Modeling

There is some debate regarding whether or not high-content gene expression datasets should be statistically filtered (to remove genes that do not respond to the treatment in a statistically significant manner in at least one dose), and the extent to which they should be filtered (i.e., whether there should be an adjustment for multiple testing), prior to modeling in BMDExpress. To address this issue, we performed BMD

modeling for the RNA-seq and microarray datasets using three different approaches. BMD modeling was applied to: (1) unfiltered gene expression data, (2) low-stringency pre-filtered data (requiring a gene to achieve an ANOVA $P < 0.05$ in at least one dose relative to control), or (3) high-stringency pre-filtered data (requiring a gene to achieve FDR $P < 0.05$ in at least one dose compared to control). The number of transcripts modeled for each approach produced by the RNA-seq platform was: (1) 11599 for unfiltered, (2) 3778 for ANOVA $P < 0.05$, (3) 2295 for FDR $P < 0.05$. The number of transcripts modeled for each approach produced by microarrays was (1) 29847 for unfiltered, (2) 2597 for ANOVA $P < 0.05$, (3) 364 for FDR $P < 0.05$.

Applying a low- or high-stringency statistical filter prior to modeling in BMDEExpress significantly decreased the means of each group of BMD(L) values (for both RNA-seq and microarray datasets; t-test, $P < 0.0001$; Fig. 4.3; Table 4.2). Data were unimodally distributed. Mode, mean and median values were affected by filtering, an affect that was consistent across platforms; shown for: pathway BMD-means (Fig. 4.4) and pathway BMDL-means (Fig. 4.5). The variety of the models used (Hill, Power, Linear, Polynomial 2^o or 3^o) to produce gene BMD(L)s was dictated by the genomics platform and was relatively unaffected by level of statistical filtering; data filtering had the effect of streamlining BMD/BMDL plots (see Appendix C, Figure C1). After considering these data, the EPA's criteria for modeling in BMDS, which stipulates that 'there should be at least one statistically or biologically significant dose-related trend in the selected endpoint' (US-EPA 2012), and the risk assessors perspective (which is typically to use the most conservative estimates), we recommend that statistically filtered

data be used. However, for the sake of comparison, we have included the results of most analyses for all three approaches.

4.5.3 Inter-Technology Comparison of BMD Values

The inter-technology relationships between BMD (or BMDL) values for genes or pathways were weak ($R^2=0.007-0.35$) regardless of level of filtering (Fig. 4.6; Figs C2-C4). However, we observed that the cloud of data points was always concentrated in the furan-dependent liver cancer ‘PoD range’, which is defined here as the range between the HCA BMD (2.6 mkd) and the HCC BMD (5.1 mkd). This trend is especially clear for the RNA-seq versus microarray plots and the filtered data (ANOVA and FDR). Therefore, while individual BMD values are not perfectly matched between platforms, their values are informative because they clustered closely together in the cancer PoD range.

We compared mean BMD values for molecular pathways that are important for furan’s carcinogenic MoA (Jackson *et al.* 2014): NRF2 Oxidative Stress Response, Xenobiotic Metabolism Signaling, ERK/MAPK Signaling, p53 Signaling, ASK1-Bax Cell Death Signaling (renamed from: 14-3-3-mediated Signaling), and Cancer Regulation by Stathmin1 (renamed from: Breast Cancer Signaling by Stathmin1). Pathway BMD-means were similar across technologies, with overlapping confidence intervals (Fig. 4.7). Moreover, they were generally within the confidence intervals of furan-induced HCA (BMDL-BMD, 1.6-2.6 mkd) or HCC (BMDL-BMD, 4.2-5.1 mkd). Importantly, not filtering the data prior to modeling pathways in BMDExpress generally produced higher BMD values with broader BMDL-BMD confidence intervals.

4.5.4 Approaches for Deriving Transcriptional Point of Departure (PoD) Values

We considered four approaches for choosing a single PoD from the long list of transcriptional BMD values: (1) the lowest BMD(L)-mean for a molecular pathway; (2) the lowest BMD(L)-mean for a molecular pathway that has been validated using a second gene expression platform; (3) the mode, mean, or median of the BMD(L)-mean values; and (4) the BMD of a key MoA-based signaling pathway (NRF2 Oxidative Stress Response pathway, in this case). The PoDs for each technology for these four approaches are summarized in Fig. 4.8. The BMD values increased from approach 1-4. The first three approaches best approximated the HCA BMD (2.6 mkd), whereas the fourth, MoA-dependent approach best approximated the malignant form of furan-induced cancer, HCC (BMD: 5.1 mkd). Although the RNA-seq and microarray experiments relied on a much larger pool of DEGs, the qPCR-produced BMD-mean values were remarkably similar. Applying a statistical filter (ANOVA or FDR) changed the identity of the most sensitive pathway; however, the resulting BMD values that were generated by approaches 1 and 2 were quite similar in the filtered and unfiltered data. When looking at the distribution of the data (Figs. 4.4, 4.5), it is clear that mode, mean, and median values of the BMD-mean were affected by statistical filtering; however, within each filter group, they were quite consistent across technologies. The BMD-mean values produced for the NRF2 Oxidative Stress Response pathway were quite similar between the filtered and unfiltered datasets and across technologies. Taken together, these four approaches to PoD derivation produced highly comparable results across platforms. Filtering the data tended to reduce the BMD (and therefore PoD) value; however, the confidence intervals of the filtered and unfiltered datasets often overlapped.

4.6 Discussion

In the present study we compared gene expression data obtained from furan-exposed mouse liver samples that were analyzed using three genomics platforms: Agilent DNA microarrays, Illumina polyA-enrichment RNA-seq, and custom RT2 Profiler PCR arrays. We examined: 1) inter-platform consistency of DEG fold change, and DEG and pathway BMD values; 2) the effect of filtering data based on statistically significant changes in gene expression prior to modeling in BMDEExpress; 3) the ability to produce transcriptional BMD values that are predictive of known cancer BMD values; and 4) different approaches for selecting a single PoD from a toxicogenomics dataset for use in risk assessment. Our findings provide insight into how toxicogenomics data from RNA-seq and microarrays align. This is critical for the transition to RNA-seq in toxicology, as the existing literature on the utility of toxicogenomics in risk assessment is predominantly derived from microarrays. Importantly, we demonstrate that important furan-dependent transcriptional changes occur in the liver (after three weeks) at the same dose range as furan-dependent hepatocarcinogenesis (after two years) regardless of technology. Therefore, our findings support the notion that transcriptional changes following sub-chronic treatments in rodents are prognostic of the harmful dose ranges at which adverse phenotypes occur.

4.6.1 Inter-Platform Comparison

Typically, when new technologies emerge they are compared against existing technologies. In genomics, qPCR is often used to validate microarrays, and now microarrays are being used to corroborate RNA-seq. However, DEG lists obtained using

these two technologies often have modest overlap (Wang *et al.* 2014, Black *et al.* 2014). Overlap is known to improve when the degree of transcriptional perturbation is increased (i.e., at higher doses), which is what we observed for furan. The direction and magnitude of change of the DEGs that did overlap was usually consistent, and inter-technology DEG lists were all positively and significantly correlated (Fig. 4.1). It is generally accepted that as the cost of next-generation sequencing continues to fall and the technical challenges inherent to processing, analyzing and storing these large datasets are overcome, RNA-seq is likely to become the dominant technology (Roy *et al.* 2011). In fact, when comparing RNA-seq to microarrays some investigators have emphasized the inter-platform differences (as opposed to similarities) to make this point. For example, Zhao *et al.* (2014a) demonstrated the richness of RNA-seq data (which can measure novel transcripts and splice variants) over microarrays, and a greater ability to quantify low- and highly-expressed transcripts (which is more difficult on microarrays due to the limited dynamic range). However, while such studies are excellent for showcasing the technological superiority of RNA-seq, the low overlap of DEGs remains troubling since we do ultimately need to evaluate whether each technology is able to detect equivalent biological effects. Reassuringly, it appears that the inter-technology differences are less pronounced in higher-order biology (such as molecular pathways and gene ontologies). Overall, our results and the results of others (Wang *et al.* 2014) suggest that these two genomics platforms are most comparable when there is a large treatment effect and higher-order endpoints are analyzed.

As a whole, individual BMD values did not correlate well between platforms, regardless of level of filtering (Fig 4.6). However, it is interesting to note that most of the

transcriptional BMDs fell within the HCA-HCC BMD range and that precision increased filter stringency (Figs. C2-C4). Black *et al.* (2014) also reported low to moderate concordance ($r= 0.19- 0.43$) between RNA-seq and microarray for pathway BMD values in a study that compared the transcriptional dose-response in rat liver following exposure to bromobenzene, a hepatotoxicant. This notwithstanding, it has been well demonstrated that transcriptional and apical BMDs are often well correlated (Thomas *et al.* 2012, Thomas *et al.* 2013c, Jackson *et al.* 2014, Clewell *et al.* 2014, Moffat *et al.* 2015). Thus, while there is a low inter-platform concordance of BMD values, they tend to remain in a range that is prognostic of the doses at which adverse apical outcomes occur.

4.6.2 Statistical Filtering

In addition to comparing the biology detected by each platform, we were also interested to know how the transcriptional dose responses compared. Each platform showed a dose-dependent increase in number of DEGs and the BMDs of these increases were quite near the already known BMD for furan-induced HCA (Table 4.1). Producing BMD values for each DEG was less straightforward; BMDExpress is flexible software that provides different options for filtering a dataset (based on statistically significant changes in gene expression) prior to modeling. However, since there is little guidance regarding the ‘correct’ approach for treating the data, the field has become divided as to what the ‘best practices’ should be. One argument is that no threshold for statistical significance of gene expression should be applied and that all genes that are detected should be modeled. This argument is based on the assumption that the greater the number of endpoints that are modeled, the more robust the BMDs will be. In contrast, others

argue that only DEGs should be modeled, since these are the endpoints that actually responded to the chemical treatment. For the latter, various thresholds for differential expression are possible in the BMDEpress software, including an FDR $p < 0.05$ or a less stringent ANOVA $p < 0.05$. Indeed, it has been our experience that discrepant recommendations relating to the need (or lack thereof) to pre-filter data are made, which was one motivation for conducting this work. To our knowledge, we are the first to thoroughly investigate whether the application of these statistical filters has an effect on gene and pathway BMD values, which is fundamental to making evidence-based decisions regarding their use.

We observed that applying a filter for differential gene expression prior to modeling in BMDEpress significantly reduces the means of the gene and pathway (mean and median) BMD(L) values (Figs. 4.3-4.5), indicating that more conservative BMD values are produced when only significantly changing transcripts are modeled. Importantly, the more stringent the filter, the better the mean and median BMD values approximated the BMD for furan-dependent HCA induction, which is not surprising given that applying these filters ensures that only the endpoints that truly respond to the treatment are considered. For this reason, we recommend that at least an ANOVA filter be applied to the data prior to modeling in BMDEpress. Indeed, for furan, after applying an ANOVA filter substantial pools of 3778 (RNA-seq) and 2597 (microarray) genes remain, which should allay concerns regarding the sufficiency of the number of endpoints modeled. Therefore, we contend that (at least in the case of furan) the argument that modeling a larger number of genes will produce a more exact BMD is erroneous.

4.6.3 Point of Departure Selection

While it is encouraging that many of the gene and pathway BMDs cluster together in the same dose range that causes furan-induced liver cancer, it is still necessary to determine which BMD should be used as the PoD for risk assessment. We explored four approaches for POD selection using our furan data (Fig. 4.8). Using the first approach, the lowest mean BMD(L) for a molecular pathway (with at least four molecules), we obtained transcriptional BMD values for each genomics platform that ranged from 1.7-2.2 mkd, which are very close to the HCA BMD (2.6 mkd). In addition, the most sensitive pathways reported here for furan involve cell death-, TNF-, and cancer-related processes, all of which are relevant to furan's MoA. These data support the idea that the lowest pathway BMD can be used as a surrogate PoD in the absence of apical data and that it is similar to the values established using traditional approaches. This approach has been used previously. For example, using five test chemicals, Thomas *et al.* demonstrated that the BMD and BMDL values of the most sensitive GO categories (measured following a 13 week chemical exposure) were predictive of PoD doses for cancer and non-cancer endpoints. They also showed that these transcriptional endpoints were well correlated with corresponding apical endpoints (Thomas *et al.* 2011a, Thomas *et al.* 2012). A recent case study of the genotoxic carcinogen benzo[a]pyrene also report similar transcriptional and apical BMDs for corresponding endpoints using the most sensitive transcriptional endpoints (Moffat *et al.* 2015). This approach of using the lowest BMD(L)-mean of a perturbed molecular pathway is useful for data-poor compounds (i.e., chemicals for which there are few, or no, apical data) because no knowledge of the MoA is required for PoD selection.

Our second proposed approach to PoD selection, to use the lowest mean BMD(L) for a molecular pathway that could also be modeled by another technology, is a safeguarded version of the first. We propose adding this validation step to the first approach in order to boost confidence in the PoD dose and avoid basing conclusions on one single assay or endpoint. In the case of furan, the most sensitive pathways for the microarray (OX40 Signaling Pathway) and qPCR (Molecular Mechanisms of Cancer) experiments remained unchanged; however, the most sensitive pathway for the RNA-seq experiment, Tumouricidal Function of Hepatic Natural Killer Cells, could not be modeled by either microarray or qPCR; thus, this pathway did not meet the criteria. The most sensitive RNA-seq pathway that could be modeled using data from at least one of the other two technologies was TNF-related Weak Inducer of Apoptosis (TWEAK) Signaling. In this case, choosing a pathway that could be modeled using data generated from at least one other technology increased the PoD from 1.7 to 1.9 mkd (Fig. 4.8B), which is slightly closer to the HCA BMD (2.6 mkd). These first two approaches to PoD selection each produce biologically plausible PoDs for furan. It appears that choosing a PoD based on one technology is sufficient; however, if there are data available from an additional platform, modeling and considering these data is worthwhile. Finally, we note that filtering for differential gene expression only marginally affected the values obtained using both ‘lowest BMD(L)’ approaches, suggesting that they are quite robust.

Historically, the regulatory tests used in risk assessment examine one (or a limited number of) endpoint(s). Therefore, regulatory thresholds have typically been set using the dose at which a single, sensitive, apical endpoint is altered. However, unlike these standard assays, toxicogenomics tests tens of thousands of endpoints (including

expression changes of individual genes, pathways, and ontologies); therefore, applying a traditional approach that favors a single, sensitive data-point and ignores the rest seems to defeat the purpose of performing a global analysis. Our third approach to PoD selection, selecting the mode, mean, or median of the BMD(L)-mean values (across all pathways), can also be applied to compounds with unknown MoAs (Fig. 4.8C). However, we believe that using one of these metrics is a more robust approach to PoD selection because these values represent all of the data (as opposed to a single, sensitive data point). In particular, the mode represents the peak or maximum of the distribution, and therefore the dose at which the majority of the transcriptional responses occur, thereby ensuring that the PoD decision is anchored to a very large proportion of the transcriptional data. The distribution of the gene and pathway BMD(L) values for furan was generally unimodal (Figs. 4.4, 4.5), which facilitates the use of the mode (presumably, if a toxicant produced a multi-modal distribution, one would choose to use the mode of the most sensitive peak). The phenomenon that there is a unimodal peak in transcriptional activity that is reflected in the distribution of BMD values has also been reported for median BMD and BMDL values of naphthalene-dependent GeneGo signaling pathways (Clewell *et al.* 2014). Therefore, our third approach, selecting the mode, mean, or median of the BMD(L)-mean values (across all pathways), offers an alternative way to estimate the system's perturbation that relies on the weight of evidence provided by the entire toxicogenomic dataset and allows for the selection of a PoD that reflects the BMD at which the majority of the transcriptional reaction to the compound occurs.

Unlike the first three approaches, our final method for PoD selection is appropriate for well-studied chemicals (such as furan). Using the BMD-mean of a MoA-

relevant pathway requires at least some understanding of the chemical's biological effects, which is an important aspect of establishing the human relevance of rodent data (Holsapple *et al.* 2006). We previously argued that chronic activation of the NRF2 transcription factor plays a pivotal role in the malignant transformation of regenerating liver tissue following furan-induced injury (Jackson *et al.* 2014). Therefore, we chose the NRF2-mediated Oxidative Stress Response pathway as the MoA-relevant transcriptional endpoint for furan-induced liver cancer (Fig. 4.8D). Interestingly, this approach was the best at predicting the dose at which the malignant form of furan-induced cancer (HCC) occurs. Therefore, using a MoA-specific BMD may be helpful aligning the BMD selected for PoD with the more severe adverse outcomes. In a recent study by Moffat *et al.* (2015), the authors compared the results of three independent risk assessments of benzo[a]pyrene (BaP): one that used only apical data, one that used only transcriptomic data, and one that was a hybrid of the two. The transcriptional PoD for BaP was selected based on BMD modeling of MoA-relevant pathways and they reported that the conclusions made using these data were consistent with the conclusions that were made in the traditional and hybrid assessments. In another study, a correlation between MoA-relevant pathway BMD values and histological BMD values was reported in which tissue- and gender-specific BMD differences observed for histological endpoints were conserved in the transcriptional BMD modeling results (Clewell *et al.* 2014). Thus, deriving a PoD from a MoA-relevant key event is a suitable approach for well-characterized compounds.

4.6.4 Conclusions

Regulatory agencies worldwide are challenged with assessing health risks for legacy chemicals (chemicals that remain in the environment long after they were first introduced), many of which have minimal toxicity information. It has been proposed that toxicogenomics is an effective screening tool for such data-poor chemicals because toxicogenomics experiments can be conducted more quickly and using fewer resources than many standard approaches (Dix *et al.* 2007). Work from our laboratory has demonstrated that toxicogenomics data are useful for both mechanistic and quantitative risk assessment. Here we show that BMD values generated for the most sensitive pathways, the transcriptional modes, means, and medians, or individual MoA-relevant pathways were highly consistent with one another, with furan-dependent cancer BMDs, and across genomics platforms. We note that statistical filtering of data prior to modeling creates BMDs that are more conservative and less likely to over-estimate the adverse-outcome that they are intended to predict. We caution that it is possible that the remarkably high concordance rates achieved here may be, at least partially, due to the limited dose range used. Therefore, we recommend that this type of study be repeated using additional chemicals and more extensive dose ranges. It seems clear that there are a number of approaches that can be effectively used to choose a PoD dose, and that most of these do not require prior knowledge of the compound's toxicity. This study represents an important step toward confidently and effectively applying toxicogenomic data to quantitative risk assessment. The comparisons made here can be used to make evidence-based decisions regarding the experimental design of future toxicogenomics studies that include BMD modeling.

4.7 Tables and Figures

Table 4.1 Gene and cancer dose response. The number of differentially expressed genes (by unique gene symbol) and their dose response (upper), and the BMDs of furan-dependent liver cancers (lower).

	Number of DEGs per dose group		
	RNA-seq	Microarray	qPCR
2 mkd	131 (12%)	17 (5%)	23 (26%)
4 mkd	221 (20%)	28 (8%)	32 (36%)
8 mkd	1041 (94%)	350 (98%)	71 (81%)
Total DEGs[†]	1113	356	88
BMD(L)[‡]	1.90 (0.53)	2.78 (1.04)	0.67 (0.37)
Cancer BMD values *			
	<i>Model</i>	<u>BMD(L)</u>	
Hepatocellular adenoma (HCA)	<i>Multistage</i>	2.6 (1.6)	
Hepatocellular carcinoma (HCC)	<i>Logistic</i>	5.1 (4.2)	
HCA or HCC	<i>Multistage</i>	2.3 (1.3)	

DEGs are defined as a gene with an FDR $p < 0.05$, fold change $> \pm 1.5$

[†]Total number of unique DEGs across all three dose groups (within each platform)

[‡]Modeled in BMDS; dichotomous data presented as percent present; BMR=10% extra risk

*Data from Moser et al. 2009

Table 4.2 Statistics for box-and-whisker plots (Fig. 4.3), which show that (within each genomics platform) means of each group are significantly different.

Analysis	Platform	Comparison	BMD					BMDL				
			Difference	Std.Err	Obs t-stat	DF	p-value	Difference	Std.Err	Obs t-stat	DF	p-value
Genes	RNA-Seq	ANOVA filtered vs Unfiltered	-0.680	0.0419	16.25	8485	0.0000	-0.297	0.0289	10.26	8075	0.0000
		FDR filtered vs Unfiltered	-1.002	0.0465	21.55	4783	0.0000	-0.479	0.0322	14.90	4472	0.0000
		FDR filtered vs ANOVA filtered	-0.321	0.0485	6.63	4886	0.0000	-0.183	0.0344	5.31	4904	0.0000
	Microarray	ANOVA filtered vs Unfiltered	-0.723	0.0499	14.47	5735	0.0000	-0.290	0.0348	8.34	5538	0.0000
		FDR filtered vs Unfiltered	-1.864	0.0800	23.31	503	0.0000	-0.918	0.0566	16.22	489	0.0000
		FDR filtered vs ANOVA filtered	-1.141	0.0832	13.71	584	0.0000	-0.628	0.0592	10.60	581	0.0000
Pathway (mean)	RNA-Seq	ANOVA filtered vs Unfiltered	-0.684	0.0612	11.18	752	0.0000	-0.312	0.0402	7.75	750	0.0000
		FDR filtered vs Unfiltered	-1.117	0.0582	19.20	636	0.0000	-0.548	0.0386	14.18	627	0.0000
		FDR filtered vs ANOVA filtered	-0.433	0.0564	7.69	554	0.0000	-0.236	0.0378	6.25	552	0.0000
	Microarray	ANOVA filtered vs Unfiltered	-0.558	0.0497	11.23	631	0.0000	-0.192	0.0354	5.44	628	0.0000
		FDR filtered vs Unfiltered	-1.925	0.0707	27.21	140	0.0000	-1.016	0.0564	18.00	126	0.0000
		FDR filtered vs ANOVA filtered	-1.367	0.0737	18.55	161	0.0000	-0.823	0.0584	14.09	144	0.0000
Pathway (median)	RNA-Seq	ANOVA filtered vs Unfiltered	-0.727	0.0692	10.51	740	0.0000	-0.419	0.0425	9.85	730	0.0000
		FDR filtered vs Unfiltered	-1.152	0.0655	17.59	642	0.0000	-0.636	0.0400	15.89	630	0.0000
		FDR filtered vs ANOVA filtered	-0.425	0.0552	7.70	558	0.0000	-0.217	0.0325	6.69	558	0.0000
	Microarray	ANOVA filtered vs Unfiltered	-0.445	0.0628	7.08	642	0.0000	-0.235	0.0444	5.30	645	0.0000
		FDR filtered vs Unfiltered	-1.862	0.0680	27.39	273	0.0000	-1.101	0.0539	20.42	210	0.0000
		FDR filtered vs ANOVA filtered	-1.417	0.0639	22.18	232	0.0000	-0.866	0.0518	16.70	186	0.0000

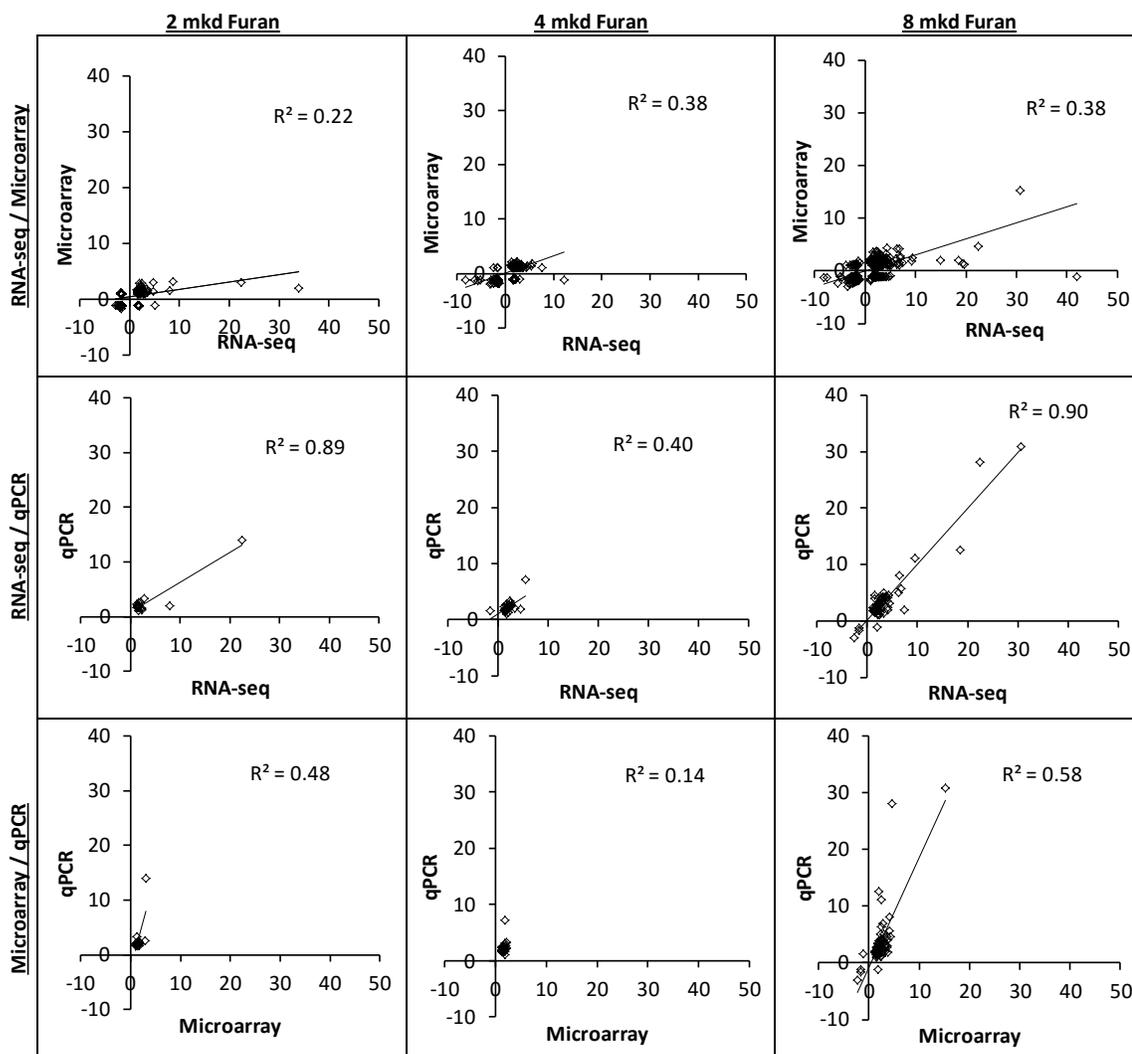


Figure 4.1 Gene expression levels are correlated between genomics platforms. Correlation analysis of the fold changes of differentially expressed genes (FDR $P < 0.05$, fold change $> \pm 1.5$ in at least one dataset) for 2 mkd (left), 4 mkd (middle), and 8 mkd (right) furan relative to controls. All regressions were significant ($P < 0.05$).

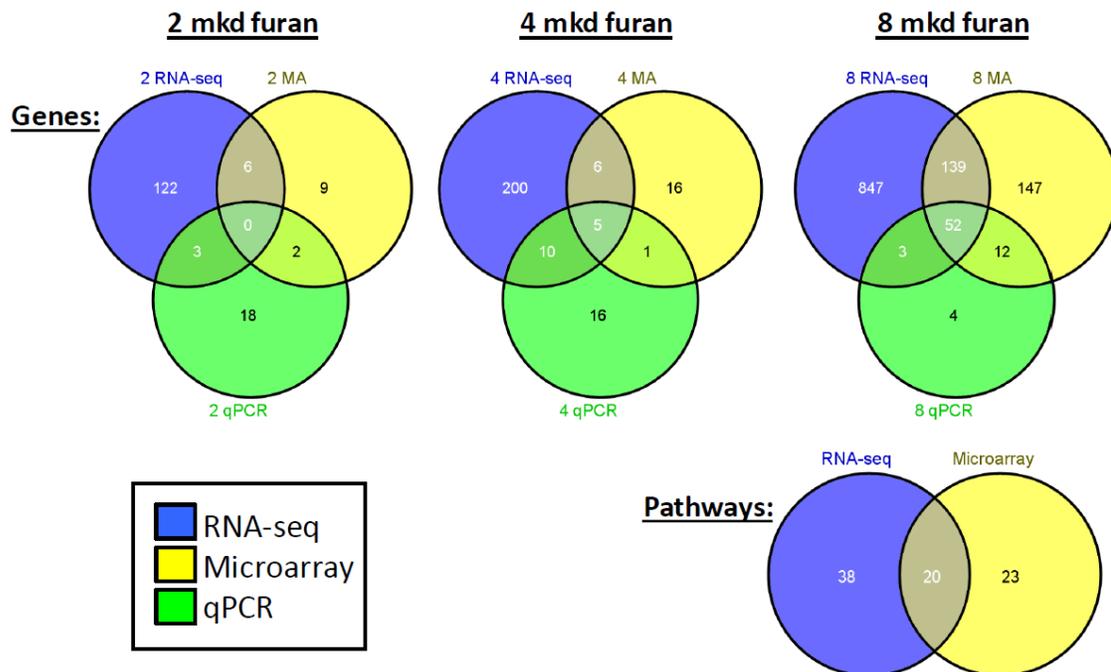


Figure 4.2 The intra-dose, inter-platform overlap of DEGs (by unique gene symbol; top) and pathways (bottom).

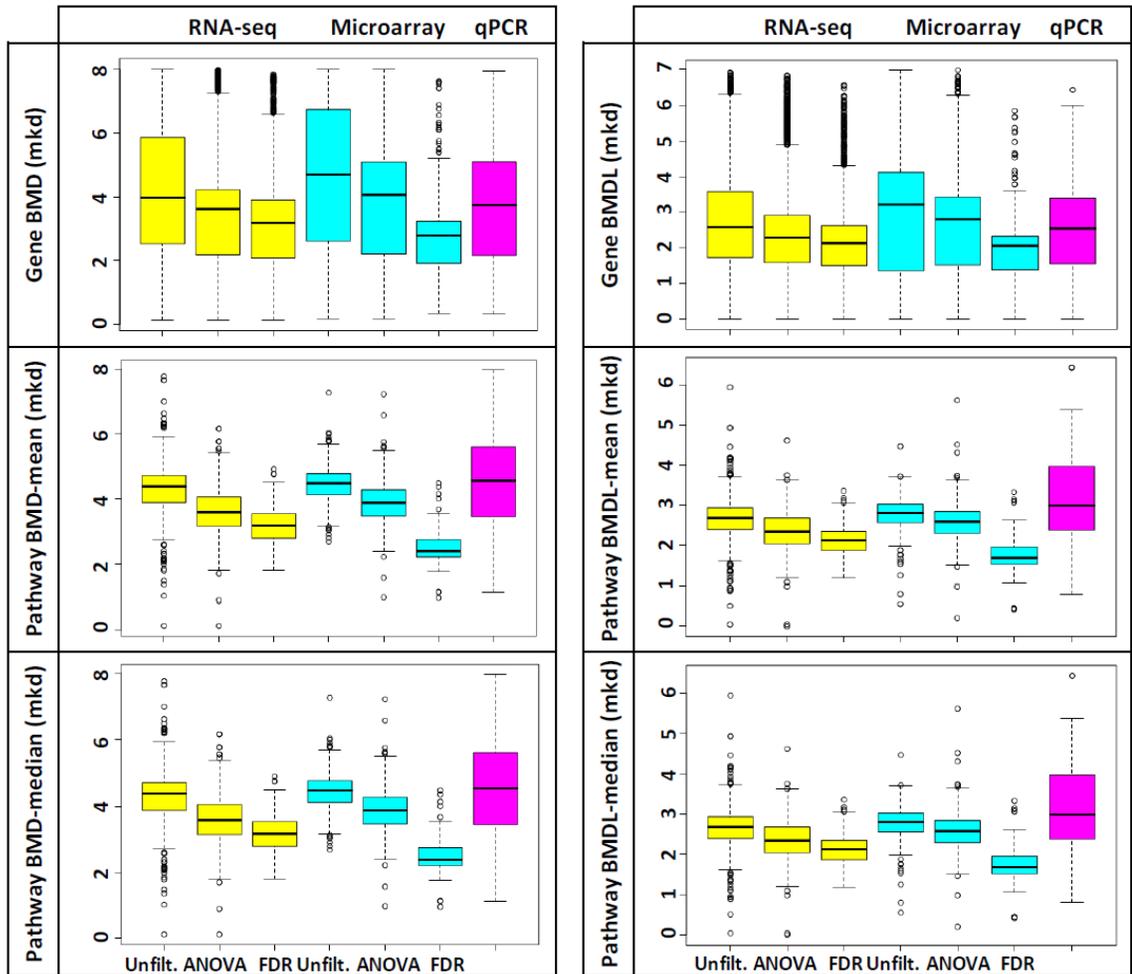


Figure 4.3 Filtering gene expression data significantly changes mean values for gene and pathway BMD(L)s. The distributions of the BMD (left) and BMDL (right) values for genes (top), pathway means (center), and pathway medians (bottom), from the RNA-seq (yellow), microarray (aqua), and qPCR (pink) experiments, were compared. Upper and lower quartiles are indicated by box, the median is indicated by the line within the box and points represent outlier BMD(L) values (upper quartile + 1.5 x the inter quartile range (IQR) or the lower quartile - 1.5 x IQR). Within each genomics platform, filtering significantly changed sample means (t-test $p < 0.0001$; Table 4.2).

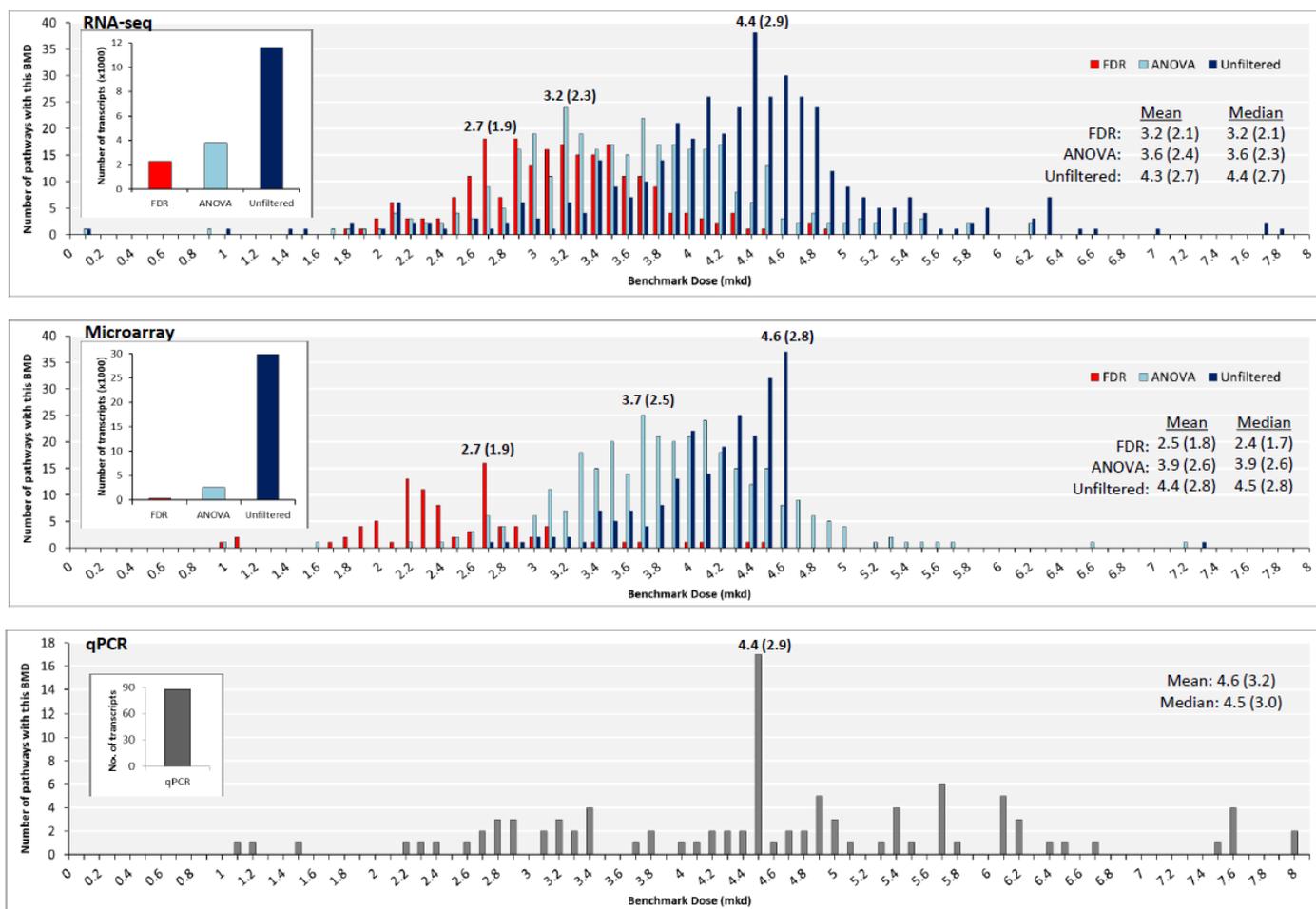


Figure 4.4 Filtering gene expression data changes the distribution of pathway BMD-mean values. Distributions of pathway BMD-mean values for RNA-seq (top), microarray (center) and qPCR (bottom). Mode BMD (BMDL) values are labeled. Modes, means and medians decrease as filtering stringency increases. Pathways were only considered in this analysis if they had four or more molecules with $P_{fit} > 0.1$. Overlain bar charts indicate the number of transcripts used to model each group.

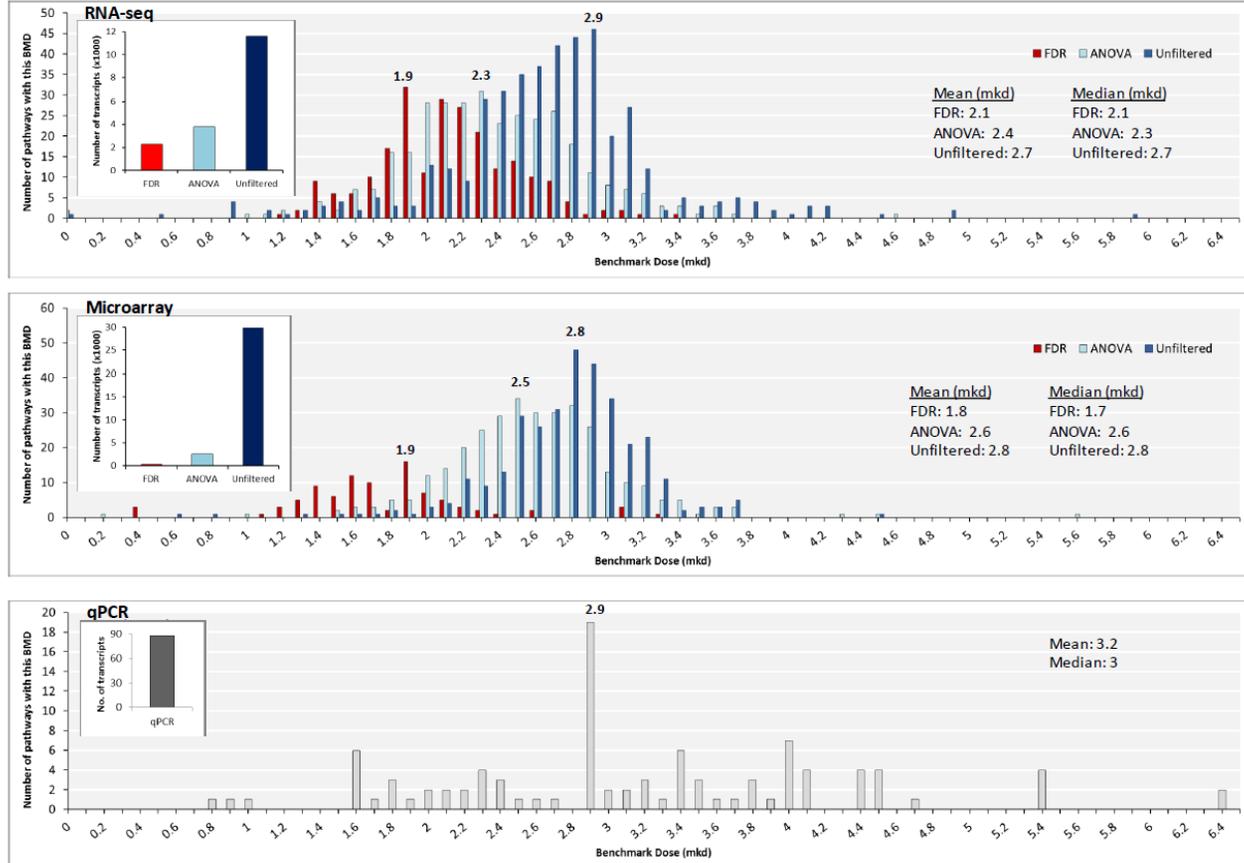


Figure 4.5 Filtering gene expression data changes the distribution of pathway BMDL-mean values. Distributions of pathway BMDL-mean values for RNA-seq (top), microarray (center) and qPCR (bottom). Mode values are labeled. Modes decrease as filtering stringency increases. Pathways were only considered in this analysis if they had four or more molecules with P fit > 0.1. Overlain are the number of transcripts used to model each group.

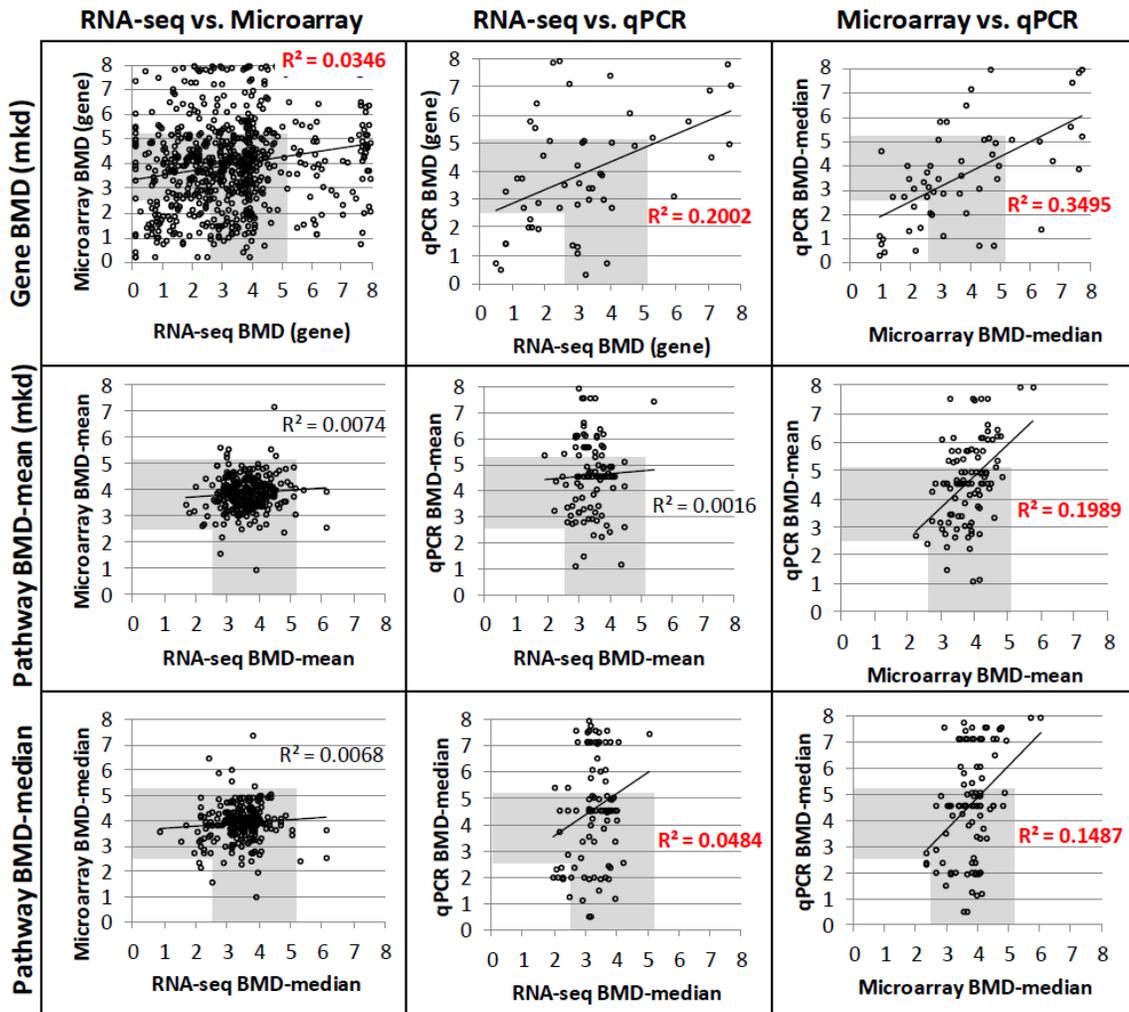


Figure 4.6 BMD values are not well correlated between genomics platforms. Inter-platform comparisons of BMDs for genes (top), pathway means (center), and pathway medians (bottom) for ANOVA filtered data. The BMD range for furan-induced HCA-HCC of 2.6-5.1 mkd is shaded in grey. Statistically significant correlations are indicated in red (regression $P < 0.05$).

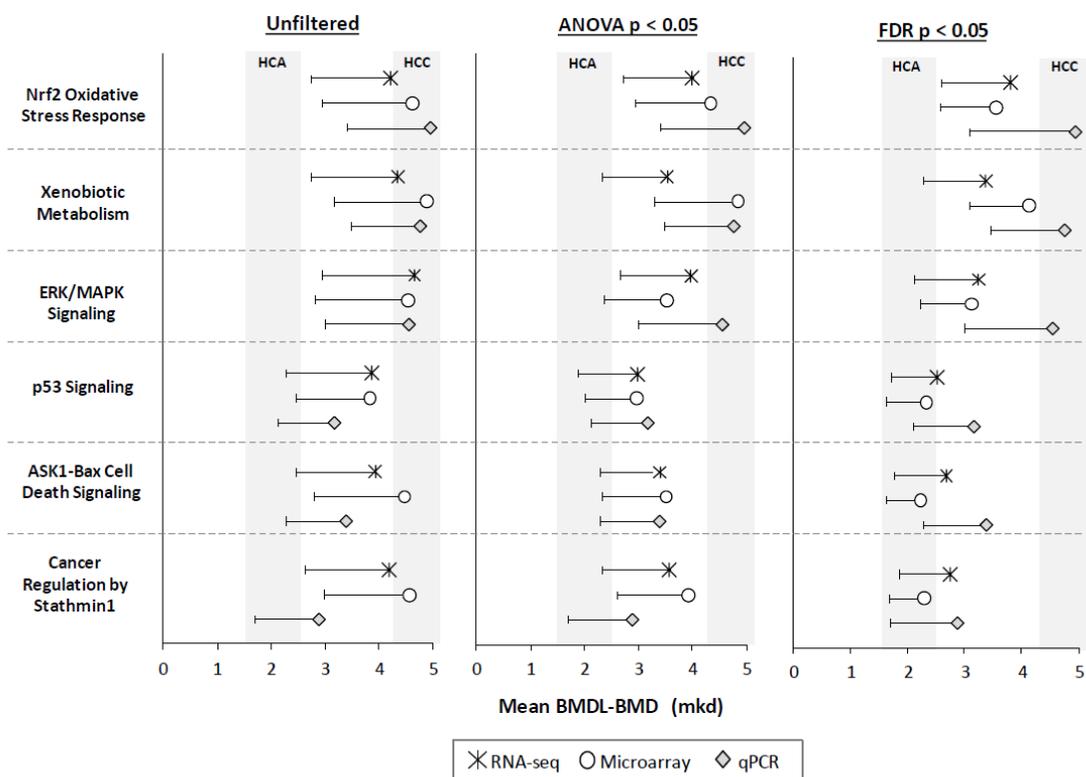


Figure 4.7 BMD-means of furan MoA pathways are consistent across platforms. Symbols indicate BMD values and whiskers indicate the lower confidence interval of the BMD. The confidence intervals (or the range between the BMDL and BMD) for HCA and HCC are shaded in grey.

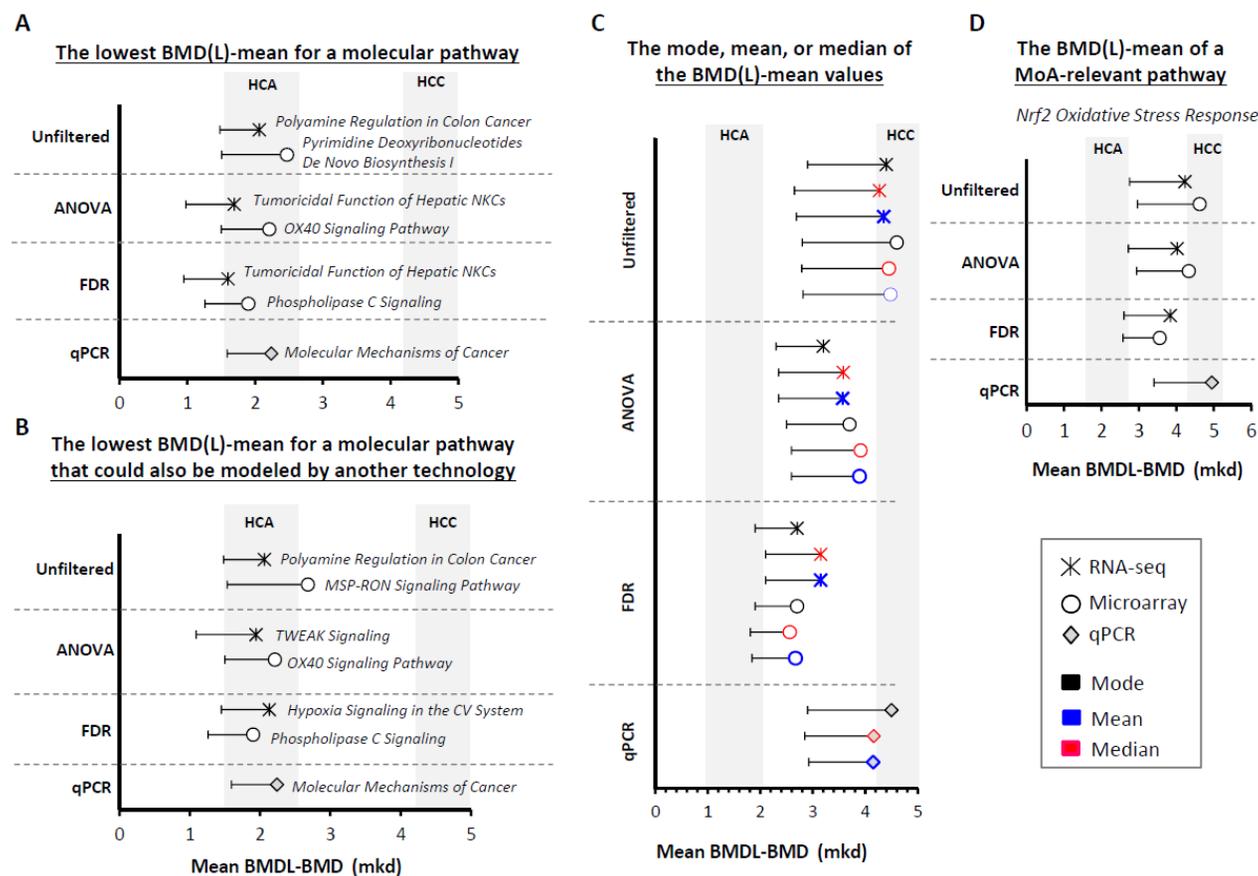


Figure 4.8 Four approaches to deriving transcriptomic PoD values. (A) The lowest BMD(L)-mean for a molecular pathway; (B) the lowest BMD(L)-mean for a molecular pathway that has been validated using a second gene expression platform; (C) the mode, mean, or median of the BMD(L)-mean values; and (D) the BMD of a key MoA-based signaling pathway. PoDs are represented as the pathway BMD-mean values (mkd) with lower confidence intervals indicated. All pathways had a minimum of four molecules that were modeled. The BMD confidence interval for each furan-dependent liver cancer is shaded in grey. NKC=Natural Killer Cells; CV=Cardiovascular; TWEAK=TNF-related weak inducer of apoptosis.

CHAPTER 5

5 CHAPTER: PREPARATION OF ARCHIVAL FORMALIN-FIXED AND PARAFFIN-EMBEDDED MOUSE LIVER SAMPLES FOR USE WITH THE AGILENT GENE EXPRESSION MICROARRAY PLATFORM

Modified from: A.F. Jackson, Williams, A., Moffat, I., Phillips, S.L., Recio, L., Waters, M.D., Lambert, I.B., Yauk, C.L. 2013. *Journal of Pharmacological and Toxicological Methods*, 68(2):260-8.

5.1 Preamble

Building on the theme of reducing the use of experimental animals in toxicity testing, we wondered if gene expression analysis could be performed using archival, FFPE tissues. This is important because globally there exists a huge number of archival FFPE tissue samples that correspond to toxicants that have been previously tested in standard assays. Performing toxicogenomic analyses on these very well characterized samples would produce robust, phenotypically-anchored molecular signatures, without the need to conduct additional animal studies. The purpose of Chapter 5 was to develop a protocol for RNA extraction from archival FFPE samples and to use these samples for global gene expression analysis using Agilent microarrays. We tested this protocol by measuring global gene expression of 16-year-old FFPE-preserved liver samples from phenobarbital-treated mice, and their matched FFPE-preserved control samples.

5.2 Abstract

Tissues samples are routinely formalin-fixed and paraffin-embedded for long-term preservation. Gene expression analysis of archival FFPE tissues may advance knowledge of the molecular perturbations contributing to disease. However, formalin causes extensive degradation of RNA. We compared RNA quality/yield from FFPE samples using six commercial FFPE RNA extraction kits. In addition, we compared four DNA microarray protocols for the Agilent 8x60K platform using 16-year-old FFPE mouse liver samples treated with phenobarbital or vehicle. Despite low quality RNA, archival phenobarbital samples exhibited strong induction of the positive control genes Cyp2b9 and Cyp2b10 by qPCR. We tested one- and two-colour microarray designs and evaluated the effects of increasing the amount of hybridized cDNA. Canonical gene responders to phenobarbital were measurably induced under each experimental condition. Increasing the amount of labelled cDNA did not improve the overall signal intensity. One-colour experiments yielded larger fold changes than two-colour and the number of differentially expressed genes varied between protocols. Gene expression changes were validated by qPCR and literature searches. Individual protocols exhibited high rates of false positives; however, pathway analysis revealed that nine of the top ten canonical pathways were consistent across experiments. Genes that were differentially expressed in more than one experiment were more likely to be validated. Thus, we recommend that experiments on FFPE samples be done in duplicate to reduce false positives. Analysis of archival FFPE samples identified pathways that are consistent with phenobarbital's mechanism of action. Therefore, we conclude that FFPE samples can be used for meaningful microarray gene expression analyses.

5.3 Introduction

Archived FFPE samples represent an invaluable resource for molecular studies of disease. Available archived tissues encompass a diversity of human and animal disease states produced from both genetic and environmental stressors (e.g., drugs and environmental chemicals) that can be used to investigate the molecular mechanisms that contribute to human disease. In the United States alone, millions of FFPE biopsy samples are stored at room temperature in tissue libraries (Bouchie 2004). Archival FFPE samples from two-year rodent cancer bioassays could be used to begin to precisely model the temporal changes in gene expression following toxic insult that are correlated with histopathological changes leading to cancer outcome in rodents. Indeed, once a large enough number of these phenotypically-anchored molecular signatures have been produced, they could be applied in comparative toxicogenomics to expedite identification of novel putative carcinogens. However, samples preserved in formalin have highly degraded RNA that may not be amenable to gene expression analysis.

Formalin is a fixative that prevents tissue decay by cross-linking and denaturing proteins. Unfortunately, formalin also causes the fragmentation of nucleic acids and cross-linking of nucleic acids and proteins. Extensive cross-linking contributes to compartmentalization of nucleic acids with proteins (which makes DNA and RNA unavailable for extraction by traditional extraction methodologies and kits). Formalin also causes the addition of mono-methylol to adenine; typically, this can result in degradation or modification of the poly-A mRNA tail such that oligo-dT (deoxythymine) primers are no longer appropriate for complementary DNA (cDNA) synthesis (Farragher *et al.* 2008, Klopfleisch *et al.* 2011). Despite these disadvantages, the

impracticalities of long-term storage of large archives of frozen (FRO) tissues have favored FFPE sample preservation.

Recent advances in protocols for genomics using FFPE tissues lend promise to the prospect of using FFPE tissue for research into the molecular basis of disease. Improvements in protocols for FFPE RNA extraction have been made (Roberts *et al.* 2009, Jacobson *et al.* 2011). Macro-molecular cross-linking can be greatly overcome by treatment with proteinase K. PolyA tail degradation can be circumvented through the use of random hexamers for cDNA synthesis (instead of using oligo-dT primers). And, since RNA extractions from FFPE tissues yield RNA polymers of less than 300 nucleotides in length that have low RNA integrity numbers (RIN ~2.0), PCR primers that amplify small oligos (~70bp) are used. Finally, whole transcriptome amplification is also conducted in order to obtain sufficient genetic material for genomic analysis. Even so, quantitative PCR (qPCR) Ct values for these samples are typically ~5 units lower than paired high quality FRO samples, and microarray signal intensities can be 50% lower (Farragher *et al.* 2008, Klopffleisch *et al.* 2011). It has also been shown that sample quality decreases linearly over storage time up to 15 years (Rentoft *et al.* 2012). Despite these experimental difficulties, consistent gene expression profiles between paired FRO and FFPE tissues have been reported.

Proof of principle that FFPE tissues can be used for molecular analyses comes primarily from research on tumours. Published comparisons between paired FRO and FFPE tumour biopsies have demonstrated comparable gene expression profiles between the two preservation techniques (Roberts *et al.* 2009). These studies have been conducted using a number of genomics platforms, including: qPCR (Antica *et al.* 2010), Affymetrix

microarrays (Lassmann *et al.* 2009, Linton *et al.* 2009, Budczies *et al.* 2011, Jacobson *et al.* 2011), and Illumina DASL and Bead-Chips (Abramovitz *et al.* 2008, Hoshida *et al.* 2008, Ravo *et al.* 2008, Di Cesare *et al.* 2010, Sadi *et al.* 2011), and Agilent Whole Human Genome 4x44K arrays (Fedorowicz *et al.* 2009). Agilent has produced a workflow for hybridizing cDNA derived from FFPE tissue and labeled in one-colour (Cyanine dye 3: Cy3) on their 4x44K platform, but no studies using archival FFPE tissues have been done using the 8x60K platform or using a two-colour microarray design (i.e., RNA samples labelled with two different dyes: Cy3 and Cy5, and hybridized to the same array). Two-colour microarray designs confer an advantage over single-colour in providing additional control of potential technical confounders. Finally, at the time of manuscript preparation, no reports had determined whether RNA derived from archival FFPE samples can be used to quantify global gene expression changes in chemically treated rodent samples using DNA microarrays.

In this study we developed a protocol for toxicogenomic analysis of archival FFPE samples on the Agilent microarray platform. Samples were 16-year-old, phenobarbital-exposed, FFPE-preserved mouse livers. Phenobarbital is an antiepileptic agent (Boothe *et al.* 2012, Pakozdy *et al.* 2012, Tulloch *et al.* 2012), a non-genotoxic rodent carcinogen, and is classified as possibly carcinogenic to humans (Group 2B) (IARC 2001). We chose to use phenobarbital in this study because it is a well-characterized chemical (Nesnow *et al.* 2009), which was important because there were no paired FRO tissues for the phenobarbital tissue blocks. Before extracting RNA from the archival samples, we compared the performance of six commercially available FFPE RNA extraction kits using freshly embedded control liver samples. We then evaluated

four protocols for hybridizing archival FFPE samples from phenobarbital-exposed mice on Agilent SurePrint mouse gene expression 8x60K arrays. Changes in transcript levels were validated by comparing the observed gene expression responses to those expected following phenobarbital exposure (i.e., literature searches), and via confirmation of changes in transcript levels using qPCR. Our goal was to develop a microarray protocol appropriate for use on archival FFPE samples that do not have paired FRO counterparts, as is the case for most FFPE samples stored in histology libraries. We investigate whether the known MoA for phenobarbital can be identified using microarray data generated from archival FFPE-preserved samples.

5.4 Materials and Methods

5.4.1 Animals and Tissues

Fresh liver tissue was collected from two healthy, untreated, 25-week-old MutaTMMouse (transgenic strain 40.6) males that were individually housed on a 12 hour light-dark cycle. Food (2014 Teklad Global standard rodent diet) and water were available *ad libitum*. Mice were bred and maintained under conditions approved by the Health Canada Animal Care Committee. Mice were euthanized by cardiac puncture under isoflurane anesthesia. The left lobe of each liver was excised and divided in three. Each of these was further divided into two; one piece from each pair was flash-frozen in liquid nitrogen and stored at -80°C and the other was fixed in 20-fold excess volume of neutral phosphate-buffered formalin (Leica Biosystems, Winnipeg, MB) for 48 hours. Fixed tissue was dehydrated and embedded in paraffin using an automated tissue processor

(Tissue Tek VIP Processor, Sakura Finetek, Torrance CA); RNA was extracted within one month of fixation.

Archival FFPE livers from phenobarbital treated B6C3F1 mice (purchased from Charles River Laboratories, Raleigh, NC; and acclimatized for 12 days prior to phenobarbital exposure) were obtained from Integrated Laboratory Systems, Inc. (ILS, Research Triangle Park, USA). 42-day-old male B6C3F1 mice were housed three per cage, in a 12 hour light-dark cycle and were fed purified tap water and NIH-07 formula (Zeigler Bros., Gardner PA) ad libitum. Mice received a sub-chronic exposure (four weeks) of 0.05% phenobarbital in drinking water (or drinking water alone; n=3) and were sacrificed one day post-treatment. Livers were fixed in neutral phosphate-buffered formalin for 36-48 hours and were stored in paraffin blocks at room-temperature for 16 years prior to RNA extraction. Experiments were conducted under the National Institute of Health (www.nih.gov) guidelines for the care and use of laboratory animals and were approved by the Chemical Industry Institute of Toxicology (www.ciit.org) Institutional Animal Care and Use Committee.

5.4.2 RNA Extraction

RNA was extracted from 2 x 5 μ m sections per freshly embedded sample. We evaluated the quality and quantity of RNA extracted from freshly preserved and freshly sectioned tissues using five commercial kits: Absolutely RNA FFPE kit (Agilent Technologies, Santa Clara, California, USA), HighPure FFPE RNA Micro kit (Roche Applied Science, Mannheim, Germany), NucleoSpin FFPE RNA kit (Macherey-Nagel, Düren, Germany), RecoverAll Total Nucleic Acid Isolation kit (Ambion, Carlsbad,

California, USA), and RNeasy FFPE RNA Extraction kit (Qiagen, Hilden, Germany). Xylene was used to deparaffinize unless a proprietary deparaffinization solution was supplied in the kit. The Qiagen, Roche, Agilent, and Ambion kits were used to evaluate corresponding slide-mounted, freshly embedded samples. We also evaluated the Roche and Agilent on-slide deparaffinization protocols (the only two companies to provide on-slide protocols). The Agilent deparaffinization procedure was subsequently used for all slide-mounted samples (Absolutely RNA FFPE kit Manual, Appendix I) because it was easier to scrape sections that were rehydrated in 70% ethanol (versus using 100% ethanol and scraping dry sections).

RNA was extracted from 1 x 10 μm slide-mounted sections for each archival sample. The three kits (Agilent, Qiagen, and Roche) that produced the best quality of extracted RNA in the freshly embedded samples, as well as the Arcturus Paradise FFPE RNA kit (Applied Biosystems, Carlsbad, California, USA), were used to extract archival RNA. RNA was extracted according to the manufacturer's protocol. Ultimately, the Qiagen kit was chosen as the best (Supplemental Protocol, section C.4.2, Table P1).

RNA yield and quality were assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA) and Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). NanoDrop spectrophotometers take absorbency measurements at 280 nm, 260 nm, and 230 nm. Ratios of these absorbencies ($A_{260}/280$ and $A_{260}/230$) are used to assess sample purity. $A_{260}/280 \approx 2.0$ and $A_{260}/230 \approx 2.0\text{-}2.2$ are considered pure for RNA samples. Bioanalyzers use the RIN as a metric for RNA purity, which has been previously described (Schroeder, et al., 2006).

5.4.3 cDNA Synthesis

We used two methods to produce cDNA depending on the downstream application. We note that both methods used random primers rather than oligo-dT primers. For qPCR, we produced cDNA from 1 µg of RNA using the High Capacity cDNA reverse transcription kit according to the supplier's protocol (Applied Biosystems). For microarrays, we produced 7-10 µg cDNA from 250 ng RNA using the Whole Transcriptome Amplification 2 (WTA2) kit according to the supplier's protocol (Sigma, St. Louis, Missouri, USA). The latter was purified using the QiaQuick PCR Clean-up kit (Qiagen): 5 x 75 µl aliquots of cDNA from the WTA2 amplification reaction were combined and purified as 2 x 187.5 µl aliquots, eluted in 2 x 30 µl nuclease-free water, and combined in a total volume of 60 µl. cDNA quality and quantity were measured with a NanoDrop spectrophotometer.

5.4.4 qPCR

Primers were designed using Primer-BLAST (Ye, et al., 2012) to amplify small oligonucleotides of ~70 base pairs (Table C5). β -actin was used as the housekeeping gene. QPCR was performed in duplicate (n=3 biological replicates) using iQ SYBR Green SuperMix (Bio-Rad, Mississauga, Ontario, Canada) on a CFX96 real-time PCR detection system (Bio-Rad). A five-point standard curve was used to determine PCR efficiency. Relative gene expression and statistical significance were calculated using the Relative Expression Software Tool program version 2.0.13 (Pfaffl *et al.* 2002).

5.4.5 Microarray

Agilent's workflow for one-colour 4x44K gene expression arrays using FFPE samples (Agilent, 2010) was modified to accommodate FFPE samples for both one-colour (1C) and two-colour (2C) 8x60K microarrays (Supplemental Protocol, section C.4.2). cDNA (produced from archival phenobarbital exposed FFPE liver sections) was labelled using the KREAtch Universal Linkage System (ULS) genomic DNA labeling kit (Agilent Technologies). To achieve 1.5-3% labelling the suggested amount of ULS-Cy3 was increased by a factor of 1.5, and ULS-Cy5 by a factor of 1.75 above the original manufacturer's suggestion (Supplemental Protocol, section C.4.2, Table P2). Non-reacted Cy-ULS was removed using KREApure columns (Agilent Technologies); we observed considerable cDNA loss through this procedure. Therefore, to obtain a sufficient amount (μg) of labelled cDNA for hybridization, we increased the amount of cDNA in the labeling reaction such that it exceeded the final amount of ULS-cDNA needed for hybridization (Supplemental Protocol, section C.4.2, Table P2). Concentration and degree of labeling were assessed using a NanoDrop spectrophotometer. Degree of labeling was calculated according to the following formula: $\text{degree of labelling} = [(340 \times \text{pmol per } \mu\text{L dye}) / (1000 \times \text{ng per } \mu\text{L cDNA})] \times 100\%$.

Hybridization mixtures were prepared according to the manufacturer's protocol (Agilent, 2010) using the Hi-RPM Gene Expression Hybridization kit (Agilent Technologies); however, component volumes were decreased to suit the 8x60K arrays (Supplemental Protocol, section C.4.2, Tables P2-P3). The addition of the Agilent-CGHBlock blocking buffer was required to minimize background signal on the array. We compared the effect of increasing the amount of labelled cDNA that was hybridized to

the array by loading normal (1X) and increased (1.5X) amounts of cDNA (Fig. 5.1; Supplemental Protocol, section C.4.2, Table P2). For two-colour arrays, a dye-swap array design was used. Samples were hybridized to Agilent SurePrint mouse gene expression 8x60K arrays for 17 hours at 65°C and 20 rpm in the Agilent SureHyb hybridization chamber. Arrays were washed and scanned according to the manufacturer's protocol. Feature extraction was accomplished using Agilent feature extraction software version 11 (one colour protocol: GE1_1100_Jul11, or two-colour protocol: GE2_1100_Jul11). The complete dataset is available through the GEO, accession number GSE40990.

5.4.6 Microarray Statistical Analysis

Non-background subtracted raw data were normalized using LOWESS (Bolstad *et al.* 2003) in R (R-Core-Development-Team 2012) for one-colour arrays. Two-colour arrays were normalized using locally weighted scatterplot smoothing (Yang *et al.* 2002). Probes with technical replicates were averaged using the median. Sample technical replicates in the dye swaps were averaged using the arithmetic mean. Probes were flagged to be in the background if the signal intensities were less than 3 trimmed standard deviations above the trimmed mean of the structural negative control 3xSLv1 background probes (these are probes which form a hairpin and therefore do not hybridize with labeled samples).

DEGs were determined using the MAANOVA library (Wu *et al.* 2003) in R. The statistical model included fixed effects of array and treatment condition and was applied to the log₂ of the absolute intensities. The F_s statistic (Cui *et al.* 2005) was used to test for treatment effects. The p-values for all statistical tests were estimated by the

permutation method using residual shuffling, followed by adjustment for multiple comparisons using the false discovery rate approach (Benjamini & Hochberg 1995). The fold change calculations were based on the least-square means (Goodnight & Harvey 1978, Searle *et al.* 1980). A gene was considered to be differentially expressed if its absolute fold change relative to control was ≥ 1.5 with a FDR $P \leq 0.05$.

Venn diagrams were produced using Venny (bioinfogp.cnb.csic.es/tools/venny/). Pathway analysis was conducted in IPA.

5.5 Results

We compared the RNA quality (integrity and purity) and quantity (μg) extracted from FFPE tissues using a total of six commercially available kits (Table 5.1). Five kits were used to extract RNA from freshly-embedded (FFPE-preserved for less than one month) and freshly sectioned (tissue sectioning was performed the morning of RNA extraction) tissues: RNeasy FFPE kit (Qiagen), Absolutely RNA FFPE kit (Agilent), RecoverAll Total Nucleic Acids Isolation kit (Ambion), HighPure FFPE RNA Micro kit (Roche), and NucleoSpin FFPE RNA kit (Macherey-Nagel). The highest RINs were obtained from the Qiagen (6.65), Agilent (6.4), and Roche (6.85) kits, and the highest yields from the Ambion (4 μg) and Qiagen (3.4 μg) kits. RNA from corresponding freshly embedded slide-mounted samples was evaluated for four kits: RNeasy FFPE kit (Qiagen), Absolutely RNA FFPE kit (Agilent), RecoverAll Total Nucleic Acids Isolation kit (Ambion), and HighPure FFPE RNA Micro kit (Roche). The highest RINs were again obtained using the Agilent (7.00), Qiagen (6.50), and Roche (6.18) kits; however, only the Qiagen kit maintained a high yield (4 μg). Four kits were used to extract RNA from

slide-mounted archival FFPE samples (FFPE-preserved ~16 years ago): RNeasy FFPE kit (Qiagen), Absolutely RNA FFPE kit (Agilent), RecoverAll Total Nucleic Acids Isolation kit (Ambion), and Arcturus Paradise Plus FFPE RNA Isolation kit (Arcturus). We observed decreased RIN values (~2.5) for all kits, but the Qiagen kit consistently produced the highest yield (4.67 – 5.89 μ g) and was the only kit that maintained an acceptable A260/230 ratio. Taken together, the Qiagen kit performed best across samples.

We validated phenobarbital responsive positive control cytochrome P450 genes (Cyp2b9 and Cyp2b10) by qPCR and found a robust increase in their expression in our FFPE phenobarbital treated samples (Table 5.2). We then compared four protocols for hybridizing FFPE-derived cDNA onto Agilent SurePrint mouse 8x60K gene expression arrays: one-colour with normal (1C 1X) and increased (1C 1.5X) amounts of cDNA, and corresponding two-colour experiments (2C 1X and 2C 1.5X) (Fig.5.1A). Mice treated with phenobarbital were compared to control mice to identify treatment-related changes in transcript levels. There was a large amount of variation in the results obtained from the protocols including: (a) the number of DEGs (Fig. 5.1B); (b) the percent of down-regulated genes (varied from 1%~40%) (Fig. 5.1C); and (c) fold-changes (larger in one-colour experiments) (Fig. 5.1D). There were only 22 DEGs identified in common across all of the protocols (Fig. 5.2A). The up-regulation of Cyp2b10 was detected using all of the protocols; however, Cyp2b9 up-regulation was only detected using the 1C 1X protocol (Table 5.2). Both experiments with 1.5X cDNA input (1C 1.5X and 2C 1.5X) had high backgrounds and failed to meet all of the Agilent quality control metrics (Tables C6-C7).

We validated perturbations in gene expression in our study by mining the published phenobarbital literature and the Comparative Toxicogenomics Database for phenobarbital studies (ctdbase.org). Of the 504 genes that were identified as perturbed by phenobarbital using any of our four protocols, 102 were supported by evidence in the phenobarbital literature, and 150 were noted as affected by phenobarbital in the Comparative Toxicogenomics Database (Table 5.2). The 1C 1X protocol had the highest rate of validation (54%), followed by the 2C 1X (42%), 1C 1.5X (40%), and 2C 1.5X (30%); 36% of the DEGs in all four protocols could be validated (Fig. 5.3). In general, using qPCR, genes appearing using any one array protocol were very difficult to validate compared to genes appearing in multiple or all protocols. For example, we validated 7 genes (gene symbols: *Ces6*, *Cyp2b10*, *Cyp2c55*, *Cyp3a11*, *Gstm2*, *Gstm3*, and *Por*) by qPCR that were identified as differentially expressed using all four protocols (100% validation rate). In contrast, of 11 genes (gene symbols: *Cyp2b9*, *Gucy2e*, *Esco2*, *Mtap1a*, *Ttc17*, *Setd1a*, *Cela2a*, *Ctrl*, *Sgta*, *Ppp3cb*, and *Hunk*) identified as differentially expressed by one protocol only, 2 (18%) were validated (Table 5.2).

Differences in the gene lists did not translate to dissimilarities in the pathways identified as perturbed for the various protocols. Pathway perturbation is identified by determining whether there is statistically significant over-representation of genes within particular molecular pathways relative to all of the genes on the microarray. In total, nine of the top ten enriched canonical pathways were consistent across all protocols (Table 5.3; Fig. 5.2B), as well as in the intersection of DEGs using our two best protocols (1C 1X and 2C 1X) (Fig. 5.2C). This suggests that pathway analysis is a more robust and reproducible approach for analyzing FFPE microarray data than gene-by-gene analyses.

In addition, although one-colour experiments tend to identify more DEGs and larger fold-changes (Fig. 5.1), two-colour experiments are equally efficacious in detecting pathway perturbations.

5.6 Discussion

In previous gene expression studies using FFPE tissues, DEGs were validated by analysis of gene expression changes in paired FRO tissues. This experimental approach requires new animal exposures and freshly embedded organs to generate paired tissue sets. We have shown that RNA quality for archival and freshly embedded FFPE samples is not equivalent (Table 5.1). Therefore, these experiments provide insufficient validation for use of archival FFPE tissues. In the present study we used bona fide 16-year-old archival FFPE tissue with no FRO ‘answer key’.

Principal differences between FFPE RNA extraction kits include variations in deparaffinization chemicals, incubation times, and the number of ethanol washes associated with the deparaffinization step. The protocols for on-slide deparaffinization are also variable. We found it best to deparaffinize with xylene according to Qiagen’s protocol for freshly sectioned tissues, and Agilent’s protocol for slide-mounted tissues. To avoid using xylene, some kits use proprietary, non-toxic deparaffinizing agents that subdivide nucleic acids and proteins in aqueous and organic phases. These agents are impractical when dealing with slide-mounted samples unless the dry paraffin-embedded section is scraped into a tube before deparaffinization. The temperature and duration of proteinase K incubation also varies between the kit protocols. Ultimately, we found that the quality and quantity of RNA extracted using the Qiagen kit was comparable to (or

exceeded) that of other kits in our experiment, and the shorter proteinase K incubation reduced the overall time needed for extraction.

Our experiments suggest that the best way to proceed in the analysis of RNA from archival FFPE tissues is to perform two technical replicates of the entire experiment. If the number of genes appearing on only one gene list is manageable, each gene should be validated using qPCR. Genes that cannot be qPCR validated should be classified as false positives. For example, in our experiment *Cyp2b9* and *Setd1a* were only identified as differentially expressed in one of the four protocols. However, both of these genes were validated using qPCR and are therefore likely to be true positives. In contrast, changes in *Cela2a*, *Ctrl*, *Sgta*, *Ppp3cb*, *Hunk*, *Gucy2e*, *Esco2*, *Mtap1a*, and *Ttc17* were only detected using one protocol, and could not be validated by qPCR. This suggests that changes in the expression of these genes were likely to be false positives (Table 5.2). An alternative approach is to assume that DEGs that do not appear in at least two separate experiments are false positives and should be removed from further analyses. Although the latter may lead to a higher false negative rate for individual genes, pathway analysis of the genes in common between our two best protocols demonstrates that the intersection shares nine of the top ten canonical pathways (Fig. 5.2C). Thus, the total number of DEGs may be lower using this approach, but the relevant pathways are identified nonetheless.

The top canonical pathways enriched among the genes identified as differentially expressed by phenobarbital treatment are consistent with its characterized mode of toxicological action. Phenobarbital is a rodent non-genotoxic hepatocarcinogen that has been effectively used in humans as a sedative and anti-epileptic drug. It is a constitutive androstane receptor (CAR) and pregnane X receptor (PXR) ligand (Ross *et al.* 2010,

Ihunnah *et al.* 2011). CAR and PXR are ligand-activated nuclear receptors that heterodimerize with retinoid X receptor (RXR) to activate gene expression of xenobiotic metabolism enzymes, including the CYP2Bs (Honkakoski *et al.* 1998, Xu *et al.* 2005) and CYP3As (Xu *et al.* 2005, Tien & Negishi 2006), respectively. In addition, CAR and PXR regulate the expression of conjugation enzymes (Tien & Negishi 2006). Since IPA canonical pathways are inherently redundant, our list of nine pathways can be resolved to explain the phenobarbital MoA. The LPS/IL-1 (lipopolysaccharide/interleukin 1) mediated inhibition of RXR function pathway is a very large pathway that includes the Xenobiotic Metabolism Signaling pathway. The latter comprises the CAR/RXR, PXR/RXR and Aryl Hydrocarbon Receptor signaling pathways, which are all upstream of the Metabolism of Xenobiotics by Cytochrome P450 and Glutathione Metabolism pathways. The Fatty Acid, Linoleic Acid, Arachidonic Acid and Tryptophan Metabolism pathways are enriched as a secondary effect because they contain endogenous cytochrome P450 substrates. Our pathway analysis of phenobarbital exposed archival FFPE tissue confirms that CAR/RXR and PXR/RXR mediated expression of xenobiotic metabolism enzymes are key events in the molecular response to phenobarbital. These data are also consistent with the evidence that phenobarbital is a non-genotoxic carcinogen as we see no enrichment for pathways such as DNA damage response or p53 signaling.

Finally, it is noteworthy that the list of DEGs generated by each of the four protocols contained lincRNAs (long intergenic non-coding RNAs); the 1C 1X, 1C 1.5X, 2C 1X and 2C 1.5X protocols detected 38, 99, 58, and 28 lincRNAs, respectively. LincRNAs are emerging as potential biomarkers of cancer (Gutschner & Diederichs

2012), and toxicity (Recio *et al.* 2013) and Appendix A; therefore, it is significant that they are detectable in FFPE samples. To our knowledge, this is the first report of lincRNAs being detected in either FFPE samples or toxicant-exposed samples. However, we note that the changes in expression detected for lincRNAs were highly variable between the protocols (thus, the lincRNA response remains to be validated).

In summary, we provide a step-by-step protocol for the analysis of gene expression from archival FFPE tissues on Agilent 8x60K microarrays (Figure 5.4; Supplemental Protocol, section C.4.2). We demonstrate that 16-year-old FFPE liver samples can be used to detect known transcript and pathway responses to phenobarbital. We recommend that, when no paired FRO tissues are available, experiments be performed in duplicate and the intersection of the gene lists be used to minimize false positives. Finally, we show that pathway analysis represents a more robust means for identifying trends in FFPE gene expression compared to a gene-by-gene analysis. We anticipate that gene expression studies using FFPE-preserved tissues will become an extremely important part of disease and toxicant MoA research as these tissues have the added benefit of having well documented apical end-point data associated with them.

5.7 Tables and Figures

Table 5.1 RNA extraction from FFPE tissues. Comparison across six commercial kits of quantity and quality of RNA extracted from freshly-embedded or archival and freshly sectioned or slide-mounted mouse liver samples (SD=standard deviation).

Commercial kit	Freshly formalin-fixed paraffin embedded (FFPE)								Archival formalin fixed paraffin embedded (FFPE)							
	Sections (n = 3)				Slides (n = 3)				Unexposed (n = 3)				Phenobarbital exposed (n = 3)			
	RIN	260/ 280	260/ 230	Yield (µg)	RIN	260/ 280	260/ 230	Yield (µg)	RIN	260/ 280	260/ 230	Yield (µg)	RIN	260/ 280	260/ 230	Yield (µg)
RNeasy mini kit (Qiagen)	8.32	2.12	2.1	9.59												
<i>SD</i>	<i>0.36</i>	<i>0.02</i>	<i>0.16</i>	<i>3.31</i>												
RNeasy FFPE kit (Qiagen)	6.65	1.97	1.6	3.4	6.5	1.94	1.44	4.04	2.52	1.95	1.77	4.67	2.52	1.97	1.95	5.89
<i>SD</i>	<i>0.3</i>	<i>0.01</i>	<i>0.08</i>	<i>0.11</i>	<i>0.35</i>	<i>0.08</i>	<i>0.18</i>	<i>1.99</i>	<i>0.03</i>	<i>0.03</i>	<i>0.08</i>	<i>0.93</i>	<i>0.03</i>	<i>0.04</i>	<i>0.1</i>	<i>2.38</i>
Absolutely RNA FFPE kit (Agilent)	6.4	1.9	0.94	1.33	7	2.03	1.19	1.47	2.48	1.95	1.24	1.46	2.55	1.98	1.03	2.36
<i>SD</i>	<i>0.71</i>	<i>0.09</i>	<i>0.13</i>	<i>0.55</i>	<i>0.09</i>	<i>0.01</i>	<i>0.12</i>	<i>0.31</i>	<i>0.03</i>	<i>0.06</i>	<i>0.06</i>	<i>0.19</i>	<i>0.05</i>	<i>0.01</i>	<i>0.68</i>	<i>0.32</i>
RecoverAll total nucleic acid isolation kit (Ambion)	5.05	2.28	3.86	4	5.28	2.04	1.14	2.09	2.43	2.07	0.29	2.53	2.5	2.06	0.53	2.77
<i>SD</i>	<i>0.61</i>	<i>0.36</i>	<i>2.72</i>	<i>5.13</i>	<i>0.35</i>	<i>0.09</i>	<i>0.74</i>	<i>0.52</i>	<i>0.03</i>	<i>0.01</i>	<i>0.07</i>	<i>0.16</i>	<i>0</i>	<i>0.02</i>	<i>0.28</i>	<i>0.6</i>
Arcturus Paradise Plus FFPE RNA Isolation kit (Arcturus)									2.5	1.96	1.6	2.97	2.47	1.98	1.46	5.37
<i>SD</i>									<i>0</i>	<i>0.05</i>	<i>0.26</i>	<i>0.79</i>	<i>0.06</i>	<i>0.04</i>	<i>0.2</i>	<i>2.45</i>
HighPure FFPE RNA Micro kit (Roche)	6.85	2.01	0.39	1.48	6.18	1.88	0.26	1.79								
<i>SD</i>	<i>0.3</i>	<i>0.02</i>	<i>0.24</i>	<i>0.58</i>	<i>0.6</i>	<i>0.11</i>	<i>0.05</i>	<i>0.17</i>								
NucleoSpin FFPE RNA kit (Macherey- Nagel)	2.72	2.01	1.51	1.31												
<i>SD</i>	<i>0.48</i>	<i>0.02</i>	<i>0.3</i>	<i>0.33</i>												

Table 5.2 Select qPCR and microarray gene expression values for PB-responsive genes. Genes that have been previously reported in the literature or the Comparative Toxicogenomics Database (CTD) are indicated.

Gene Symbol	Real time PCR			SurePrint G3 mouse GE 8 × 60 K microarray (Agilent)								Reference	CTD?
	FC	P	Result	1C, 1 x		1C, 1.5 x		2C, 1 x		2C, 1.5x			
				FC	P	FC	P	FC	P	FC	P		
<i>Ces6</i>	3.35	0.014	Up	2.57	P < 0.0001	2.1	P < 0.0001	1.96	P < 0.0001	1.54	P < 0.0001	(Ross <i>et al.</i> 2009) (Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Ross <i>et al.</i> 2009, Tamasi <i>et al.</i> 2009, Tamasi <i>et al.</i> 2009, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011)	N
<i>Cyp2b10</i>	361.5	P < 0.0001	Up	9.78	P < 0.0001	5.87	P < 0.0001	2.72	P < 0.0001	1.61	P < 0.0001	(Konno <i>et al.</i> 2010, Ross <i>et al.</i> 2010, Ross <i>et al.</i> 2010, Ross <i>et al.</i> 2010)	Y
<i>Cyp2c55</i>	10.72	P < 0.0001	Up	8.18	P < 0.0001	7.44	P < 0.0001	4.98	P < 0.0001	1.74	P < 0.0001	(Ross <i>et al.</i> 2009, Ross <i>et al.</i> 2009, Ross <i>et al.</i> 2010)	Y
<i>Cyp3a11</i>	1.86	0.07	Up	2.3	0.012	1.99	P < 0.0001	1.98	P < 0.0001	1.57	P < 0.0001	(Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Ross <i>et al.</i> 2010, Zheng <i>et al.</i>	Y
<i>Gstm2</i>	2.9	P < 0.0001	Up	2.24	0.025	1.95	0.011	1.83	P < 0.0001	1.52	P < 0.0001		Y

												2011, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011)	
<i>Gstm3</i>	2.52	0.03	Up	8.18	P < 0.0001	8.42	P < 0.0001	6.36	P < 0.0001	2.22	P < 0.0001	(Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Ross <i>et al.</i> 2010, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011)	Y
<i>Por</i>	5.42	P < 0.0001	Up	5.51	P < 0.0001	4.8	P < 0.0001	3.19	P < 0.0001	1.53	P < 0.0001	(Stahl <i>et al.</i> 2005, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011)	Y
<i>Cyp2b9</i>	157.5	P < 0.0001	Up	2.09	P < 0.0001							(Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Stahl <i>et al.</i> 2005, Ross <i>et al.</i> 2010, Zheng <i>et al.</i> 2011, Zheng <i>et al.</i> 2011)	Y
<i>Gucy2e</i>			DA	12.62	0.04								N
<i>Esco2</i>			DA	2.18	0.03								N
<i>Mtap1a</i>			DA	17.35	0.037								N
<i>Ttc17</i>	1.10	0.101	NC	14.27	0.034								N
<i>Setd1a</i>	1.5	0.037	Up					1.64	0.0115				N
<i>Cela2a</i>	1.02	0.513	NC					2.04	P < 0.0001				N
<i>Ctrl</i>	1.1	0.099	NC					–	P < 0.0001				N

					2.04	
<i>Sgta</i>	1.07	0.464	NC	2.08	0.042	N
<i>Ppp3cb</i>			DA	6.07	P < 0.0001	N
<i>Hunk</i>			DA	10.52	P < 0.0001	N

Bold text indicates that p-values are significant ($p < 0.05$).

FC = Fold change; NC = no change; DA = Didn't amplify; CTD = Comparative Toxicogenomics Database; Y = yes; N = no

Table 5.3 Top 10 enriched molecular pathways. Pathways are ordered alphabetically; the gene cut-offs are FC > ±1.5 and p < 0.05.

C, 1 ×; 1C, 1.5 ×; 2C, 1.5 ×		2C, 1 ×		Pathways found using genes present in both 1C, 1 × and 2C, 1 × protocols	
Pathway	P	Pathway	P	Pathway	P
Arachidonic acid metabolism	P < 0.0001	Acute phase response signaling	P < 0.001	Arachidonic acid metabolism	P < 0.0001
Aryl hydrocarbon receptor signaling	P < 0.0001	Arachidonic acid metabolism	P < 0.0001	Fatty acid metabolism	P < 0.0001
Fatty acid metabolism	P < 0.0001	Fatty acid metabolism	P < 0.0001	Glutathione Metabolism	P < 0.0001
Glutathione metabolism	P < 0.0001	Glutathione metabolism	P < 0.0001	Linoleic acid metabolism	P < 0.0001
Linoleic acid metabolism	P < 0.0001	Linoleic acid metabolism	P < 0.0001	LPS/IL-1 mediated inhibition of RXR function	P < 0.0001
LPS/IL-1 mediated inhibition of RXR function	P < 0.0001	LPS/IL-1 mediated inhibition of RXR function	P < 0.0001	Metabolism of xenobiotics by cytochrome P450	P < 0.0001
Metabolism of xenobiotics by cytochrome P450	P < 0.0001	Metabolism of xenobiotics by cytochrome P450	P < 0.0001	NRF2-mediated oxidative stress response	P < 0.0001
PXR/RXR activation	P < 0.0001	PXR/RXR activation	P < 0.0001	PXR/RXR activation	P < 0.0001
Tryptophan metabolism	P < 0.0001	Tryptophan metabolism	P < 0.0001	Tryptophan metabolism	P < 0.0001
Xenobiotic metabolism signaling	P < 0.0001	Xenobiotic metabolism signaling	P < 0.0001	Xenobiotic metabolism signaling	P < 0.0001

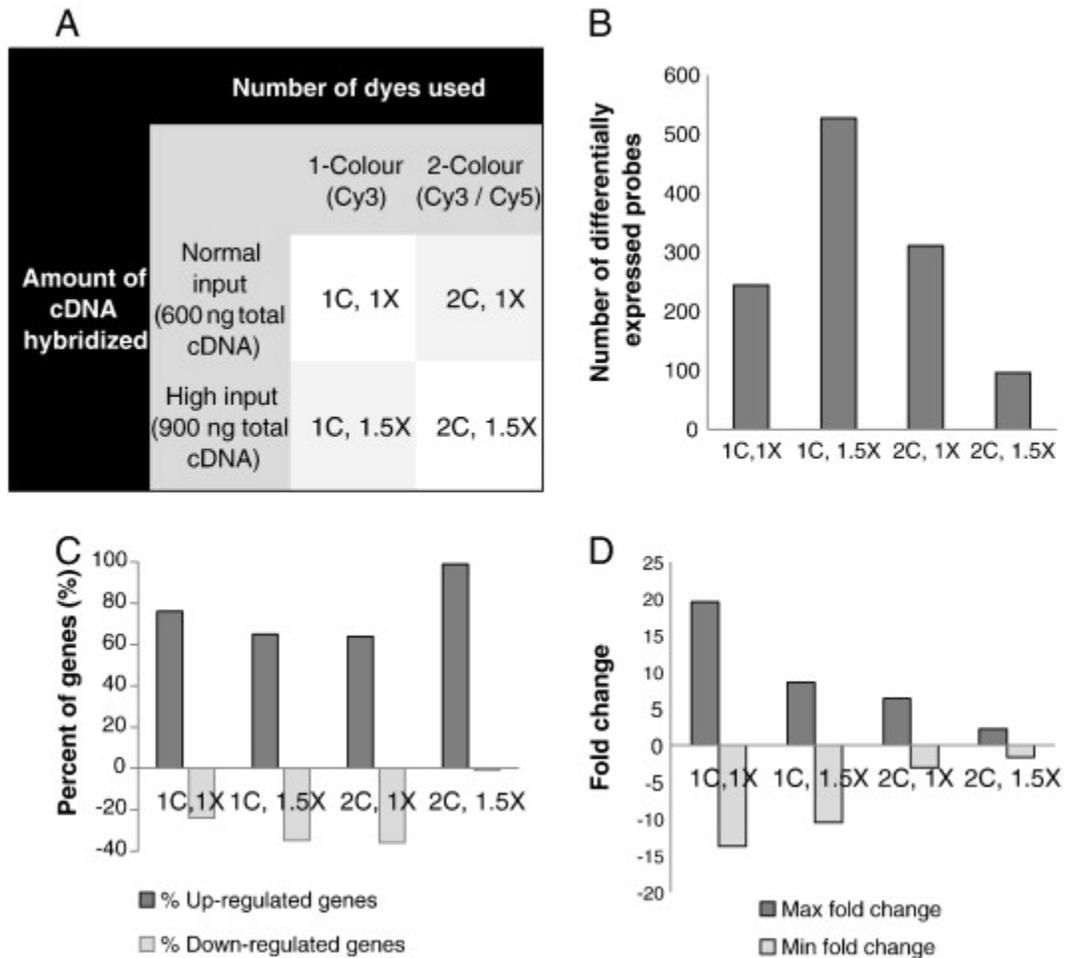


Figure 5.1 Qualitative descriptions of the four differentially expressed probe lists generated by each microarray protocol. (A) Experimental design for four array protocols. (B) The number of significant ($P \leq 0.05$) probes with a fold change of greater than ± 1.5 . (C) The percentage of differentially expressed probes in each list that is up- or down-regulated. (D) The range of intensity of fold-change represented by the maximum and minimum changing probes for each data set.

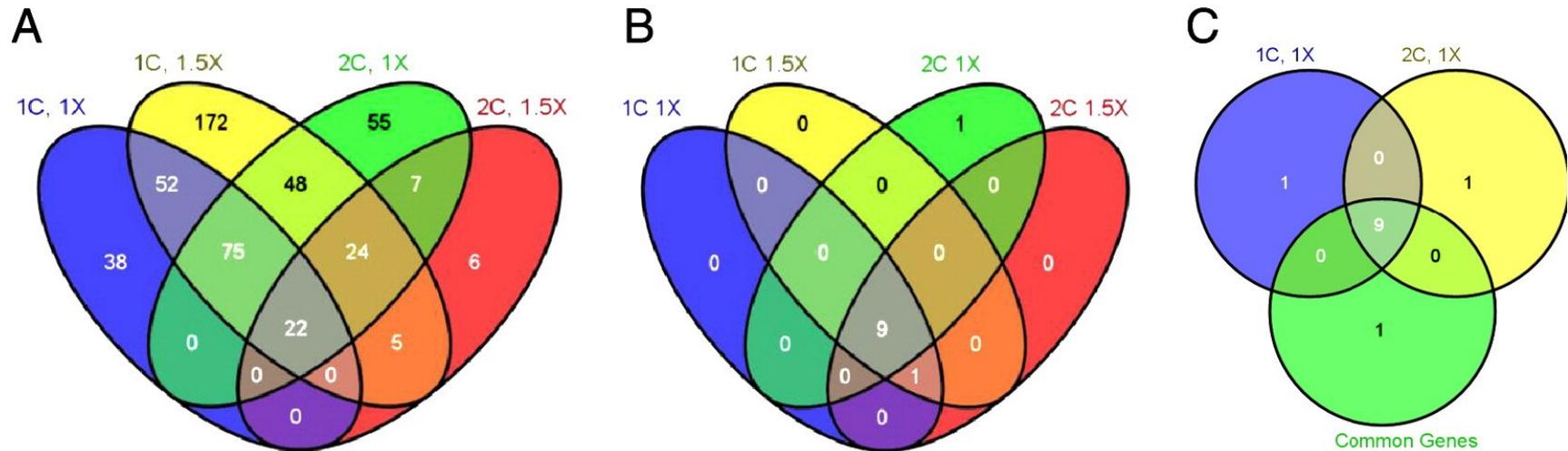


Figure 5.2 Comparison of the DEGs identified by each microarray protocol. Venn diagrams representing the overlaps of (A) DEGs from each protocol, (B) top 10 IPA canonical pathways enriched for within each gene list, and (C) top 10 canonical pathways enriched in the gene lists from 1C 1 ×, 2C 1 ×, and the genes common to both 1C 1 × and 2C 1 ×.

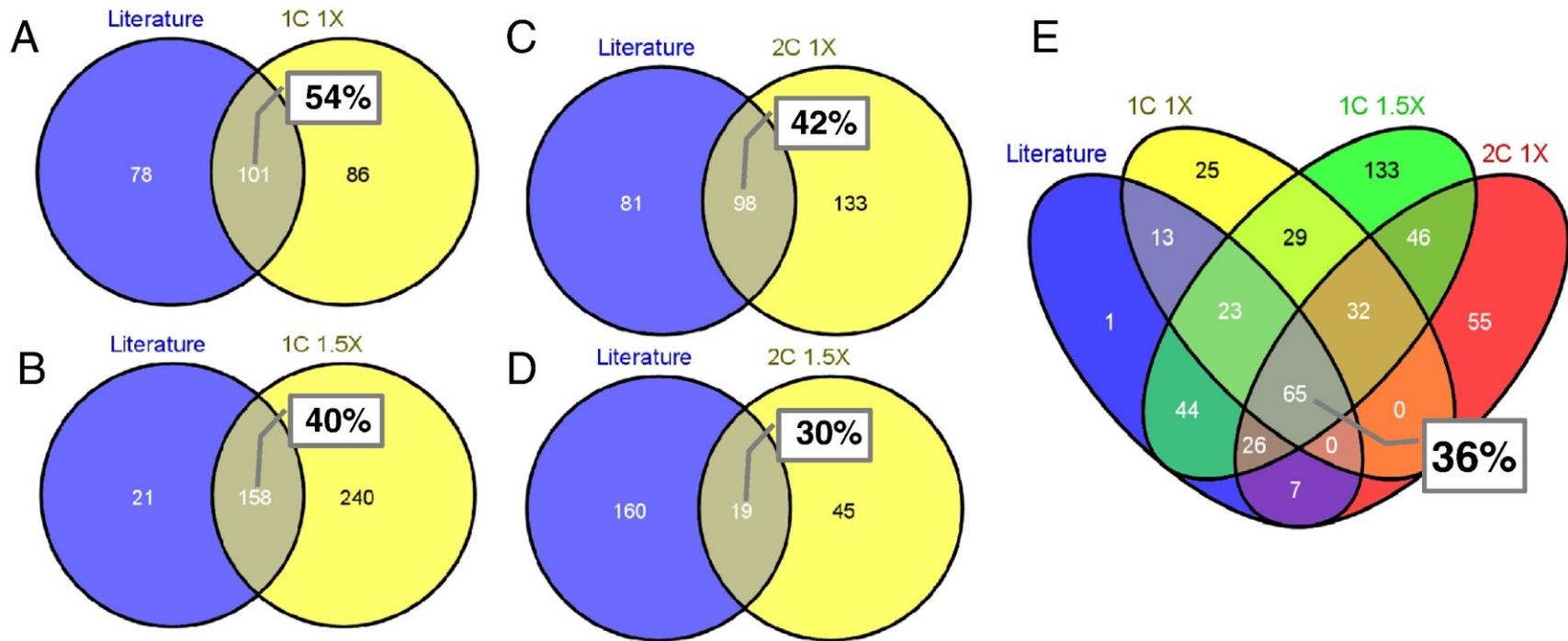


Figure 5.3 Literature validation of each protocol. (A) 54% of DEGs appearing in the 1C 1 × protocol were also found in published phenobarbital studies, as were (B) 40% of the DEGs appearing in the 1C 1.5 × protocol, (C) 42% of the DEGs appearing in the 2C 1 × protocol, and (D) 30% of the DEGs appearing in the 2C 1.5 × protocol. (E) 36% of the 179 phenobarbital responsive genes reported in the literature also appeared in our top three protocols.

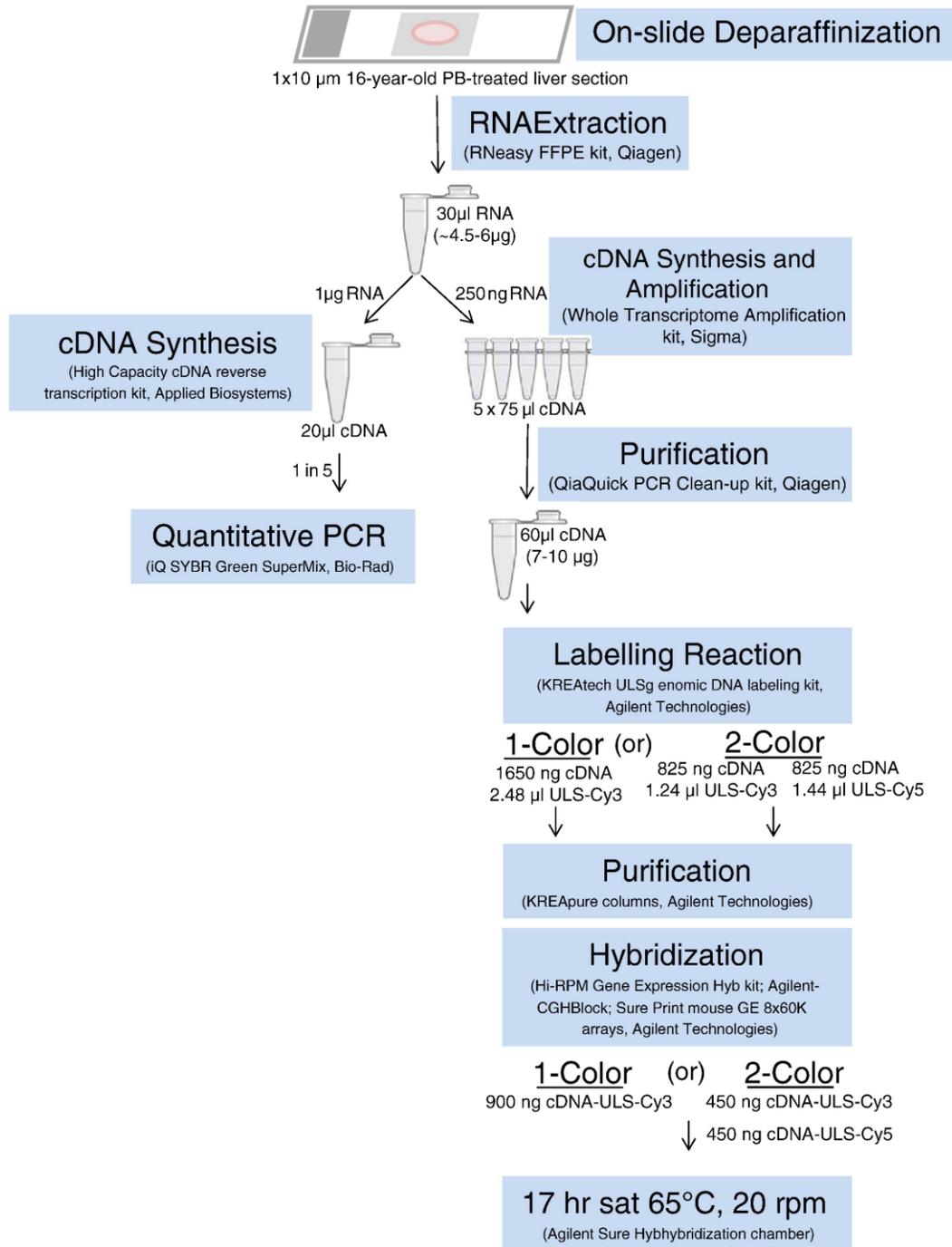


Figure 5.4 Schematic of experimental steps taken for analysis of gene expression in archival, slide-mounted FFPE tissue samples.

CHAPTER 6

6 CHAPTER: MINING THE ARCHIVES: A CROSS-PLATFORM ANALYSIS OF GENE EXPRESSION PROFILES IN ARCHIVAL FORMALIN-FIXED PARAFFIN-EMBEDDED TISSUE

Modified from: A.F. Webster, Zumbo, P., Fostel, J., Gandara, J., Hester, S.D., Recio, L., Williams, A., Wood, C.E., Yauk, C.L., Mason, C.E. *Toxicol. Sci.* (ePub Sept. 15, 2015)
doi: 10.1093/toxsci/kfv195

6.1 Preamble

Chapter 6 brings together the furan MoA (Chapter 2) and the FFPE gene expression protocol (Chapter 5), and tests two additional RNA-seq protocols for analysis of FFPE tissues. Gene expression profiles from paired FRO and FFPE furan-treated liver samples are compared. We also look at two additional factors that affect the quality of RNA extracted from FFPE tissues: time spent in formalin, and age of paraffin block.

6.2 Abstract

FFPE tissue samples represent a potentially invaluable resource for transcriptomic-based research into the molecular basis of disease. However, use of FFPE samples in gene expression studies has been limited by technical challenges resulting from degradation of nucleic acids. Here we evaluated gene expression profiles derived from FRO and FFPE mouse liver tissues using two DNA microarray protocols and two whole transcriptome sequencing (RNA-seq) library preparation methodologies. The ribo-depletion protocol outperformed the other three methods by having the highest correlations of DEGs and best overlap of pathways between FRO and FFPE groups. We next tested the effect of sample time in formalin (18 hours or 3 weeks) on gene expression profiles. Hierarchical clustering of the datasets indicated that test article treatment, and not preservation method, was the main driver of gene expression profiles. Meta- and pathway analyses indicated that biological responses were generally consistent for 18-hour and 3-week FFPE samples compared to FRO samples. However, clear erosion of signal intensity with time in formalin was evident, and DEG numbers differed by platform and preservation method. Lastly, we investigated the effect of age in FFPE block on genomic profiles. RNA-seq analysis of 8-, 19-, and 26-year-old control blocks using the ribo-depletion protocol resulted in comparable quality metrics, including expected distributions of mapped reads to exonic, UTR, intronic, and ribosomal fractions of the transcriptome. Overall, our results suggest that FFPE samples are appropriate for use in genomic studies in which frozen samples are not available, and that ribo-depletion RNA-seq is the preferred method for this type of analysis in archival and long-aged FFPE samples.

6.3 Introduction

Archival tissues housed in tissues libraries (such as the NTP histopathology archival library) represent a unique and exciting opportunity for conducting retrospective, and phenotypically anchored toxicogenomic analyses. Recently, whole-transcriptome RNA sequencing (RNA-seq) methodologies have been developed as a more precise and less biased measurement of transcript levels that may overcome issues associated with highly fragmented or degraded RNA. Compared to conventional DNA microarrays, RNA-seq enriches for many additional fragments as it is not restricted to predefined probes and has (in principle) no limitations to dynamic range (SEQC/MAQC-III Consortium 2014, Li *et al.* 2014b). Previous work has shown concordance between gene expression profiles produced by microarrays and by RNA-seq for FRO and FFPE samples (Marioni *et al.* 2008, Malone & Oliver 2011, Wang *et al.* 2014). Recent studies investigating the use of RNA-seq in FFPE samples have also shown promising results (Hedegaard *et al.* 2014, Auerbach *et al.* 2014, Zhao *et al.* 2014b). To extend these findings, the HESI Committee on Application of Genomics to Mechanism-Based Risk Assessment coordinated a working group to assess the utility of various gene expression profiling methods for FFPE tissues. The main goals of this group were to evaluate current RNA-seq and microarray methods in paired FRO and FFPE samples, and investigate different preanalytical and experimental factors influencing FFPE results.

In order to evaluate genomics methods for FFPE tissues, we compared gene expression profiles from paired FRO and FFPE-preserved liver tissues from female B6C3F1 mice that had been exposed for three weeks to a carcinogenic dose of furan (8 mkd). Gene expression profiling was completed using four protocols: Illumina Hi-seq

RNA-seq (ribo-depletion and polyA-enrichment formats) and Agilent SurePrint G3 Mouse GE 8x60K microarrays (one- and two-colour formats). Next, we investigated the effect of tissue time-in-formalin on the quality of transcriptomic data given that different institutions have different time-in-formalin protocols. For this aim, we used furan-exposed paired mouse liver samples that had been preserved in formalin for either 18 hours or 3 weeks. Finally, to better understand the effects of time-in-paraffin, we used the RNA-seq ribo-depletion protocol to evaluate rat liver samples that were obtained from the National Toxicology Program (NTP) archives and had been in storage for up to 26 years. Our results provide a comprehensive comparison of genomics methodologies for FFPE samples, which allow us to make recommendations for the most appropriate approach for the study of archival specimens.

6.4 Methods

6.4.1 Frozen and FFPE Tissue Samples

FRO and FFPE samples for the microarray to RNA-seq comparison, and time-in-formalin analysis were obtained from a recent study of female mice treated with furan (CAS No. 110-00-9) (> 99% pure) (Sigma-Aldrich Chemical Co., Milwaukee, WI). Doses of 0 or 8 mkg furan dissolved in Mazola corn oil were prepared as described in Chapter 2. Mice were housed, dosed, and sacrificed as described in Chapter 2. Livers were removed and trimmed, and portions were either flash-frozen or fixed in a standard solution of neutral phosphate-buffered formalin. Frozen samples were stored at or below -70°C. Fixed samples were kept in formalin for 18 hours or 3 weeks, and then transferred to 100% ethanol. After ~72 hours, tissues were processed and then embedded in paraffin

using standard histologic procedures. FFPE tissue blocks were stored at room temperature for less than one year before RNA was isolated for this study.

For the time-in-paraffin analysis, FFPE blocks of control liver tissue were obtained from the NTP archive collection. The corresponding slides from each block were re-evaluated on histopathology to confirm that the selected tissue was morphologically normal. Samples were obtained from three different studies in control F344 rats in which tissues had been stored as FFPE blocks for different lengths of time: (1) 8 years (13-week controls for beta-picoline exposure); (2) 19 years (two-year controls for oxazepam exposure); and (3) 26 years (13-week controls for trichlorfon exposure). The difference in study length affects both the age of the animal at sacrifice and also the allowable time in formalin. According to NTP study protocol, for a 13-week study the tissues may only stay in formalin for up to 6 weeks, while for a two-year study fixed tissues are permitted to remain in formalin for up to 6 months. Recently-collected rat liver was used as FRO control.

6.4.2 RNA Extraction and cDNA Synthesis

Total RNA was extracted from FRO tissues using the RNeasy Midi RNA Extraction kit (Qiagen, Mississauga, ON, Canada) as described in Chapter 2.

Total RNA was extracted from slide-mounted furan and archival FFPE tissues using the FFPE RNeasy kit (Qiagen). One 10 um section was used per FFPE extraction per animal for the 18 hours-in-formalin group. Due to low yields of RNA, two 10 um sections were used per FFPE extraction per animal for the 3 weeks-in-formalin group. RNA extracted from FFPE tissues was assessed for integrity using an Agilent 2100

Bioanalyzer (Agilent Technologies Inc., Mississauga, ON, Canada), quantitated using a NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA), and stored at -80°C (Table C3). For cDNA synthesis, RNA (250 ng per sample) was reverse transcribed and amplified using the Whole Transcriptome Amplification kit 2 (WTA2; Sigma, St. Louis, Missouri, USA). cDNA was purified using the QiaQuick PCR clean-up kit (Qiagen).

6.4.3 Microarray Analysis

Microarray analysis for the FRO tissues was performed using one-colour and two-colour protocols (all microarray experiments were conducted in the Yauk laboratory, Ottawa, Canada).

The one-colour FRO protocol was carried out according to the One-Colour Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling Protocol, version 6.6 (Agilent Technologies, 2012). Briefly, using the RNA Spike-In Kit, One-Colour (Agilent Technologies) and the Low Input Quick Amp Labeling Kit, One-Colour (Agilent Technologies), 200 ng of RNA from each sample was used to synthesize, amplify, and Cy3-label cRNA. We labeled 4 samples each for the FRO, 18 hours in formalin FFPE, and 3 weeks in formalin FFPE (control animals 2, 3, 4, 6, and treated animals 41, 42, 43, 44); RNA from animal 3 in the FFPE-3wks group did not label so the number of biological replicates for the FFPE-3wks group was reduced to 3 (we included an additional technical replicate of sample 2 in this group). Labeled cRNA was purified using the RNeasy Mini Kit (Qiagen) and quantified using a NanoDrop Spectrophotometer. Hybridization mixes were prepared using the Gene Expression

Hybridization Kit (Agilent Technologies). Cy3-labeled cRNA (600 ng) was hybridized to SurePrint G3 Mouse GE 8 x 60 K microarrays (Agilent Technologies) at 65 °C for 17 hours at 10 rpm.

The two-colour FRO protocol was previously described in Chapter 2. Briefly, using a reference design, 200 ng of sample RNA and 200 ng of mouse universal reference RNA (Stratagene, Agilent Technologies Inc.) were labeled with Cy5 and Cy3, respectively, using the Low Input Quick Amp Labeling Kit (Agilent Technologies). Labeled cRNA was purified and quantified as described above. cRNA (300 ng per sample) and reference cRNA were hybridized to SurePrint G3 Mouse GE 8 x 60 K microarrays (Agilent Technologies) at 65 °C for 17 hours at 10 rpm.

Microarray analysis for FFPE tissues was performed according to our published one-colour and two-colour cDNA input protocols (Chapter 5). For the one-colour FFPE protocol, 1650 ng of cDNA was labeled with 2.48 µl Cy3 dye using the Genomic DNA ULS Labeling Kit (Agilent Technologies). For the two-colour FFPE protocol, 825 ng of cDNA was labeled with 1.44 µl Cy5 dye, and 825 ng of reference cDNA (made from a pool of all sample cDNA) was labeled with 1.24 µl Cy3 using the Genomic DNA ULS Labeling Kit (Agilent Technologies). In both protocols, excess dye was removed using KREApure columns (Agilent Technologies), labeled product was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc.), and degree of labeling was calculated using the following calculation: degree of labeling = $[(340 \times \text{pmol per } \mu\text{l dye}) / (\text{ng per } \mu\text{l cDNA} \times 1000)] \times 100\%$ (with an acceptable range of 1.5-3%).

Hybridization mixtures were prepared using the Gene Expression Hybridization kit (Agilent Technologies) and CGHBlock (Agilent Technologies). For the one-colour

FFPE protocol, 600 ng of Cy3-labelled cDNA was hybridized to SurePrint G3 Mouse GE 8x60K microarrays (Agilent Technologies). For the two-colour FFPE protocol, 300 ng each of sample and reference cDNA was hybridized to SurePrint G3 Mouse GE 8x60K microarrays (Agilent Technologies). Hybridization for both protocols occurred in a randomized block design, at 65°C for 17 hours at 20 rpm in the dark in an Agilent SureHyb hybridization chamber.

All slides were washed with Wash Buffers 1 and 2 (Agilent Technologies), and scanned at 5 µm resolution on an Agilent G2505B scanner. Feature extraction was accomplished using Agilent feature extraction software version 11. The complete microarray dataset is available through the GEO, accession number: GSE62843.

6.4.4 Microarray Statistical Analysis

One-Colour: Non-background subtracted median signal intensities were cyclic-LOWESS normalized (Bolstad *et al.* 2003) in R (R-Core-Development-Team 2012) using the affy library (Gautier *et al.* 2004). Differential gene expression analysis between the control and treated samples was applied using the MAANOVA library (Wu *et al.* 2003) in R. The F_s statistic (Cui *et al.* 2005) was used to test for differential expression. The p-values from these tests were estimated using the permutation method with residual shuffling, and adjusted for multiple comparisons by using the FDR approach (Benjamini & Hochberg 1995). Least square means (Goodnight & Harvey 1978, Searle *et al.* 1980) were used to estimate the fold change for each pairwise comparison tested.

Two-Colour: A reference design (Kerr 2003) was employed for the two-colour analysis. Non-background subtracted median signal intensities were LOWESS-

normalized (Yang *et al.* 2002) in R (www.R-project.org) using the MAANOVA library. The MAANOVA library was used for the DEG analysis using the log₂ of the relative intensities. The p-values were estimated using the permutation method with residual shuffling, and FDR-adjusted p-values were computed using the FDR approach. Fold change estimates were computed as in the one-colour analysis.

Correlation analysis: Correlations in Fig. 6.1A were determined using the regression function in the Microsoft Excel ‘Data Analysis’ tool pack.

6.4.5 Microarray Bioinformatic Analysis

Cluster analysis: The universal reference used in the two-colour FRO data was different than the pooled sample reference used in the FFPE two-colour analyses. For the cluster analysis, the data from the two-colour and one-colour experiments were normalized to the median of the time-matched controls. By applying this normalization the FRO and FFPE data could then be analyzed jointly. Heatmaps were constructed using a similarity metric based on 1-correlation. The correlations were estimated using the Spearman approach. Genes were considered significant with a 2-fold change and an FDR $P < 0.05$. The row dendrogram for the heatmaps was estimated by pooling the one-colour and two-colour data in order to have the same row order for the two plots to help facilitate the comparison of the two heatmaps.

Pathway analysis: IPA was used to identify enriched molecular pathways. Gene expression data was analyzed using an IPA Core Analysis with a gene expression threshold of fold change $\geq \pm 1.5$ and FDR $P \leq 0.05$. Enrichment of IPA canonical pathways was determined based on the number of DEGs in the dataset that also appear in

that pathway. Pathway significance thresholds were determined in IPA using a right-tailed Fisher's exact test. Only mapped DEGs were considered for Fig. 6.2 (unmapped DEGs in Fig. C5).

NextBio: To identify chemical effectors and disease states that produce changes to gene expression that are similar to those produced following exposure to furan, gene expression profiles from all microarray data sets were mined against a large genomic database repositories using NextBio. NextBio uses proprietary algorithms that use pairwise gene signature correlations and rank-based enrichment statistics to produce scores, where the chemical or disease with the highest similarity is given a score of 100 and the rest are normalized accordingly (Kupersmidt *et al.* 2010). The meta-analyses included up to the top 250 DEGs (mapped) from each group with an FDR $P < 0.05$.

6.4.6 RNA-Sequencing Methods

RNA sequencing was performed on the FRO and FFPE tissues using two different protocols: ribosomal RNA-depleted-sequencing (rmRNA-seq; conducted in the Mason laboratory, NYC, USA) and poly-A-selected RNA-sequencing (mRNA-seq; conducted at Genome Quebec, Montreal, Canada). The rmRNA-seq and mRNA-seq protocols were carried out according to Illumina's TruSeq Total RNA Sample Prep Kit and Illumina's TruSeq RNA Sample Preparation Kit v2, respectively. For rmRNA-seq library construction, ribosomal RNA was removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads (Human/Mouse/Rat). For mRNA-seq library construction, poly-A-containing mRNA was purified using oligo-dT-attached magnetic beads. Following purification, rRNA-depleted or poly-A-enriched RNA was

chemically fragmented into small pieces using divalent cations under elevated temperature. For all FFPE samples, we used double the input when compared to fresh-frozen. The RNA fragments were copied into first strand cDNA using reverse transcriptase and random hexamer primers, followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. Adapters were subsequently ligated to the cDNA fragments, and then purified and enriched for with PCR to create a final cDNA library. The final cDNA libraries were single-end sequenced on a HiSeq2000. Ribo-depletion experiments had a read length of 42 bases and a read depth of 10-60M reads. Poly-A-enrichment experiments had a read length of 50 bases and a read depth of 3.8-11M reads. It has previously been reported that the required sequencing depth to be microarray-equivalent is 5M (Black *et al.* 2014). The complete dataset is available through the Gene Expression Omnibus <http://www.ncbi.nlm.nih.gov/geo/> accession number: GSE62843.

6.4.7 RNA-Sequencing Bioinformatic Analysis

The RNA-sequencing data were processed using Illumina's RTA software and converted into FASTQ files using Illumina's CASAVA pipeline. The FASTQ files were aligned to either the UCSC mm9 mouse genome (for the furan study) or the rn5 rat genome (for the age-in-block study) with STAR (Dobin *et al.* 2013), a universal RNA-seq aligner, using default parameters. Sequences that mapped to more than one locus were excluded from further analysis. Gene expression values were calculated based on composite gene models of NCBI's Reference Sequence (RefSeq) gene annotation models (retrieved via UCSC's Table Browser 5/7/14) using featureCounts (Liao *et al.* 2014), a program for assigning sequence reads to genomic features. Composite gene models for

each gene consisted of the union of the exons from the set of all transcript isoforms from that gene. Expression values were also calculated for a subset of RefSeq genes, which corresponded to the intersection between the RefSeq gene models and the Agilent SurePrint G3 Mouse Microarray targets. The intersection between the RefSeq gene models and the Agilent SurePrint G3 Mouse Microarray targets was determined as follows. Microarray probe sequences corresponding to Agilent SurePrint G3 Mouse Microarrays were downloaded from NCBI (accession number: GPL13912). The probe sequences were aligned to the mouse genome using STAR. Probe sequences that mapped to more than one locus (5.65%) were excluded from downstream consideration. Probe sequences that mapped to just one locus were intersected with composite gene models from NCBI's RefSeq database using featureCounts. Probe sequences that unambiguously overlapped with no more than one RefSeq composite gene model (53.9%) were paired with that gene model; the remaining probe sequences were discarded. The pairs were collated into a unique, non-redundant list of RefSeq genes, which was used in subsequent differential gene expression analyses. All sets of gene counts were normalized using voom (Law *et al.* 2014), which performs a LOWESS regression to estimate the mean-variance relation and transforms the gene counts into the appropriate log form for linear modeling. Differential gene expression analysis was then performed for each set of transformed gene counts using limma (Ritchie *et al.* 2015) for the following set of contrasts: FRO Control versus FRO High Dose, FFPE-18hr Control versus FFPE-18hr High Dose, and FFPE-3wk Control versus FFPE-3wk High Dose.

6.4.8 Online Data Access

The complete dataset is available through the GEO. The superseries accession number is GSE62843.

6.5 Results

6.5.1 Gene Expression Profiles

Our initial aim was to evaluate transcriptomic profiles obtained using different protocols. The mean RNA integrity numbers (RIN) for the FRO, FFPE-18hr, and FFPE-3wk samples were 9.3 (range: 8.9-9.5), 2.1 (range: 1.9-2.3), and 2.2 (range: 1.3-2.5), respectively. RNA integrity values are reported in Table C8. In order to make fair comparisons between RNA-seq and microarray datasets, RNA-seq reads were only considered if they could be mapped back to a transcript corresponding to a probe on the Agilent SurePrint G3 Mouse GE 8x60K microarray. Mapping produced a list of 20,167 RefSeq genes. This list was also used to retrospectively filter the list of microarray-derived DEGs. The numbers of DEGs in the full and mapped gene lists are listed in Table 6.1. With respect to the FRO groups, more DEGs were detected in the FFPE-18hr groups, and fewer were detected in the FFPE-3wk in all experiments apart from the poly-A-enrichment RNA-seq. One-colour DNA microarrays consistently detected about twice the number of DEGs as two-colour arrays across all sample types. Compared to microarray protocols, RNA-seq DEG numbers were intermediate (FRO and FFPE-3wk) or modestly lower (FFPE-18hr) for the ribo-depletion protocol, and either higher (FRO) or lower (FFPE-18hr and FFPE-3wk) for the poly-A protocol. Applying a 2-fold cut-off reduced

the number of DEGs by at least 75% across FRO and FFPE groups, limiting the ability to perform pathway analysis in some groups.

Next, we evaluated gene expression correlations by fold-change and sample type (Fig. 6.1A). The R^2 correlation coefficients between paired FRO and FFPE samples ranged from 0.29 to 0.93 depending on protocol; all correlations were statistically significant (regression $P < 0.05$). The correlation between FRO and FFPE-18hr samples was markedly higher for both ribo-depletion and poly-A RNA-seq (but not microarray) protocols compared to correlations between FRO and FFPE-3wk samples, indicating an erosion of signal with time in formalin at the read depths used. Overall, the ribo-depletion method clearly provided the highest correlation between FRO and FFPE samples (at both 18hr and 3wk).

Hierarchical clustering of the datasets demonstrated that furan treatment was the main driver of gene expression profiles (Fig.6.1B). A decrease in this signal with time-in-formalin is evident from the microarray cluster diagrams, which show a decrease in signal intensity in the furan groups from FRO to FFPE-18hr to FFPE-3wk (the increased amount of black on the heatmaps). Only five genes were up-regulated across all 12 datasets (*Osgin1*, *Tnfrsf12a*, *Srxn1*, *Ephx1*, and *Chka*). Genes up-regulated in 11/12 datasets were: *Cdkn1a*, *Jun*, *Dusp6*, *Hmox1*, *Brd2*, *Ugt2b35*, *Gclc*, and *Trp53inp1*; and in 10/12 datasets: *Tubb2a*, *Phlda3*, *Gdf15*, *Gsta2*, *Zmat3*, *Ugdh*, *Rcan1*, *Rbp1*, *Apoc2*, and *Ces2e*. For the four FRO datasets (ribo-depletion, poly-A-enrichment, 1-colour and 2-colour), the overlap of shared DEGs was: 78 (4/4 FRO datasets), 101 (3/4), and 206 (2/4). For the four FFPE-18hr groups, these numbers were: 54 (4/4), 63 (3/4), and 807 (2/4). And, for the four FFPE-3wks groups: 7 (4/4), 16 (3/4), and 112 (2/4); demonstrating an

erosion of quality with time in formalin. The DEG overlap for the three preservation methods (FRO, FFPE-18hr, and FFPE-3wk) for ribo-depletion was: 86 (3/3) and 165 (2/3); for poly-A-enrichment: 43 (3/3) and 233 (2/3); one-colour: 45 (3/3) and 167 (2/3); and, two-colour: 18 (3/3) and 76 (2/3). Therefore, once again, the ribo-depletion protocol was the most consistent between groups.

6.5.2 Pathway Analysis

Pathway analysis of each DEG list resulted in a large number of enriched pathways (Table 6.1), which were ranked for significance in IPA using the $-\log(\text{p-value})$ of a Fisher's Exact test. To simplify inter-group comparisons, we first considered the top 15 pathways from each group by sample type (FRO, FFPE-18hr, FFPE-3wk). Using FRO as a reference group, the overlap was: 10/15 (ribo-depletion), 5/15 (polyA-enrichment), 4/15 (one-colour microarray), 2/15 (two colour microarray) (Fig. 6.2). Once again, the greatest consistency among sample types was found with the ribo-depletion method. We then compared all enriched pathways between FRO and FFPE for each platform. The percent of FRO pathways that also occurred in FFPE-18hr were: 49% (ribo-depletion), 53% (poly-A-enrichment), 59% (1-colour), and 52% (2-colour) shared pathways. The percent of FRO pathways that also occurred in FFPE-3wk were: 41% (ribo-depletion), 18% (poly-A-enrichment), 24% (1-colour), and 23% (2-colour) shared pathways (Fig. 6.2, Venn diagrams). Overall, ribo-depletion and poly-A-enrichment were about equal at 18-hours; however, ribo-depletion was better at 3-weeks. One-colour out-performed 2-colour microarrays at both time-points.

Enriched pathways relevant to the MoA for furan (Jackson *et al.* 2014) were consistently enriched in FFPE samples (e.g., NRF2-mediated Oxidative Stress Response, Xenobiotic Metabolism, and p53 Signaling). Other relevant pathways related to cell cycle (G2/M DNA damage checkpoint), cell death (TWEAK signaling), inflammatory response (LPS/IL-1 inhibition of RXR function), and various cancer pathways were less consistently present across groups. Only the NRF2 Oxidative Stress Response and p53 Signaling pathways were shared across all four experimental set-ups, and only the former was enriched in all 12 datasets. The pathway analysis using the unmapped gene lists for each group can be seen in Fig. C5.

6.5.3 Meta-Analysis of Chemicals and Diseases

We performed a meta-analysis using NextBio to identify chemicals that produce similar changes in gene expression to furan in order to determine if these similarities were detectable in the FFPE furan samples. This meta-analysis was conducted using the data generated by the best-performing protocols from each platform: ribo-depletion RNA-seq and one-colour microarray protocols. We performed two chemical meta-analyses. In the first, expression profiles of the top 250 furan-dependent DEGs were compared to those produced in all mouse chemical studies (across the entire NextBio database). Furan was most similar to thioacetamide (on both platforms and across samples), followed by 1-5-naphthalenediamine, dimethylnitrosamine, methapyrilene, and diethylnitrosamine (Table 6.2). In the second, we compared our datasets with the changes produced following a 3-week exposure to 26 different chemicals in female B6C3F1 mice by Thomas *et al.* (2011b); this latter study was preselected as a more targeted study

subset with the same exposure time, mouse strain, and sex as the current furan study. Compared to the Thomas *et al.* study, furan produced gene expression changes most similar to malathion, benzofuran, methylene chloride, and 1,5-naphthalenediamine. There was a remarkable degree of overlap between FRO and FFPE samples, including identical top five rankings for FRO and FFPE-18hr groups across all NextBio datasets. Overall, the ribo-depletion study produced the best results.

A third meta-analysis was performed to identify liver and biliary disease states that produce similar gene expression profiles to furan exposure (Table 6.3). The most correlated studies involved differential gene expression following partial hepatectomy and exposure to carbon tetrachloride (a CYP2E1 substrate). Both of these are highly consistent with the known MoA for furan, which includes xenobiotic metabolism by CYP2E1, cytotoxicity, and regenerative proliferation of the liver. The FRO, FFPE-18hr, and FFPE-3wk groups shared 9/10 (ribo-depletion RNA-seq) or 10/10 (one-colour microarray) of the most correlated studies, indicating consistency across the two platforms and three preservation methods for the highest ranking DEGs.

6.5.4 RNA-Sequencing Sample Profiles

To examine the impact of library preparation on FRO and FFPE transcriptomic profiles, we prepared libraries using either poly-A-enrichment or ribo-depletion protocols, and sequenced the libraries on an Illumina HiSeq2000 instrument. In order to evaluate the different libraries, we examined different aspects of the RNA-seq data using quality metrics that have been previously developed (Wang *et al.* 2012), including gene-body coverage uniformity, guanine-cytosine (GC) content distribution, and the

distribution of the reads across genic structures (UTRs, exons, introns, or intergenic regions).

We observed that the gene body coverage (calculated as a percentage of reads that covered each nucleotide position of all RefSeq genes scaled to 100 bins) was dependent on both the library preparation method and the sample type (Fig. 6.3A). In particular, because poly-A-enrichment is 3' biased, and because FFPE samples are more degraded than their FRO counterparts, the uniformity of gene coverage was biased toward the 3' end of the gene for poly-A-enriched FFPE libraries. On the other hand, ribo-depleted RNA samples consistently showed more uniform gene coverage than did poly-A-enriched samples; however, after 3 weeks in fixation, even ribo-depleted samples began to show 3' bias, indicating that 3 weeks in fixation is too long if one is concerned about detecting full-length RNA transcripts.

Additionally, we observed that GC content distributions from each sample followed a bell curve, regardless of library preparation protocol (Fig. 6.3B). However, several differences were observed, namely in the widths and heights of those bell curves, likely indicating a difference in the population that was being sampled. Accordingly, the distribution of reads across genic structure (calculated as the fraction of the reads which mapped to exons, UTRs, introns, or intergenic regions) revealed differences between the two library preparation protocols and sample types (Fig. 6.3C). It is noteworthy that a higher read depth is required for ribo-depletion experiments (compared to poly-A-enrichment) in order to detect equivalent numbers of protein coding genes. In the current experiment the sequencing depth of the ribo-depletion experiment (10-60M reads) was greater than that used in the polyA-enrichment experiment (3.8-11M reads); however, the

latter maintained a larger proportion reads mapped to exonic regions (Fig. 6.3C). Overall, ribosomal RNA-depleted libraries consistently showed a larger proportion of introns being sequenced compared to poly-A-enriched samples, as expected (Li et al., 2014). Poly-A-enriched samples, on the other hand, showed increased exon and 3' UTR sequencing. Compared to FRO samples, poly-A-enriched FFPE samples showed an increasing number of reads mapping to the 3' UTR as the underlying RNA population degraded (i.e., as a function of time-in-formalin), in line with observed changes in gene body uniformity. FFPE ribo-depleted samples, on the other hand, showed an increasing proportion of reads mapping to introns as a function of fixation time as compared to FRO.

6.5.5 Long-Term Archival Samples

To examine whether these techniques would be applicable to a wide range of FFPE samples, we tested an additional set of FFPE samples. These archival FFPE samples were obtained from the NTP and had been banked for 8-26 years. It was not possible to obtain archival furan-exposed samples, so the decades-old samples that were tested here represented the control tissue blocks from toxicity studies for beta-picoline (8 years), oxazepam (19 years), and trichlorfon (26 years). These samples were derived from rat; however, it is unlikely that formalin would interact with the genome in a species-specific manner and thus we would expect to see the similar results for mouse. We sequenced each set of samples using the ribo-depletion protocol and compared the expression profiles to standard metrics established by the FDA's SEQC RNA-seq Consortium (Li et al. 2014a).

We observed that the ribo-depletion protocol was very successful in creating expected distributions of mapped reads to the exonic, UTR, intronic, and ribosomal fractions of the transcriptome (Fig. 6.3D). The only aspect of the mapped reads that was high was the percent of mapped reads that were intergenic (violet), which could indicate degradation from the long-banked samples, although we found this trend across all samples. Moreover, the percentage of guanosine-cytosine (GC) from each sample followed a normal distribution (Fig. 6.3E), and did not show any significant shift as a function of treatment or time. Overall, these results demonstrate the potential applicability of these techniques for very old FFPE samples.

6.6 Discussion

Archival FFPE samples provide a unique opportunity to directly link molecular profiles and histopathologic or clinical outcomes without the cost and time associated with new studies. Resulting profiles provide detailed mapping of target pathways, identification of biomarkers for both environmental chemicals and pharmaceutical agents, and insight for understanding of MoAs. However, better methods for transcriptomic analysis of FFPE tissues are needed. Several recent proof-of-concept studies have evaluated RNA-seq as a novel way to derive gene expression profiles from FFPE or degraded RNA samples and reported promising results (Linton *et al.* 2012, Spencer *et al.* 2013, Auerbach *et al.* 2014, Hedegaard *et al.* 2014, Zhao *et al.* 2014b). Our goal was to expand on these studies by examining specific questions related to the use of RNA-seq in FFPE tissues, including characterizing the performance versus microarray platform, discerning the optimal library preparation method, and examining the effects of time-in-

formalin and age-in-block. The ribo-depletion RNA-seq method with a relatively brief (18 hour) fixation time provided the strongest correlation between FRO and FFPE tissues at a global transcriptomic level. For highly altered DEGs concordance was also observed across FRO and FFPE samples for both RNA-seq and microarray platforms. Increased time-in-formalin and age-in-paraffin resulted in loss of signal intensity and other indicators of degradation, but not to the extent that key molecular signatures were lost.

Our first aim was to compare both platforms (two protocols each) in order to develop a pipeline for future FFPE experiments. Across platforms, ribo-depletion RNA-seq provided the strongest gene fold change correlations between FRO and FFPE (Fig. 6.1) and the highest overlap of top pathways (Fig. 6.2). Within platforms, ribo-depletion outperformed poly-A-enrichment RNA-seq, and one-colour outperformed two-colour microarrays. Compared to the poly-A-enrichment protocol, the ribo-depletion protocol produced a more even coverage across all genes and also revealed more low-expressing transcripts. Meta-analyses of the ribo-depletion datasets also showed high concordance in top chemical (Table 6.2) and disease (Table 6.3) signatures. The advantage of the ribo-depletion approach over the poly-A-approach is that it does not rely on the poly-A tail of the mRNA molecules, which is known to be highly modified and degraded by formalin (Farragher *et al.* 2008, Klopfleisch *et al.* 2011). The strength of the RNA-seq platform over microarrays is that it has (essentially) an unlimited dynamic range, and reads are not confined by predetermined probe sequences. However, because microarrays have been in use for much longer, the data normalization, processing, and analysis associated with them is more established and straightforward. Collectively, these findings support the ribo-depletion approach for future FFPE genomics experiments.

Our second aim was to examine how formalin fixation impacts RNA quality and the detection of a chemical signature. Formalin generates crosslinks between cellular macromolecules and introduces various adducts and other covalent modifications. Sample fixation time varies (from hours to weeks) across laboratories; therefore, an important consideration in planning a retrospective FFPE genomics study is to understand how the initial fixation in formalin will affect the quality of the final gene expression data. Our findings indicate that longer fixation times decrease RNA yields from paraffin-embedded tissues (herein 3 weeks in formalin resulted in half the RNA yield/section compared to 18 hours) and increase RNA input needed during library preparation (see methods). This latter effect is likely due to “non-functional” RNA (e.g. adducts or damaged bases) that could be potentially mitigated by treatments like PreCR or other enzymatic and chemical methods. Extended time in formalin (3 weeks) led to lower correlation with FRO samples (RNA-seq) (Fig. 6.1A), detection of fewer DEGs (RNA-seq and microarray) (Table 6.1), and erosion of signal intensity (Fig. 6.1B) (microarray). It is unclear whether this erosion in signal occurs early, or if signal continues to decline beyond 3 weeks. Given that time in formalin varies, future studies should investigate samples held in formalin for longer durations.

Samples that stayed in formalin for only 18 hours produced a larger number of DEGs than their FRO counterparts, especially for the microarray protocols. Reasons for this are not clear, but the same phenomenon has been observed previously, e.g., Hedegaar *et al.* (2014), suggesting that FFPE samples have an increased propensity for producing false positive genes. To account for this, future experiments may consider including technical replicates and only treating DEGs that appear in both replicates as true

positives. Also, RNA-seq is more variable at lower read depths; thus, deeper sequencing is another way to increase signal intensity and reduce false-positives. Moreover, our results suggest that longer incubations in formalin may reduce DEG numbers due to lower signal intensity. These data indicate that FFPE genomics studies should, whenever possible, favour samples that have spent limited time in formalin. Our findings highlight the fact that FFPE tissue blocks may vary widely in RNA quality (despite similar RIN profiles), pointing to a need for improved quality metrics specific to FFPE samples.

An important future application of FFPE genomics is the identification of key molecular features in a chemical or drug MoA. Despite inter-platform and inter-preservation differences in DEG lists, the higher-level biology (pathways and meta-analyses) detected in the FFPE groups in our study (particularly the FFPE-18hr) corresponded well with paired FRO groups and the expected furan-induced molecular changes. Concordance was strongest for the ribo-depletion RNA-seq protocol. The test article in this study, furan, is a well-characterized hepatocarcinogen with an extensive toxicological literature. We recently characterized early molecular key events in the MoA of furan in female mice using FRO livers and identified a key role for the Nrf2 Oxidative Stress Response pathway (Recio *et al.* 2013, Jackson *et al.* 2014). In the current study, this pathway was consistently enriched across all groups and platforms. Pathways enriched in the FRO groups were typically well represented in the corresponding FFPE-18hr groups, while the FFPE-3wk groups showed fewer response pathways. These findings indicate that samples fixed in formalin for shorter periods of time provide more consistent pathway-level results.

In addition to pathway analysis, we performed two meta-analyses using NextBio software. This approach showed higher inter-group consistency than the pathway analysis, likely because it does not group gene expression profiles into specific events (such as pathways). Instead, comparisons are made between entire biosets generated from different chemical treatments or disease states, allowing comparisons between data-poor and data-rich compounds. Here, furan gene expression changes were most similar to those induced by thioacetamide and carbon tetrachloride, which are also metabolized in the liver by CYP2E1 (Manibusan *et al.* 2007, Kang *et al.* 2008). Moreover, the most similar disease state was liver regeneration following partial hepatectomy, consistent with previous studies showing that the carcinogenic MoA for furan includes cytotoxicity followed by dysregulated regenerative proliferation (Fransson-Steen *et al.* 1997, Moser *et al.* 2009). These meta-analyses indicate that FFPE samples can provide valuable expression-based information for prioritization of chemicals and next-tier studies.

Lastly, we investigated effects of age-in-paraffin block on RNA quality metrics. No clear effects of age-in-block were observed, supporting the idea that older FFPE samples could be used for transcriptome profiling (Fig. 6.3). Other researchers have reported RNA-sequencing for 9-year-old and 20-year-old samples (Meng *et al.* 2013, Hedegaard *et al.* 2014), but to our knowledge, these are the oldest samples tested. While the data from these samples did not have the matching FRO samples for comparison, the data suggest that changes in RNA quality parameters due to age-in-block may be limited. Indeed, decades-old samples appear to be a viable option for FFPE genomics, but additional studies are needed to define optimal preanalytical and bioinformatics approaches.

While there are differences between the genomic profiles of FRO and FFPE samples, our study demonstrates that: (1) RNA-seq with ribo-depletion library preparation is the preferred method for transcriptomic analysis of FFPE tissues; (2) FFPE samples can be used to obtain reliable transcriptomic information related to toxicant MoA; and (3) the RNA profile of a compound is affected by the amount of time in formalin (although this effect becomes less pronounced with respect to higher-level analyses). Time-in-paraffin did not have a large effect on distributions of mapped reads across control FFPE samples. These findings provide evidence that FFPE samples can be used in genomics studies in absence of existing FRO samples (or when new animal studies are not feasible). Additional studies could also evaluate whether higher read depth may compensate in part for the loss of signal observed in FFPE samples formalin-fixed for longer periods of time, which could impact work from archival samples in quantitative applications such as transcriptomic benchmark dose estimation. This type of work represents an important opportunity for performing retrospective, phenotypically-anchored research that characterizes chemical- and disease-dependent genomic changes. These chemical/disease signatures can be applied and reapplied (via meta-analysis) to the many data-poor environmental chemicals in need of assessment. This information would also provide a critical bridge between emerging *in vitro* data and health effects established in prior toxicologic and epidemiologic studies, accelerating the input and value of higher-throughput models in safety assessment, biomarker development, and drug-discovery.

6.7 Tables and Figures

Table 6.1 Number of differentially expressed genes (by unique gene symbol) in each microarray and RNA-seq experiment using different filtering thresholds. The corresponding number of enriched IPA pathways (with $p < 0.05$ and at least 4 molecules) is indicated in brackets. Lists that were used to derive enriched pathways are indicated by bold text.

		FRO				FFPE							
						18 hours				3 week			
fold change		± 1.5		± 2		± 1.5		± 2		± 1.5		± 2	
$p \leq 0.05$		<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>unadj.</i>	<i>FDR</i>	<i>unadj.</i>
Mapped RefSeq	<i>RNA-seq</i>												
	Ribo-	307 (37)	768	129	207	576 (44)	814	189	248	204 (25)	928	77	451
	PolyA+	609 (55)	675	189	199	438 (36)	741	160	249	128 (11)	1447	51	812
	<i>Microarray</i>												
	1 colour	443 (35)	489	103	111	1656 (66)	1847	302	323	362 (20)	739	72	118
2 colour	197 (35)	234	47	53	798 (90)	798	77	77	149 (29)	152	12	12	
Full RefSeq	<i>RNA-seq</i>												
	Ribo-	316 (44)	836	133	226	619 (46)	883	208	265	212 (25)	986	76	483
	PolyA+	659 (58)	728	206	217	455 (37)	783	165	257	136 (11)	1519	58	870
	<i>Microarray</i>												
	1 colour	667 (28)	811	157	173	3543 (133)	3753	626	640	491 (40)	887	69	105
2 colour	339 (32)	405	84	93	1854 (196)	1853	192	188	266 (48)	272	25	24	

Table 6.2 Chemical Signature Analysis. The expression values of the top 250 mapped DEGs (FC > 1.5, FDR P < 0.05 for furan-treated mice versus respective controls) stratified by sample group (FRO, FFPE-18hr, FFPE-3wk) were compared to liver gene expression signatures from all NextBio mouse studies (top) and GSE18858 (Thomas *et al.* 2011; bottom). In the latter, female B6C3F1 mice were exposed to one of 26 chemicals for 13 weeks. The top five most correlated datasets for each group are indicated by bold text (where 1 indicates the most correlated study, 2 the second most correlated, etc.); only positively correlated chemicals with at least 3 supporting studies were included (all correlations are P < 0.05, which was calculated in NextBio using a Fisher's exact test).

	<i>Ribo-depletion RNA-seq</i>			<i>One-colour microarray</i>		
	Time in formalin			Time in formalin		
	FRO	18hrs *	3wks *	FRO	18hrs	3wks
All NextBio						
Thioacetamide	1	1	2	1	1	1
1,5-Naphthalenediamine	2	2	1	6	31	3
Dimethylnitrosamine	3	3	5	3	35	25
Methapyrilene	4	4	3	2	2	2
Diethylnitrosamine	5	5	13	5	26	10
Anastrozole	6	14	14	4	17	6
Tunicamycin	7	19	6	7	10	31
Bleomycin	8	9	91	-	-	-
Benzo(a)pyrene	9	4	11	-	-	-
Bromobenzene	10	18	8	8	15	7
GSE18858						
	FRO	18hrs *	3wks	FRO	18hrs *	3wks *
Malathion (14800ppm)	1	1	1	1	1	1
Benzofuran (240mkd)	2	2	2	3	2	2
Methylene chloride (4000ppm)	3	6	14	5	14	14
1,5-Naphthalenediamine (2000ppm)	4	3	3	4	3	3
Iodoform (93mkd)	5	-	4	2	4	4
Methylene chloride (3000ppm)	6	5	10	6	10	10
Methylene chloride (2000ppm)	7	12	11	9	11	11
1,4-Dichlorobenzene (600mkd)	8	4	5	7	5	5
1,2,3-Trichloropropane (60mkd)	9	11	13	10	13	13
1,4-Dichlorobenzene (500mkd)	10	10	8	8	8	8

Kendall's rank correlation for all NextBio (* indicates a significant correlation with respect to FRO):

- Ribo- RNA-seq: FRO vs. 18hrs (tau=0.674; **p=0.0071**); FRO vs. 3wks (tau=0.511; **p=0.0466**)

- 1C microarray: FRO vs. 18hrs (tau=0.214; p=0.5484); FRO vs. 3wks (tau=0.429; p=0.1789)

Kendall's rank correlation for GSE18858:

- Ribo- RNA-seq: FRO vs. 18hrs (tau=0.556; **p=0.0446**); FRO vs. 3wks (tau=0.467; p=0.0726)

- 1C microarray: FRO vs. 18hrs (tau=0.6; **p=0.0167**); FRO vs. 3wks (tau=0.6; **p=0.0167**)

Table 6.3 Disease Signature Analysis. The expression values of the top 250 mapped DEGs ($FC > 1.5$, $FDR p < 0.05$) in each group (FRO, FFPE-18hr, FFPE-3wk) determined by RNA-seq or microarray analysis were compared to gene expression signatures derived from publically available mouse liver and biliary disease datasets using NextBio. The top 5 most correlated datasets for each group are indicated by bold text (where 1 indicates the most correlated study, 2 the second most correlated, etc.); only positively correlated studies were included. All correlations have a $p < 0.05$, which was calculated in NextBio using a Fisher's Exact Test.

Treatment	Time of tissue collection	Mouse details	GEO	FRO	FFPE	
					18hrs *	3wks *
<i>Ribo-depletion RNA-seq</i>						
Partial hepatectomy	48hrs post-PH	5-6m/CB6F1	GSE20427	1	1	2
Partial hepatectomy	48hrs post-PH	25-27m/CB6F1	GSE20427	2	3	4
Partial hepatectomy	18hrs post-PH	Adult/wildtype C57BL/6	GSE51801	3	2	5
CCl4-exposure	2d post-exposure	BALB/c (IL4 IL13 dKO)	GSE45002	4	4	1
CCl4-exposure	2d post-exposure	BALB/c (wildtype)	GSE45002	5	6	6
Partial hepatectomy	2hrs post-PH	25-27m/CB6F1	GSE20427	6	5	3
Partial hepatectomy	2hrs post-PH	5-6m/CB6F1	GSE20427	7	9	-
Partial hepatectomy	24hrs post-PH	Tob1-null	GSE21836	8	8	7
Partial hepatectomy	38hrs post-PH	5-6m/CB6F1	GSE20427	9	7	9
Partial hepatectomy	24hrs post-PH	5-6m/CB6F1	GSE20427	10	10	8
<i>One-colour microarray</i>						
				FRO	18hrs	3wks *
Partial hepatectomy	48hrs post-PH	5-6m/CB6F1	GSE20427	1	3	2
Partial hepatectomy	48hrs post-PH	25-27m/CB6F1	GSE20427	2	4	4
CCl4-exposure	2d post CCl4-exposure	BALB/c (wildtype mice)	GSE45002	3	9	5
CCl4-exposure	2d post CCl4-exposure	BALB/c (IL4/IL13 dKO mice)	GSE45002	4	2	3
Partial hepatectomy	18hrs post-PH	Adult/wildtype C57BL/6	GSE51801	5	1	1
Partial hepatectomy	38hrs post-PH	5-6m/CB6F1	GSE20427	6	7	6
Partial hepatectomy	24hrs post-PH	Tob1-null (Tob1 is a repressor of liver regeneration)	GSE21836	7	5	8
Ehrlichia chaffeensis induced Hepatitis	15d Wakulla strain	SCID	GSE8966	8	8	9
Partial hepatectomy	24hrs post-PH	5-6m/CB6F1	GSE20427	9	6	7
Partial hepatectomy	2hrs post-PH	25-27m/CB6F1	GSE20427	10	10	10

PH = partial hepatectomy; CCl4 = carbon tetrachloride; m=month

All samples are: mouse; liver; treated vs. control

Kendall's rank correlation (where * indicates a significant correlation with respect to FRO):

- Ribo- RNA-seq: FRO vs. 18hrs ($\tau=0.778$; $p=0.0009$); FRO vs. 3wks ($\tau=0.611$; $p=0.0247$)

- 1C microarray: FRO vs. 18hrs ($\tau=0.378$; $p=0.1557$); FRO vs. 3wks ($\tau=0.644$; $p=0.0091$)

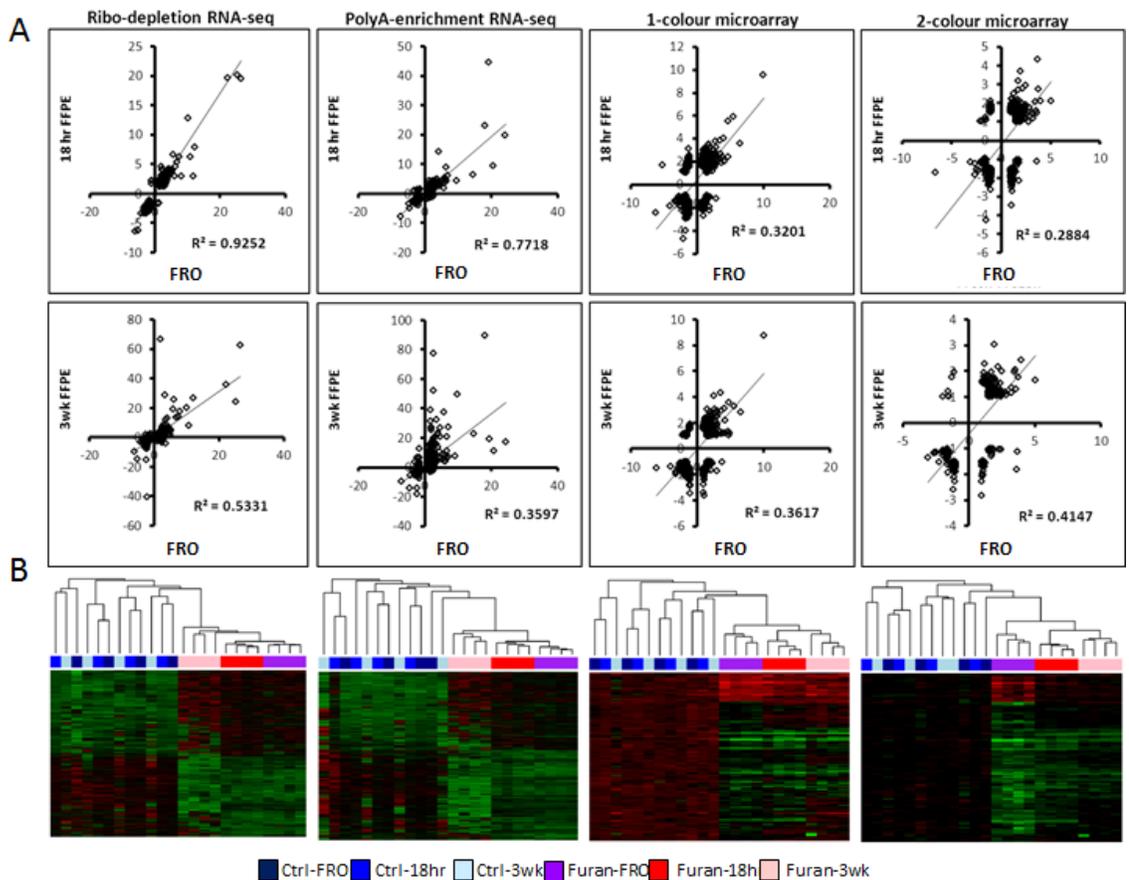


Figure 6.1 Inter-group comparisons of differential gene expression. Expression profiles following exposure to furan (8 mkd) in FRO, FFPE-18hr, or FFPE-3wk liver samples evaluated using different DNA microarray and RNA-seq protocols. (A) Correlation analysis of DEG fold changes for FRO versus FFPE-18hr (top row) and FRO versus FFPE-3wk (bottom row) ($P < 0.0001$ for all linear regressions). Genes were significantly changed in at least one list (FDR $P \leq 0.05$, fold change $> \pm 1.5$ compared to control). (B) Hierarchical clustering of all DEGs (FDR $P \leq 0.05$, fold change $> \pm 2$).

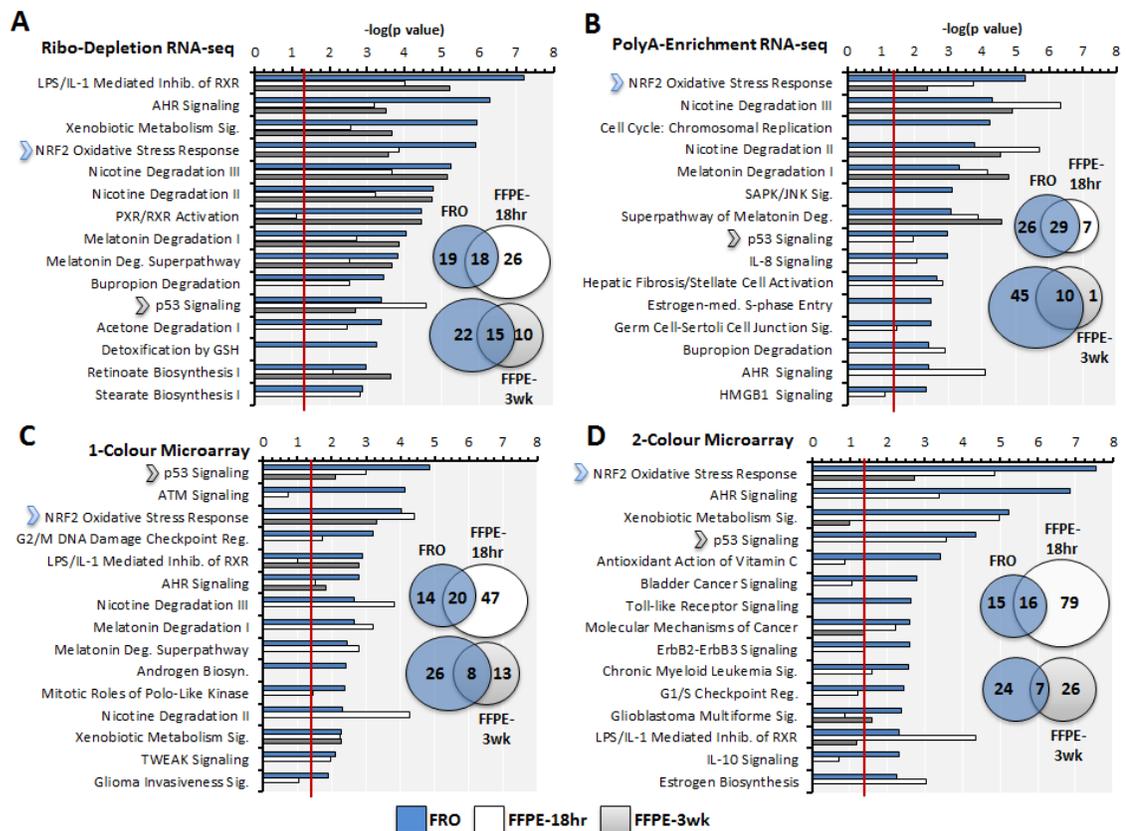


Figure 6.2 Pathway analysis. (A) Ribo-depletion RNA-seq, (B) poly-A-enrichment RNA-seq, (C) one-colour microarrays, and (D) two-colour microarrays. The top 15 pathways for the fresh-frozen (FRO) groups (blue) are listed vertically. The level of enrichment of the corresponding FFPE groups (white-FFPE-18hrs; grey-FFPE-3wks) is plotted. Pathway significance was calculated in IPA using a Fisher Exact test; $-\log(p\text{-value}) = 1.3$ corresponds to a $p = 0.05$ (indicated in red). Venn diagrams indicate the overlap of each full list. For all analyses, only mapped DEGs with $FDR\ p < 0.05$ and fold change $> \pm 1.5$ were used and only pathways with at least four DEGs were considered. Nrf2 Oxidative Stress Response was the only pathway that was enriched in all 12 groups and p53 was enriched in 11/12 groups (indicated by chevrons).

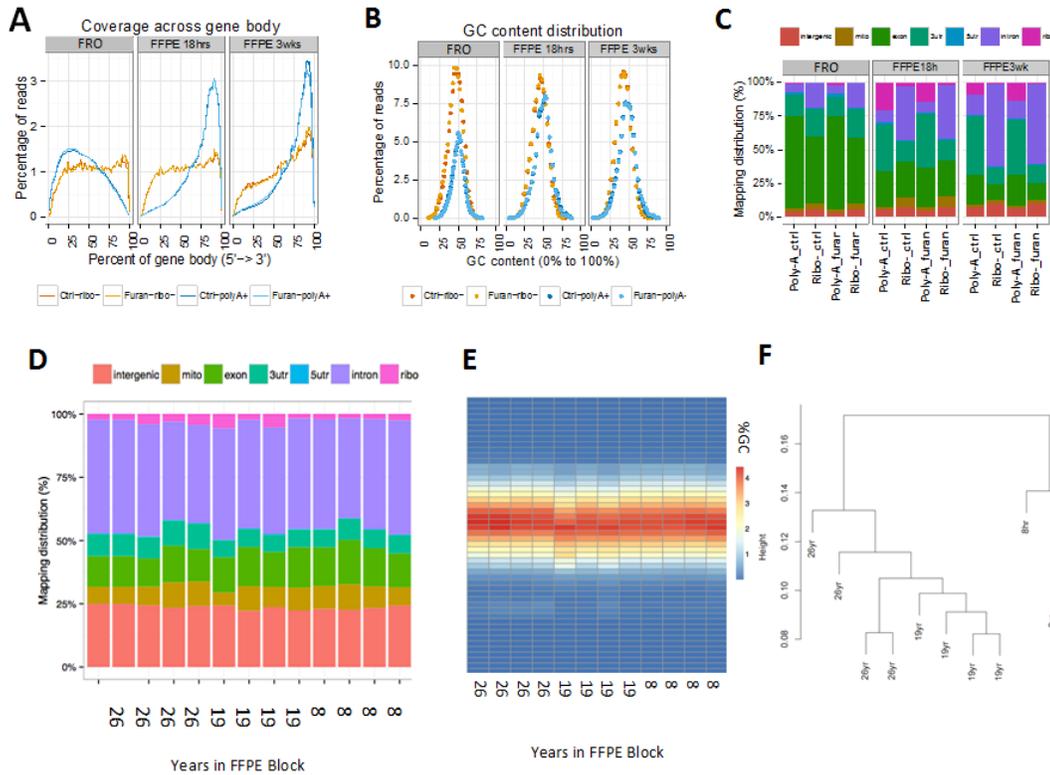


Figure 6.3 Examination of RNA-seq data quality metrics. (A-C) Mapping distributions and GC content distributions for the furan-treated sampled. (A) The percentage of reads that covers each nucleotide position of all of genes scaled to 100 bins, from 5' UTR to 3' UTR for FRO and FFPE samples. The ribo-depleted RNA samples consistently showed more uniform gene coverage than did their poly-A–selected counterparts. (B) GC content distributions for FRO and FFPE samples. (C) The percentage of FRO and FFPE reads that map to various genic regions. Overall, we detect more intronic reads from ribo-depleted RNA samples than from poly-A–enriched libraries; this trend increased as a function of time-in-formalin. (D-E) Mapping distributions and GC content distributions, expressed as a percentage, for 8-, 19-, and 26-year-old control FFPE tissues. (F) Inter-group Kendall's tau comparison.

CHAPTER 7

7 CHAPTER: FINAL DISCUSSION

7.1 Summary of Study Outcomes

The broad goal of this thesis was to explore toxicogenomics as an approach that produces informative toxicity data more quickly and using fewer experimental animals than standard approaches. Using furan as a model compound we demonstrated that these results can be obtained through quantitative, predictive, and mechanistic toxicogenomics approaches using several genomics technologies. Sub-chronic toxicogenomics testing was able to predict the dose at which adverse effects in long-term tests arise, predict hazard, and provide insight into MoA. We also demonstrated that these data can be generated from archival FFPE samples, whose use can accelerate the development of phenotypically anchored, predictive chemical signatures. The outcomes of the specific hypotheses tested in and experimental objectives of each chapter are described briefly below.

Chapter 2

Hypothesis: Furan-dependent transcriptional changes following a 21-day exposure will provide predictive and mechanistic information regarding furan's MoA and will be prognostic of the dose at which furan-dependent liver cancer occurs in mice.

Outcome: The hypothesis was supported. We demonstrated that the BMDs of important transcriptional endpoints were consistent with the BMDs for furan-dependent liver cancer. We demonstrated that similarities in global gene expression patterns are predictive of adverse outcomes (in this case, liver cancer) by comparing the

transcriptional profiles of furan-exposed mouse livers to other publically available transcriptomic datasets. Bioinformatics software (IPA, NextBio) revealed pathway perturbations and similarities to known toxicants that provided mechanistic understanding of furan's MoA. Specifically, pathway enrichment analyses demonstrated a key role of NRF2 oxidative stress response in furan's MoA, which enabled the development of a model for NRF2 in balancing cytotoxicity with regenerative proliferation to maintain a healthy liver at doses below 2 mkd, but resulting in liver cancer at doses above 4 mkd.

Conclusions: Toxicogenomics is a useful tool for understanding and assessing carcinogens. Although the data obtained from toxicogenomics studies are very different from results obtained in standard toxicity and cancer assays, toxicogenomics can be used to predict carcinogenic dose, identify disease hazards, and can be used to derive mechanistic information.

Chapter 3

Hypothesis: BrdU treatment does not mask furan-dependent changes in gene expression.

Outcome: The hypothesis was supported. Furan-dependent changes in gene expression were detectable in BrdU-treated mice. Further, there were only minor differences in gene expression profiles between the two cohorts of mice (\pm BrdU).

Conclusions: Our results indicate that toxicogenomics studies can be carried out in +BrdU-treated animals. Importantly, this means that the –BrdU animal cohorts can be eliminated from future toxicogenomics studies, thereby halving the number of experimental animals required.

Chapter 4

Objectives:

- To compare BMD values for furan modeled using data from three genomics platforms: 2-colour Agilent 8 x 60K SurePrint Gene Expression microarrays, PolyA-enrichment RNA-sequencing (Illumina HiSeq2000), and qPCR arrays (SA BioSciences).
- To understand the effects of statistical filtering of genomics data on BMDExpress output values.
- To compare four approaches for choosing a transcriptional PoD.

Outcomes:

- Overall trends in BMD values obtained from different platforms were similar; however, individual values were not well correlated.
- Increasing statistical stringency significantly decreased gene and pathway mean, median and mode BMD values, and produced more conservative BMD values.
- All four approaches for choosing a transcription PoD that we tested produced values that were consistent with the cancer BMD values. Therefore, we concluded that

transcriptional profiling provides a robust approach for identifying a PoD that represents the dose at which significant molecular perturbations occur.

Conclusions: Quantitative toxicogenomics (i.e., benchmark dose modeling) is a relatively new approach. Currently there is no consensus regarding how the data should be handled or presented. Using furan as an example, this study demonstrated that fundamental experimental factors, such as genomics platform and statistical treatment of data, can impact BMD values, which should become an important consideration during the design of future quantitative toxicogenomic experiments. Importantly, transcriptional and cancer BMDs were generally consistent.

Chapter 5

Objective: To develop a protocol for RNA extraction and microarray analysis of FFPE samples. Specifically, to compare the quality of RNA produced from six commercially available FFPE RNA extraction kits and the quality of data from four one- and two-colour microarray protocols.

Outcome: We demonstrate that using an appropriate RNA extraction and DNA microarray protocol, a molecular signature consistent with phenobarbital's established MoA was clearly observed in 16-year-old archived mouse tissues. Our recommended protocol for gene expression analysis of FFPE tissue samples is RNA extraction using the RNeasy FFPE kit (Qiagen) and gene expression analysis using one-colour microarrays (Agilent).

Chapter 6

Hypotheses:

1. Furan-dependent changes in gene expression can be detected in FFPE liver samples.
2. Ribo-depletion RNA-seq will outperform polyA-enrichment RNA-seq, and 1-colour microarrays will outperform 2-colour microarrays for global gene expression profiling of FFPE tissue samples.
3. Increased time in formalin will negatively impact gene expression data.
4. Increased age of tissue block will negatively impact gene expression data.

Outcomes:

- Hypothesis 1 was supported. While there were differences between the FRO and FFPE samples, important furan-dependent transcriptional changes were detectable in the FFPE samples.
- Hypothesis 2 was supported. The best protocol was the ribo-depletion RNA-seq approach.
- Hypothesis 3 was supported. The samples that were fixed in formalin for 18 hours produced better gene expression data than did those that were fixed in formalin for three weeks.
- Hypothesis 4 was not supported. RNA-seq quality control metrics did not appear to be influenced by the age of the FFPE tissue block.

Conclusions: For the field of predictive toxicogenomics to reach its full potential it is absolutely necessary that a large collection of robust, phenotypically anchored molecular

signatures be available against which novel datasets can be compared. Using archival tissues to build these gene expression libraries would be beneficial because: (1) the samples are well characterized; (2) most are from chemical trials that would be impractical to repeat; (3) using existing samples is less expensive and quicker than generating new samples; and, (4) using existing samples represents a strategy for reducing experimental animal usage in toxicity studies. Our furan case study demonstrated that FFPE samples maintain important, chemical-dependent gene expression patterns and can feasibly be used in toxicogenomics studies.

7.2 General Conclusions and Contributions to Scientific Knowledge

7.2.1 Furan Case Study: Major Conclusions

Overall, this thesis provides an excellent example of the utility of toxicogenomics data for providing chemical-specific quantitative, predictive, and mechanistic data (Fig. 7.1). This section will describe the main findings of each of these approaches for furan.

Quantitative toxicogenomics is an extremely useful tool because it can be completed quickly and it provides concrete BMD values that indicate the dose at which the transcriptome becomes perturbed. The BMD values can be used to identify the compound's PoD, which can then be used to calculate the MoE and set acceptable exposure limits. This approach does not provide any indication as to what adverse health outcome the compound might produce, nor does it provide any mechanistic insight. Using furan-dependent changes in gene expression we observed that transcriptional BMD values are consistent with cancer BMD values, regardless of genomics platform used. The furan-dependent cancer PoD is 2.6 mkd (for HCA), which is remarkably similar to

the modes of the pathway BMD values, which were 2.7, 3.2, and 4.4 mkd for RNA-seq, microarray, and qPCR, respectively (Fig. 4.2). We further observed that modeling data for genes that have been altered by the treatment in a statistically significant way produces more conservative values (compared to modeling the entire dataset). Finally, we calculated MoE values using transcriptional BMDLs and note that these values fall between those calculated for HCA and HCC (Table 2.3). Quantitative toxicogenomics is a relatively new approach and the concordance between cancer and transcriptional BMD values should be confirmed using additional chemicals; however, the results presented here are very promising.

Predictive toxicogenomics begins to answer the question of what adverse health outcomes are produced by the compound, and can be accomplished using open-source resources or paid software. For the former, the gene expression data of the test compound can be compared to any publically available dataset (e.g., any dataset in the NCBI's GEO) using a hierarchical clustering package (Fig. 2.1). This is an extremely powerful approach; however, it requires advanced bioinformatics and biostatistical skill sets. An alternative approach is to use commercially available software. For example, NextBio software is a user-friendly interface that allows users to compare their gene expression datasets to all of the gene expression datasets in the NextBio knowledgebase, which houses a large number of disease- and chemical-specific expression profiles. Using both of these approaches, we report that following a 21-day exposure to carcinogenic doses of furan, the changes in global gene expression that are produced are most similar to those produced in spontaneous liver cancer, NF-kappaB-mediated inflammation, thioacetamide exposure, carbon tetrachloride exposure, and liver regeneration following partial

hepatectomy. These correlations are consistent with different aspects of furan's known MoA. Determining which diseases and chemicals produce changes in gene expression that are most similar to those produced by the test compound is especially useful for read-across (when the effects of well-studied chemicals are used to infer the effects of a data-poor chemical). Like quantitative toxicogenomics, predictive toxicogenomics can be done relatively quickly but, unlike the quantitative approach, it also provides some insight into the potential health hazards of the compound.

Mechanistic toxicogenomics provides a deeper understanding of the molecular effects of the compound, but can be time consuming. Therefore, when time is limiting (e.g., in rapid screening applications), it is possible to assess a compound using only the quantitative and predictive approaches. In fact, it has been argued that mechanistic understanding of a compound is not always required for the use of toxicogenomics data in chemical risk assessment (Thomas *et al.* 2013c). However, mechanistic data are important for establishing human health relevance (Holsapple *et al.* 2006, Kienhuis *et al.* 2011). The compound's MoA is developed using functional and pathway analyses (using software such as IPA and DAVID) and intensive literature review. The predominant molecular pathway that was perturbed by furan was the NRF2 Oxidative Stress Response Pathway (Figs. 2.3, 6.2). Mechanistic analyses are limited by a number of parameters, including: the expertise of the researcher, and the amount of ambiguity intrinsic to the dataset. Examples of the latter include: (1) increased expression of genes with conflicting functions; enrichment of pathways with conflicting (or unrelated) functions; (2) functional redundancy across genes or pathways; (3) the fact that gene expression data cannot (directly) capture post-translationally controlled processes (e.g., phosphorylation

during MAPK signaling cascades, or protein cleavage during caspase cascades) and thus activation of these processes must be inferred through changes in downstream targets; and, (4) the fact that mRNA that is extracted from whole tissue is extracted from a heterogeneous cell population (e.g., liver is comprised of hepatocytes, BECs, Kupffer cells, etc.) that may react to the treatment in different ways. Ultimately, accurate interpretation of toxicogenomics data requires substantial understanding of molecular biology and cell physiology. It is possible that MoA development will become swifter once the Adverse Outcome Pathway (AOP) wiki is fully populated (discussed in 7.4.2). Because a mechanistic approach is the most time-consuming of the three, it is important to have clearly defined goals at the outset of toxicogenomic studies with respect to the required outputs. Specifically, researchers and risk assessors should determine whether assessment deadlines permit a mechanistic analysis and whether the mechanistic information will significantly improve their ability to assess the compound.

7.2.2 Toxicogenomics: A Cost-Benefit Analysis

Using a toxicogenomic approach, the associated costs are much lower than a two-year cancer bioassay. Our furan case study used 50 mice (compared to ~800 mice and rats used by the NTP), that were exposed to four doses of furan plus control (compared to the three or two doses plus control in rats and mice, respectively), for three weeks (compared to two years). Approximately one year was required to complete the project (Chapter 2), from lab work, to quantitative, predictive and mechanistic analyses, to manuscript publication. The microarray experiment cost approximately \$10,000 (compared to \$2-4 million for a two-year rodent bioassay). Alternatively, conducting this

experiment using RNA-seq would cost approximately \$5,000 (~\$200/sample); although costs vary with depth of sequencing. Even with these improved efficiencies, it would not be practical to test all compounds using toxicogenomics; therefore, a strategy for prioritization of compounds for toxicogenomics testing is required.

7.3 How Toxicogenomics Fits into the Current Toxicity Testing Paradigm

7.3.1 Tiered Testing

“Toxicity Testing in the 21st Century: A Vision and a Strategy” is a ‘call to action’ publication produced by the National Academy of Sciences (USA, 2007) that criticizes standard approaches for toxicity testing and proposes a provocative plan for the future of chemical testing that is heavily reliant on high-throughput, systems biology approaches that are analyzed using sophisticated bioinformatics. Further, the vision proposes that toxicological assays should be primarily conducted *in vitro*, with a preference for human cell lines (Krewski *et al.* 2010). The report also emphasizes the identification and use of predictive toxicity pathways and dose-response modeling in the assessment of novel, putative toxicants. In the wake of this publication the toxicology community remains in a state of upheaval as researchers collaborate (and compete) to establish a novel testing paradigm that can be universally applied (and agreed upon). Recently, a three tiered paradigm for chemical testing and prioritization was proposed in which toxicogenomics appears in tier two (Thomas *et al.* 2013b). This tiered strategy is heavily reliant on the calculation of MoE values, which are used to include/exclude compounds from next-tier testing based upon the quantification of perceived risk. A MoE is the ratio of the PoD dose to the estimated exposure; a high MoE indicates that the

harmful dose is much larger than the expected exposure dose (which corresponds to low risk to human health). MoE cut-offs are typically set in a chemically specific manner. Well characterized chemicals tend to have lower MoE cut-offs (100-1000) because there is less uncertainty associated with them; whereas, poorly characterized chemicals have higher, more conservative MoE cut-offs (10,000) that correspond to a greater number of uncertainty factors. It is noteworthy that the tiered test strategy was developed for testing all types of toxicants; however, the focus of this discussion will be its application to carcinogens.

Tier one is composed of five components that rely entirely on *in vitro* and *in silico* assays. (1) Chemicals are divided into “selective” (compounds with a clear MoA, e.g., receptor mediated) and “nonselective” (compounds with an ambiguous MoA) using high-throughput *in vitro* assays that test chemical selectivity against a large number of biological targets and were developed for ToxCast (<http://epa.gov/ncct/toxcast/>) (Judson *et al.* 2010, Knudsen *et al.* 2011, Judson *et al.* 2014). ToxCast comprises over 700 assays that measure a compound’s activity both in living cells, and in non-cellular assays that use purified proteins, with the goal of understanding how the compound would behave in a more complex biological context. Nonselective compounds (i.e., compounds that are active in a large number of the ToxCast assays) are expected to make up the majority of the test articles. (2) Next, chemicals are divided into “genotoxic” and “non-genotoxic” MoAs based on the results of *in vitro* genotoxicity assays (typically the Ames test and the mammalian cell micronucleus test). (3) IVIVE pharmacokinetic modeling is used to derive an oral equivalent dose for a key event (selective MoA) or most sensitive *in vitro* assay (non-selective MoA), which can be used as the PoD (4). Together with a human

exposure estimate (derived from biomonitoring data or probabilistic modeling), the PoD is used to calculate a chemical-specific MoE (5), which is used to determine whether the compound should continue into tier two testing. Compounds that demonstrate a concerning MoA and relevant MoE move to tier two testing.

Tier two testing is comprised of short-term *in vivo* assays and, like tier one, has five components. (1) *In vivo* PoD values are derived for compounds with nonselective MoAs via BMD modeling of transcriptomic data from the eight tissues that are established as the primary sites of cancer in the two-year rodent bioassay (i.e., the liver, lung, kidney, stomach, mammary gland, urinary bladder, hematopoietic system, and vascular system, which account for majority of cancers in this assay); whereas, (2) *in vivo* PoD values of compounds with selective MoAs are derived using dose-response data from *in vivo* assays that are relevant to the MoA (transcriptomic or otherwise). (3) *In vivo* and *in vitro* pharmacokinetic studies are used to identify the discrepancy between the applied and the internal dose, as well as to identify/predict potential primary metabolites. (4) Tier one human exposure estimates are verified against any novel data. (5) MoE values are recalculated. Chemicals with MoEs below the cut-off continue to tier three testing; whereas, those with MoE values above the cut-off are not considered to pose sufficient risk to human health to warrant further testing. The authors estimate that additional testing could be avoided for 97% or 85% of chemicals by applying cut-offs of 100 or 1000, respectively. In the current study of furan, toxicogenomics data was used to produce BMD(L)s that were used to calculate transcriptional MoEs (Table 2.3). These MoE values were above 1000 and thus suggest that exposure to furan is not a risk to human health. In this framework, these MoE values would have excluded furan from tier

three testing; i.e., standard, chronic *in vivo* testing (e.g., the two-year rodent cancer bioassay). The authors of this framework do not propose any changes to the current long-term *in vivo* testing methodologies.

Together, this framework promotes efficient chemical testing and reduction of animal use, and is consistent with the “Toxicity Testing in the 21st Century” vision (Krewski *et al.* 2010, Settivari *et al.* 2015). It proposes an effective way to exclude chemicals from upper tier testing, which makes doing these tests more realistic. Finally, it builds a weight of evidence (WoE) that uses toxicogenomic data together with a large amount of data from short-term *in vitro* tests, such that informed, evidenced-based decisions can be confidently made regarding chemical toxicity (while avoiding the use of chronic *in vivo* assays). However, there are limitations to this paradigm. Interpretation of the tier one test results remains challenging; therefore, there is concern that triaging chemicals based on these tests alone might be error-prone. Further, the genetic homogeneity of tier two animals does not adequately reflect the genetic diversity in human populations. Using *in vitro* tests and animal models together with tests in human cell lines begins to address these issues. The parallelogram approach, reviewed in: Kienhuis *et al.* (2011), compares *in vitro* rodent data (top left corner of the parallelogram) to *in vivo* rodent data (top right) in order to identify animal-specific effects; *in vitro* rodent data to *in vitro* human data (bottom left) to identify inter-species effects (and *in vitro* artifacts); and makes predictions regarding the *in vivo* human response (bottom right). This approach has been applied to toxicogenomics studies of coumarin (Kienhuis *et al.* 2009a) and acetaminophen (Kienhuis *et al.* 2009b). The coumarin study was especially interesting because coumarin is highly toxic to rats, but

not to humans; therefore, the parallelogram approach was used to tease apart species-specific differences in chemical effect. Alternatively, the acetaminophen study produced similar results in both species. Another option is to use humanized animal models to test specific endpoints (e.g., the hCYP2E1 mouse (Cheung *et al.* 2005). Indeed, humanized mouse models for CYP1A1/CYP1A2, CYP2E1, CYP2D6, CYP3A4, CYP3A7, pregnane X receptor, and peroxisome proliferator-activated receptor alpha have been generated (Cheung & Gonzalez 2008), which could be quite useful in chemical risk assessment. Ultimately, the results produced in tiered assays must be extrapolatable to humans *in vivo*.

7.3.2 Guidelines for using Toxicogenomics Data in Formal Risk Assessment

The implementation of toxicogenomics in standard toxicity testing will require: (1) excellent communication between the scientists that generate the data and the risk assessors who apply it to decision-making, (2) a transition toward highly collaborative, multi-disciplinary teams that include biologists, bioinformaticians, and risk assessors, and (3) a clear and standardized approach for the production, analysis and interpretation of toxicogenomics datasets. To this end, Bourdon-Lacombe *et al.* (2015) recently published a comprehensive technical guide for evaluating and using gene expression data in human health risk assessment. The publication provides extensive practical guidance on interpreting toxicogenomic data, together with numerous published examples of high quality toxicogenomic studies. Despite the strong track record of technologies like qPCR and microarray (with RNA-seq catching up quickly), they report that a surprisingly low amount of these data have been used in formal chemical risk assessments. The authors

surveyed risk assessments that were completed from 2000-2013 by three groups: Health Canada's Existing Substances Risk Assessment Bureau (ESRAB), Health Canada's Guidelines for Canadian Drinking Water Quality program, and EPA's (IRIS) program; these groups used gene expression data in 2% (5/209), 0%, and 20% (12/59) of their assessments, respectively, with most of these within the last 5 years. The authors provide a strong rationale for the incorporation of genomics data into human health risk assessment, as well as the practical guidance as to how to proceed.

Bourdon-Lacombe *et al.* (2015) provide guidance on how to evaluate the quality of genomics data, which is important for establishing inclusion criteria for using genomics data (from qPCR, microarray, and RNA-seq) from both *in vitro* and *in vivo* experiments in risk assessments. In particular, they stress the importance of sound experimental design, including the use of appropriate controls, adequate sample sizes, and controlling for confounding effects (for example: the use of a randomized block design in a microarray experiment to control for time of experiment and array location). Next they list the criteria that are required for a well-done genomics study, including: the use of high quality RNA (RNA from FFPE samples is not considered in this manuscript), species appropriate microarrays or reference genomes, appropriate data normalization, statistical corrections for multiple comparisons, and setting appropriate thresholds for statistical comparisons. Most importantly, Bourdon-Lacombe *et al.* provide practical examples (from the published literature) for how to include genomics data in a risk assessment. Many of these strategies are employed in this thesis, for instance: (1) they describe how unsupervised hierarchical clustering and PCA can be used to compare individual samples, treatment groups, or studies, in order to gain a better understanding of

which data sets are similar and which are dissimilar; (2) they point out that molecular pathway, network, and upstream regulator enrichment (using software like DAVID and IPA) are useful approaches for gaining mechanistic understanding of a toxicant; (3) they describe how meta-analyses using other disease- and toxicant-specific datasets can be used to narrow down MoA; and, (4) they explain how the dose-response of genomics data can be used to identify PoD. An important aspect for establishing human health relevance of a toxicant that is discussed by Bourdon-Lacombe *et al.* (but not explored in the current thesis) is the comparison of gene expression profiles produced in animal and human cells, which is particularly useful when there is a lack of epidemiological data for a given toxicant. Because the guidelines by Bourdon-Lacombe and colleagues are comprehensive and written in accessible language, they are expected to be very important for facilitating the uptake of toxicogenomics data into human health risk assessment.

For implementation of the ‘Toxicity in the 21st Century’ vision, implementation of the 3Rs, and timely assessment of environmental hazards, changes are needed in the current chemical testing paradigm. Case studies such as the one presented in this thesis are important for providing evidence of the utility of toxicogenomics data in human health risk assessment. Moreover, the tiered testing strategy (discussed in 7.3.1) and technical guidance framework are important for the actual application of these data. To date, a number of toxicogenomics case studies have been completed with promising results; e.g.: (Ellinger-Ziegelbauer *et al.* 2009, Nesnow *et al.* 2009, Bourdon *et al.* 2012, Labib *et al.* 2012, Yao *et al.* 2012, Bourdon *et al.* 2013, Maertens *et al.* 2013). While it is always possible to complete further validation exercises, there do not appear to be any

clear advantages to delaying the inclusion of (high quality) genomics data in present-day chemical risk assessments.

7.4 Future Directions

7.4.1 FFPE Genomics

While genomic analysis of archival tumour biopsies has been done, performing toxicogenomic analysis of FFPE samples is still a very new (and somewhat controversial) idea. Tissue samples archived from past toxicity studies are very precious and difficult to obtain because they typically represent assays that would be difficult or impractical to repeat. However, because these samples are very well characterized, they are an extremely valuable resource for retrospective, phenotypically anchored toxicogenomics studies. While additional case studies should be done in order to confirm that transcriptional profiling of FFPE samples produces biologically accurate data, the results obtained in this thesis suggest that these samples can be used to develop predictive signatures that could be used to help characterize novel, data-poor chemicals (using a predictive toxicogenomics approach). The success of this work depends on storing the FFPE genomics datasets in a publically accessible format (such as a GEO superseries) so that all research groups conducting toxicogenomics experiments have access to the same set of standard chemical signatures (which, ideally, would include the 1500 cancer bioassay compounds). In addition, any corresponding apical data should also be made publically available. Importantly, the apical data need to be linked to the transcriptional data in a user-friendly way, to facilitate rapid analysis and decision-making. Altogether, archival samples from toxicity studies represent a massive opportunity to accelerate the

use of toxicogenomics in chemical risk assessment. Ultimately, while the use of the tissues for genomics analyses might seem controversial, its huge potential should not be overlooked.

During the course of this thesis, the field of genomics has moved extremely quickly. We began with a microarray-centric approach for conducting FFPE genomics (Chapter 5); however, we turned to next generation sequencing as its accessibility and affordability improved (Chapter 6). Moving forward, we have determined that the ribo-depletion RNA-seq approach is most amenable to degraded FFPE samples. Importantly, ribo-depletion avoids the issue of the degraded poly-A tail (which is problematic in poly-A enrichment RNA-seq) and it does not rely on pre-defined probe sequences (which is the case for microarrays). These attributes increase the likelihood of capturing as much of the transcriptome as possible, despite it being degraded. We do recommend, however, the addition of a technical replicate to the experiment in order to eliminate false positive genes, which appeared to be an issue in our experiments (Table 6.1), and expect that this approach would reduce the differences observed between paired samples. We assessed two variables that were likely to impact the quality of the genomics data: time in formalin and age of block. The former affected the quality of the data, whereas, surprisingly, the latter did not; these are important considerations for sample selection in future experiments. We expect that as next generation sequencing technologies continue to improve, so too will their capacity to deal with degraded, archival samples.

7.4.2 An Adverse Outcome Pathway for Cyp2E1 Substrates

The Organization for Economic Cooperation and Development (OECD) launched the Adverse Outcome Pathway (AOP) program in 2012, which is a truly collaborative effort that is crowd-sourcing experts from around the world to build step-by-step pathways that link a molecular initiating event (MIE) between a toxic substance and organism to an adverse outcome (AO). The AOPs are freely available online through a curated wiki (aopkb.org/aopwiki). This web-based format, which is more flexible than standard publications, allows each AOP to exist as a living document that can be edited and expanded as new information becomes available. As the AOP knowledgebase grows, AOPs are expected to become an important resource in chemical risk assessment.

An AOP is a chemically agnostic sequence of events that link a MIE to its AO through a series of key events (KE), which are sequentially linked by key event relationships (KERs). Importantly, an AOP is a simplified, linear pathway, not a detailed MoA. A key event must be measurable and essential to the pathway, but is not necessarily sufficient to cause the adverse outcome. A WoE is developed for each of these KEs and KERs using modified Bradford-Hill criteria that include: biological plausibility, empirical support, and essentiality (Becker *et al.* 2015). Further, KEs that are shared between AOPs are also linked, thereby forming large meta-networks of possible adverse outcomes (Knapen *et al.* 2015). The interactive nature of the AOPs and AOP networks is expected to facilitate the assessment of novel chemicals (for which there are limited data) by enabling researchers and risk assessors to fill-in-the-blanks by extrapolating from known chemical effects to possible adverse outcomes using the documented relationships found in existing AOPs. The AOP program also represents a

strategy toward reducing animal testing by directing hypotheses of chemically dependent toxicological effects, or preventing it all together (Landesmann *et al.* 2013, Burden *et al.* 2015a). Once a large enough number of AOPs is produced, they will be an important resource for chemical risk assessment.

Building a full complement of AOPs that fully populate the wiki will take some time for a number of reasons. The most obvious is that writing an AOP involves a thorough review of the literature, which is time consuming. What's more, it is a voluntary exercise that is typically outside of the AOP builder's normal job responsibilities. Further, there are very specific, non-standard formatting requirements that are related to the online structure of the wiki that are challenging to learn and follow (Villeneuve *et al.* 2014a, Villeneuve *et al.* 2014b). Finally, to ensure that the material that appears on the AOP wiki is of high quality, there is an extensive review process that must be satisfied. Therefore, it might be some time before these pathways become populated enough to be routinely used in risk assessment. In addition, a recent study of three AOPs demonstrated that their application in an applied setting is highly dependent on the data-richness of the AOP (Patlewicz *et al.* 2015). The AOP wiki currently lists 17 AOPs that are complete and open for commenting, and 64 AOPs that are under development. Published AOPs include a wide variety of etiologies, including: alkylation of DNA in male premeiotic germ cells leading to heritable mutations (Yauk *et al.* 2015), acetylcholinesterase inhibition leading to acute mortality (Russom *et al.* 2014), skin sensitization through covalent protein binding (MacKay *et al.* 2013), modulation of retinoic acid homeostasis leading to neural tube and axial defects (Tonk *et al.* 2014), and drug-mediated bile salt export pump inhibition leading to cholestatic liver injury (Vinken *et al.* 2013).

We have submitted a proposal that was accepted by the OECD to produce an AOP for Cyp2E1 activation leading to liver cancer (Fig. 7.2), which is the most imminent future direction for the work presented in this thesis. Briefly, the MIE is the binding of Cyp2E1 with a chemical substrate, which increases the Cyp2E1 protein half-life and produces reactive primary metabolites. KER1: uncoupling of the Cyp2E1 catalytic cycle, results in KE1: oxidative stress. KE2: cytotoxicity, is produced directly by KE1 and indirectly through primary metabolite-dependent cytotoxicity (iKER1). Cytotoxicity leads to the activation of a number of signaling pathways (KER3) that produce additional cytotoxicity, inflammation (iKER2), and regenerative proliferation (KE3). The result of dysregulated regenerative proliferation is liver cancer (the AO).

7.5 Concluding Remarks

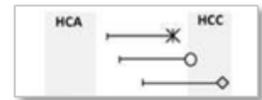
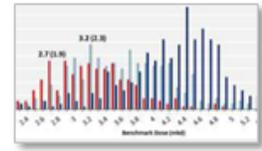
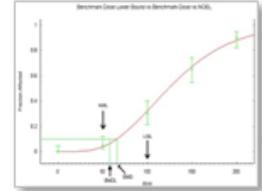
The work presented in this thesis represents an important step toward the routine use of toxicogenomics data in chemical risk assessment. Further, we explored concrete strategies toward the reduction of animal use in toxicity testing. While current approaches for chemical carcinogen identification are unsustainable, there remain legitimate technical and bureaucratic barriers impeding the transition toward newer methods. Since caution is an integral part of the risk assessment culture, strong toxicogenomics case studies that are clearly presented are crucial to building the trust required to shift existing paradigms. Altogether, toxicogenomics has enormous potential to improve and accelerate chemical risk assessment.

7.6 Figures

Figure 7.1 A summary of quantitative, predictive, and mechanistic toxicogenomics.

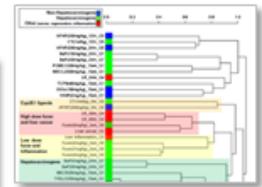
QUANTITATIVE TOXICOGENOMICS

1. Model the dose response of the number of DEGs (BMDS)
2. Model the dose response of the entire gene expression dataset (BMDExpress)
 - Use only the gene expression data that have responded to the treatment in a statistically significant manner (by applying an ANOVA or FDR filter)
 - Visualize the distribution of the gene and pathway BMD values
3. Choose a PoD based on the dose range where the majority of the endpoints depart from background
4. Calculate an MoE



PREDICTIVE TOXICOGENOMICS

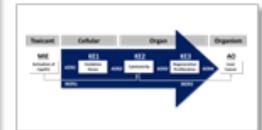
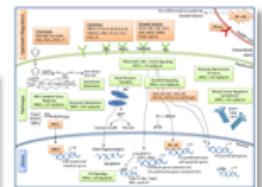
1. Download publically available gene expression datasets (from GEO) and perform hierarchical clustering with your dataset
2. Query the NextBio Pharmaco- and Disease- Atlases to identify diseases and chemicals that produce similar gene expression changes



Treatment	Close to Toxicity			
	QED	PKC	PKA	PKB
RNA-depletion RNA-seq				
Partial Hepatocellular Carcinoma	0.020427	1	1	2
Partial Hepatocellular Carcinoma	0.020427	2	2	4
Partial Hepatocellular Carcinoma	0.020427	3	3	5
CCl4 exposure	0.049302	4	4	1
CCl4 exposure	0.049302	5	5	6
Partial Hepatocellular Carcinoma	0.020427	6	6	8
Partial Hepatocellular Carcinoma	0.020427	7	7	1

MECHANISTIC TOXICOGENOMICS

1. Perform pathway and functional analyses using bioinformatics softwares (DAVID, IPA...) to develop a MoA for the compound
2. Query AOPs to form hypotheses on the possible adverse health effects of the test compound
3. Build human health relevance



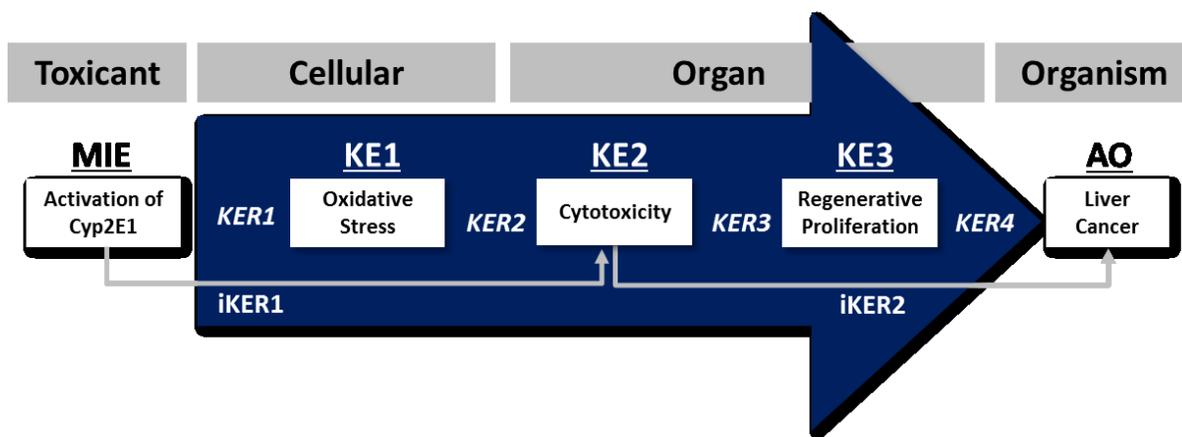


Figure 7.2 Adverse Outcome Pathway for Cyp2E1 Activation Leading to Liver Cancer.

References

- Abbas T & Dutta A** 2009 P21 in cancer: intricate networks and multiple activities. *Nature Reviews. Cancer* **9** 400-414.
- Abramovitz M, Ordanic-Kodani M, Wang Y, Li Z, Catzavelos C, Bouzyk M, W. SG, Jr, Moreno CS & Leyland-Jones B** 2008 Optimization of RNA extraction from FFPE tissues for expression profiling in the DASL assay. *Biotechniques* **44** 417-23.
- Alcorn JA, Feitelberg SP & Brenner DA** 1990 Transient induction of c-jun during hepatic regeneration. *Hepatology* **11** 909-915.
- Ameri K & Harris AL** 2008 Activating transcription factor 4. *The International Journal of Biochemistry & Cell Biology* **40** 14-21.
- Amir M, Liu K, Zhao E & Czaja MJ** 2012 Distinct functions of JNK and c-Jun in oxidant-induced hepatocyte death. *Journal of Cellular Biochemistry* **113** 3254-3265.
- Anders S, Pyl PT & Huber W** 2015 HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31** 166.
- Antica M, Paradzik M, Novak S, Dzebro S & Dominis M** 2010 Gene expression in formalin-fixed paraffin-embedded lymph nodes. *Journal of Immunological Methods* **359** 42-6.
- Arnér ESJ & Holmgren A** 2000 Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry* **267** 6102-6109.
- Ashkenazi A** 2002 Targeting death and decoy receptors of the tumour-necrosis factor superfamily. *Nature Review: Cancer* **2** 420-430.
- Auerbach SS, Phadke DP, Mav D, Holmgren S, Gao Y, Xie B, Shin JH, Shah RR, Merrick BA & Tice RR** 2014 RNA-Seq-based toxicogenomic assessment of fresh frozen and formalin-fixed tissues yields similar mechanistic insights. *Journal of Applied Toxicology* **35** 766-780.
- Auerbach SS, Shah RR, Mav D, Smith CS, Walker NJ, Vallant MK, Boorman GA & Irwin RD** 2010 Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. *Toxicology and Applied Pharmacology* **243** 300-314.
- Bass BP, Engel KB, Greytak SR & Moore HM** 2014 A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-

embedded (FFPE) tissue: how well do you know your FFPE specimen? *Archives of Pathology and Laboratory Medicine* **138** 1520-1530.

Becker RA, Chambers JM & Wilks AR 1998 The New S Language.

Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, Van Der Burg B, Villeneuve DL, Watanabe H & Barton-Maclaren TS 2015 Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. *Regulatory Toxicology and Pharmacology* **72** 514-537.

Behrens A, Sibilio M, David JP, Möhle-Steinlein U, Tronche F, Schütz G & Wagner EF 2002 Impaired postnatal hepatocyte proliferation and liver regeneration in mice lacking c-jun in the liver. *EMBO Journal* **21** 1782-1790.

Benford D, Bolger PM, Carthew P, Coulet M, DiNovi M, Leblanc JC, Renwick AG, Setzer W, Schlatter J, Smith B, Slob W, Williams G & Wildemann T 2010 Application of the Margin of Exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. *Food and Chemical Toxicology* **48** 2-24.

Benjamini Y & Hochberg Y 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 289-289-300.

Berasain C, Castillo J, Perugorria MJ, Latasa MU, Prieto J & Avila MA 2009 Inflammation and liver cancer: new molecular links. *Annals of the New York Academy of Sciences* **1155** 206-221.

Bhat VS, Hester SD, Nesnow S & Eastmond DA 2013 Concordance of transcriptional and apical benchmark dose levels for conazole-induced liver effects in mice. *Toxicological Sciences* **136** 205-215.

Black MB, Budinsky RA, Dombkowski A, Cukovic D, LeCluyse EL, Ferguson SS, Thomas RS & Rowlands JC 2012 Cross-species comparisons of transcriptomic alterations in human and rat primary hepatocytes exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicological Sciences* **127** 199-215.

Black MB, Parks BB, Pluta L, Chu TM, Allen BC, Wolfinger RD & Thomas RS 2014 Comparison of microarrays and RNA-seq for gene expression analyses of dose-response experiments. *Toxicological Sciences* **137** 385-403.

Bolstad BM, Irizarry RA, Åstrand M & Speed TP 2003 A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185-193.

Boothe DM, Dewey C & Carpenter DM 2012 Comparison of phenobarbital with bromide as a first-choice antiepileptic drug for treatment of epilepsy in dogs. *Journal of the American Veterinary Medical Association* **240** 1073-1083.

Bouchie A 2004 Coming soon: a global grid for cancer research. *Nature Biotechnology* **22** 1071-3.

Bourdon JA, Halappanavar S, Saber AT, Jacobsen NR, Williams A, Wallin H, Vogel U & Yauk CL 2012 Hepatic and pulmonary toxicogenomic profiles in mice intratracheally instilled with carbon black nanoparticles reveal pulmonary inflammation, acute phase response, and alterations in lipid homeostasis. *Toxicological Sciences* **127** 474-484.

Bourdon JA, Williams A, Kuo B, Moffat I, White PA, Halappanavar S, Vogel U, Wallin H & Yauk CL 2013 Gene expression profiling to identify potentially relevant disease outcomes and support human health risk assessment for carbon black nanoparticle exposure. *Toxicology* **303** 83-93.

Bourdon-Lacombe JA, Moffat ID, Deveau M, Husain M, Auerbach S, Krewski D, Thomas RS, Bushel PR, Williams A & Yauk CL 2015 Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals. *Regulatory Toxicology and Pharmacology* **72** 292-309.

Brunner AL, Beck AH, Edris B, Sweeney RT, Zhu SX, Li R, Montgomery K, Varma S, Gilks T, Guo X, Foley JW, Witten DM, Giacomini CP, Flynn RA, Pollack JR, Tibshirani R, Chang HY, van de Rijn M & West RB 2012 Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biology* **13** R75.

Budczies J, Weichert W, Noske A, Muller BM, Weller C, Wittenberger T, Hofmann HP, Dietel M, Denkert C & Gekeler V 2011 Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays. *Journal of Histochemistry and Cytochemistry* **59** 146-157.

Burden N, Sewell F, Andersen ME, Boobis A, Chipman JK, Cronin MT, Hutchinson TH, Kimber I & Whelan M 2015a Adverse Outcome Pathways can drive non-animal approaches for safety assessment. *Journal of Applied Toxicology* **35** 971-975.

Burden N, Sewell F & Chapman K 2015b Testing chemical safety: what is needed to ensure the widespread application of non-animal approaches? *PLoS Biology* **13**.

Burka LT, Washburn KD & Irwin RD 1991 Disposition of ¹⁴[C] furan in the male F344 rat. *Journal of Toxicology and Environmental Health* **34** 245-257.

Byrns MC, Predecki DP & Peterson LA 2002 Characterization of nucleoside adducts of cis-2-butene-1,4-dial, a reactive metabolite of furan. *Chemical Research in Toxicology* **15** 373-9.

Byrns MC, Vu CC, Neidigh JW, Abad JL, Jones RA & Peterson LA 2006 Detection of DNA adducts derived from the reactive metabolite of furan, cis-2-butene-1,4-dial. *Chemical Research in Toxicology* **19** 414-420.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A & Rinn JL 2011 Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development* **25** 1915-1927.

Call KM & Thilly WG 1991 5-Azacytidine inhibits the induction of transient TK-deficient cells by 5-bromodeoxyuridine. A novel hypothesis for the facilitation of hypermethylation by 5-bromodeoxyuridine. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **248** 101-114.

Carfagna MA, Held SD & Kedderis GL 1993 Furan-induced cytolethality in isolated rat hepatocytes: correspondence with in vivo dosimetry. *Toxicology and Applied Pharmacology* **123** 265-273.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest ARR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasaki Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SPT, Kruger AF, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schönbach C, Sekiguchi K, Semple CAM, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegene E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusica V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura

K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J & Hayashizaki Y 2005 The transcriptional landscape of the mammalian genome. *Science* **309** 1559-1563.

Carthew P, DiNovi M & Setzer RW 2010a Application of the margin of exposure (MoE) approach to substances in food that are genotoxic and carcinogenic: example: furan (CAS No. 110-00-9). *Food and Chemical Toxicology* **48** S69-74.

Carthew P, DiNovi M & Woodrow Setzer R 2010b Application of the Margin of Exposure (MOE) approach to substances in food that are genotoxic and carcinogenic: example: CAS No: 105650-23-5 PhIP (2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine). *Food and Chemical Toxicology* **48** S98-105.

Cederbaum A 2009 Nrf2 and antioxidant defense against CYP2E1 toxicity. *Expert opinion on drug metabolism & toxicology* **5** 1223-1244.

Chaisson ML, Brooling JT, Ladiges W, Tsai S & Fausto N 2002 Hepatocyte-specific inhibition of NF- κ B leads to apoptosis after TNF treatment, but not after partial hepatectomy. *Journal of Clinical Investigation* **110** 193-202.

Chen LJ, Hecht SS & Peterson LA 1997 Characterization of amino acid and glutathione adducts of cis-2-butene- 1,4-dial, a reactive metabolite of furan. *Chemical Research in Toxicology* **10** 866-874.

Chen Q & Cederbaum AI 1998 Cytotoxicity and apoptosis produced by cytochrome P450 2E1 in Hep G2 cells. *Molecular pharmacology* **53** 638-648.

Cheung C & Gonzalez FJ 2008 Humanized mouse lines and their application for prediction of human drug metabolism and toxicological risk assessment. *The Journal of Pharmacology and Experimental Therapeutics* **327** 288-299.

Cheung C, Yu A-, Ward JM, Krausz KW, Akiyama TE, Feigenbaum L & Gonzalez FJ 2005 The CYP2E1-humanized transgenic mouse: Role of CYP2E1 in acetaminophen hepatotoxicity. *Drug Metabolism and Disposition* **33** 449-457.

Chu WM 2013 Tumor necrosis factor. *Cancer Letters* **328** 222-225.

Churchwell MI, Scheri RC, Von Tungeln LS, Gamboa da Costa G, Beland FA & Doerge DR 2015 Evaluation of Serum and Liver Toxicokinetics for Furan and Liver DNA Adduct Formation in Male Fischer 344 Rats. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*.

- Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris KV, Morillon A, Rozowsky JS, Gerstein MB, Wahlestedt C, Hayashizaki Y, Carninci P, Gingeras TR & Mattick JS** 2011 The reality of pervasive transcription. *PLoS Biology* **9**.
- Clewell HJ, Efremenko A, Campbell JL, Dodd DE & Thomas RS** 2014 Transcriptional responses in the rat nasal epithelium following subchronic inhalation of naphthalene vapor. *Toxicology and Applied Pharmacology* **280** 78-85.
- Cressman DE, Greenbaum LE, Haber BA & Taub R** 1994 Rapid activation of post-hepatectomy factor/nuclear factor κ B in hepatocytes, a primary response in the regenerating liver. *Journal of Biological Chemistry* **269** 30429-30435.
- Crews C, Hasnip S, Roberts DPT & Castle L** 2007 Factors affecting the analysis of furan in heated foods. *Food Additives and Contaminants* **24** 108-113.
- Crews C, Roberts D, Laurysen S & Kramer G** 2009 Survey of furan in foods and coffees from five European Union countries. *Food additives & contaminants. Part B, Surveillance* **2** 95-98.
- Cui X, Hwang JT, Qiu J, Blades NJ & Churchill GA** 2005 Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* **6** 59-75.
- Cui Y & Paules RS** 2010 Use of transcriptomics in understanding mechanisms of drug-induced toxicity. *Pharmacogenomics* **11** 573-585.
- Dang CV** 2012 MYC on the path to cancer. *Cell* **149** 22-35.
- Davis JA, Gift JS & Zhao QJ** 2011 Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1. *Toxicology and Applied Pharmacology* **254** 181-191.
- Denicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, Mangal D, Yu KH, Yeo CJ, Calhoun ES, Scrimieri F, Winter JM, Hruban RH, Iacobuzio-Donahue C, Kern SE, Blair IA & Tuveson DA** 2011 Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475** 106-110.
- Di Cesare S, Nantel A, Marshall JC, Fernandes BF, Anteckka E, Orellana ME, Abourbih D, Saornil AM & N. BM, Jr** 2010 Expression profiling of formalin-fixed paraffin embedded primary human uveal melanomas using DASL matrices. *Journal of Cancer Research and Clinical Oncology* **136** 577-586.
- DiDonato JA, Mercurio F & Karin M** 2012 NF-kappaB and the link between inflammation and cancer. *Immunological reviews* **246** 379-400.

Diehl AM, Yin M, Fleckenstein J, Yang SQ, Lin HZ, Brenner DA, Westwick J, Bagby G & Nelson S 1994 Tumor necrosis factor-alpha induces c-jun during the regenerative response to liver injury. *The American Journal of Physiology* **267** G552-61.

Ding W, Petibone DM, Latendresse JR, Pearce MG, Muskhelishvili L, White GA, Chang C-, Mittelstaedt RA, Shaddock JG, McDaniel LP, Doerge DR, Morris SM, Bishop ME, Manjanatha MG, Aidoo A & Heflich RH 2012 In vivo genotoxicity of furan in F344 rats at cancer bioassay doses. *Toxicology and Applied Pharmacology* **261** 164-171.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW & Kavlock RJ 2007 The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicological Sciences* **95** 5-12.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See L-, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guig R & Gingeras TR 2012 Landscape of transcription in human cells. *Nature* **489** 101-108.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR 2013 STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29** 15-21.

Dunnnett CW 1955 A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association* **50** 1096-1096-1121.

Durling LJ, Svensson K & Abramsson-Zetterberg L 2007 Furan is not genotoxic in the micronucleus assay in vivo or in vitro. *Toxicology Letters* **169** 43-50.

Eferl R, Sibilina M, Hilberg F, Fuchsbichler A, Kufferath I, Guertl B, Zenz R, Wagner EF & Zatloukal K 1999 Functions of c-Jun in liver and heart development. *The Journal of Cell Biology* **145** 1049-1061.

Eferl R & Wagner EF 2003 AP-1: A double-edged sword in tumorigenesis. *Nature Reviews Cancer* **3** 859-868.

EFSA 2005 EFSA - Scientific Opinion of the Scientific Committee: Opinion of the Scientific Committee on a request from EFSA related to A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic. **2013**.

Ellinger-Ziegelbauer H, Aubrecht J, Kleinjans JC & Ahr HJ 2009 Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicology Letters* **186** 36-44.

Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A & Ahr HJ 2008 Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutation Research* **637** 23-39.

Endoh M, Miki K, Hossain MN, Fujii M & Ayusawa D 2007 5-Bromodeoxyuridine increases transient expression of ectopic genes in human cells. *Bioscience, Biotechnology and Biochemistry* **71** 1098-1102.

Farragher SM, Tanney A, Kennedy RD & Paul Harkin D 2008 RNA expression analysis from formalin fixed paraffin embedded tissues. *Histochemistry and Cell Biology* **130** 435-45.

Fedorowicz G, Guerrero S, Wu TD & Modrusan Z 2009 Microarray analysis of RNA extracted from formalin-fixed, paraffin-embedded and matched fresh-frozen ovarian adenocarcinomas. *BMC Medical Genomics* **2** 23.

FitzGerald MJ, Webber EM, Donovan JR & Fausto N 1995 Rapid DNA binding by nuclear factor κ B in hepatocytes at the start of liver regeneration. *Cell Growth and Differentiation* **6** 417-427.

Fransson-Steen R, Goldsworthy TL, Kedderis GL & Maronpot RR 1997 Furan-induced liver cell proliferation and apoptosis in female B6C3F1 mice. *Toxicology* **118** 195-204.

Gan L, Guo K, Li Y, Kang X, Sun L, Shu H & Liu Y 2010 Up-regulated expression of stathmin may be associated with hepatocarcinogenesis. *Oncology Reports* **23** 1037-1043.

Gates LA, Lu D & Peterson LA 2012 Trapping of cis-2-butene-1,4-dial to measure furan metabolism in human liver microsomes by cytochrome P450 enzymes. *Drug Metabolism and Disposition* **40** 596-601.

Gates LA, Phillips MB, Matter BA & Peterson LA 2014 Comparative metabolism of furan in rodent and human cryopreserved hepatocytes. *Drug metabolism and disposition: the biological fate of chemicals* **42** 1132-1136.

Gautier L, Cope L, Bolstad BM & Irizarry RA 2004 Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20** 307-315.

Gill S, Kavanagh M, Barker M, Weld M, Vavasour E, Hou Y & Cooke GM 2011 Subchronic Oral Toxicity Study of Furan in B6C3F1 Mice. *Toxicologic Pathology* **39** 787-794.

Gonzalez FJ 2005 Role of cytochromes P450 in chemical toxicity and oxidative stress: studies with CYP2E1. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **569** 101-110.

Goodman MF, Hopkins RL, Lasken R & Mhaskar DN 1985 The biochemical basis of 5-bromouracil- and 2-aminopurine-induced mutagenesis. *Basic Life Sciences* **31** 409-423.

Goodnight JH & Harvey WR 1978 Least-Squares Means in the Fixed-Effects General Linear Models. **R-103**.

Gutschner T & Diederichs S 2012 The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biology* **9** 703-719.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL & Lander ES 2009 Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458** 223-227.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE & Lander ES 2011 LincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477** 295-300.

Guttman M & Rinn JL 2012 Modular regulatory principles of large non-coding RNAs. *Nature* **482** 339-346.

Hai T, Wolfgang CD, Marsee DK, Allen AE & Sivaprasad U 1999 ATF3 and stress responses. *Gene Expression* **7** 321-335.

Han ZG 2012 Functional genomic studies: insights into the pathogenesis of liver cancer. *Annual Review of Genomics and Human Genetics* **13** 171-205.

Hanahan D & Weinberg RA 2000 The hallmarks of cancer. *Cell* **100** 57-70.

Hanahan D & Weinberg RA 2011 Hallmarks of cancer: the next generation. *Cell* **144** 646-674.

Harrill AH, Ross PK, Gatti DM, Threadgill DW & Rusyn I 2009 Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. *Toxicological Sciences* **110** 235-243.

Hasnip S, Crews C & Castle L 2006 Some factors affecting the formation of furan in heated foods. *Food Additives and Contaminants* **23** 219-227.

Hayes JD & McMahon M 2006 The double-edged sword of Nrf2: Subversion of redox homeostasis during the evolution of cancer. *Molecular Cell* **21** 732-734.

Hedegaard J, Thorsen K, Lund MK, Hein A-K, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtröder K, Kruhøffer M, Hager H, Knudsen B, Andersen CL, Sørensen KD, Pedersen JS, Ørntoft TF & Dyrskjøt L 2014 Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS ONE* **9**.

Hess J, Angel P & Schorpp-Kistner M 2004 AP-1 subunits: quarrel and harmony among siblings. *Journal of Cell Science* **117** 5965-5973.

Hickling KC, Hitchcock JM, Oreffo V, Mally A, Hammond TG, Evans JG & Chipman JK 2010 Evidence of oxidative stress and associated DNA damage, increased proliferative drive, and altered gene expression in rat liver produced by the cholangiocarcinogenic agent Furan. *Toxicologic Pathology* **38** 230-243.

Hilberg F, Aguzzi A, Howells N & Wagner EF 1993 c-Jun is essential for normal mouse development and hepatogenesis. *Nature* **365** 179-181.

Holsapple MP, Pitot HC, Cohen SH, Boobis AR, Klaunig JE, Pastoor T, Dellarco VL & Dragan YP 2006 Mode of action in relevance of rodent liver tumors to human cancer risk. *Toxicological Sciences* **89** 51-56.

Honkakoski P, Zelko I, Sueyoshi T & Negishi M 1998 The nuclear orphan receptor CAR-retinoid X receptor heterodimer activates the phenobarbital-responsive enhancer module of the CYP2B gene. *Molecular and Cellular Biology* **18** 5652-8.

Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, Moore J, Wrobel MJ, Lerner J, Reich M, Chan JA, Glickman JN, Ikeda K, Hashimoto M, Watanabe G, Daidone MG, Roayaie S, Schwartz M, Thung S, Salvesen HB, Gabriel S, Mazzaferro V, Bruix J, Friedman SL, Kumada H, Llovet JM & Golub TR 2008 Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *New England Journal of Medicine* **359** 1995-2004.

Hothorn T, Bretz F & Westfall P 2008 Simultaneous inference in general parametric models. *Biometrical Journal* **50** 346-363.

Hsieh SY, Huang SF, Yu MC, Yeh TS, Chen TC, Lin YJ, Chang CJ, Sung CM, Lee YL & Hsu CY 2010 Stathmin1 overexpression associated with polyploidy, tumor-cell invasion, early recurrence, and poor prognosis in human hepatoma. *Molecular Carcinogenesis* **49** 476-487.

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T & Rinn JL 2010 A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142** 409-419.

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerød A, Børresen-Dale A-, Kim SK, Van De Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ & Chang HY 2011 Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genetics* **43** 621-629.

IARC 1995 Dry cleaning: Some chlorinated solvents and other industrial chemicals. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* **63** 33-475.

IARC 2001 Some Thyrotropic Agents: Phenobarbital and its Sodium Salt. **79** 161.

Ihunnah CA, Jiang M & Xie W 2011 Nuclear receptor PXR, transcriptional circuits and metabolic relevance. *Biochimica et Biophysica Acta* **1812** 956-63.

IRIS 2012 Furan (CASRN 110-00-9); <http://www.epa.gov/iris/subst/0056.htm>.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U & Speed TP 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4** 249-264.

Jackson AF, Williams A, Moffat I, Phillips SL, Recio L, Waters MD, Lambert IB & Yauk CL 2013 Preparation of archival formalin-fixed paraffin-embedded mouse liver samples for use with the Agilent gene expression microarray platform. *Journal of Pharmacological and Toxicological Methods*.

Jackson AF, Williams A, Recio L, Waters MD, Lambert IB & Yauk CL 2014 Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan. *Toxicology and Applied Pharmacology* **274** 63-77.

Jacobson TA, Lundahl J, Mellstedt H & Moshfegh A 2011 Gene expression analysis using long-term preserved formalin-fixed and paraffin-embedded tissue of non-small cell lung cancer. *International Journal of Oncology* **38** 1075-1081.

Jaeschke H, Gores GJ, Cederbaum AI, Hinson JA, Pessayre D & Lemasters JJ 2002 Mechanisms of hepatotoxicity. *Toxicological Sciences* **65** 166-176.

Jaeschke H, Xie Y & McGill MR 2014 Acetaminophen-induced Liver Injury: from Animal Models to Humans. *Journal of clinical and translational hepatology* **2** 153-161.

Jalali S, Jayaraj GG & Scaria V 2012 Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. *Biology Direct* **7**.

James W & Stein C 1961 Estimation with Quadratic Loss. **1** 361-379.

Jemal A, Bray F, Center MM, Ferlay J, Ward E & Forman D 2011 Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61** 69-90.

Jochum W, Passequé E & Wagner EF 2001 AP-1 in mouse development and tumorigenesis. *Oncogene* **20** 2401-2412.

Judson R, Houck K, Martin M, Knudsen T, Thomas RS, Sipes N, Shah I, Wambaugh J & Crofton K 2014 In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast programme. *Basic and Clinical Pharmacology and Toxicology* **115** 69-76.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM, Rotroff DM, Shah I, Richard AM & Dix DJ 2010 In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. *Environmental Health Perspectives* **118** 485-492.

Kang JS, Wanibuchi H, Morimura K, Wongpoomchai R, Chusiri Y, Gonzalez FJ & Fukushima S 2008 Role of CYP2E1 in thioacetamide-induced mouse hepatotoxicity. *Toxicology and Applied Pharmacology* **228** 295-300.

Karin M 2006 Nuclear factor-kappaB in cancer development and progression. *Nature* **441** 431-436.

Karin M 2009 NF-kappaB as a critical link between inflammation and cancer. *Cold Spring Harbor Perspectives in Biology* **1**.

Karin M & Shaulian E 2001 AP-1: linking hydrogen peroxide and oxidative stress to the control of cell proliferation and death. *IUBMB life* **52** 17-24.

Kedderis GL, Carfagna MA, Held SD, Batra R, Murphy JE & Gargas ML 1993 Kinetic analysis of furan biotransformation by F-344 rats in vivo and in vitro. *Toxicology and Applied Pharmacology* **123** 274-282.

Kellert M, Wagner S, Lutz U & Lutz WK 2008 Biomarkers of furan exposure by metabolic profiling of rat urine with liquid chromatography-tandem mass spectrometry and principal component analysis. *Chemical Research in Toxicology* **21** 761-8.

Kensler TW, Roebuck BD, Wogan GN & Groopman JD 2011 Aflatoxin: a 50-year odyssey of mechanistic and translational toxicology. *Toxicological sciences : an official journal of the Society of Toxicology* **120 Suppl 1** S28-48.

Kerr MK 2003 Design considerations for efficient and effective microarray studies. *Biometrics* **59** 822-828.

Kew MC 2013 Aflatoxins as a cause of hepatocellular carcinoma. *Journal of gastrointestinal and liver diseases : JGLD* **22** 305-310.

Khalaf N, White D, Kanwal F, Ramsey D, Mittal S, Tavakoli-Tabasi S, Kuzniarek J & El-Serag HB 2015 Coffee and Caffeine Are Associated With Decreased Risk of Advanced Hepatic Fibrosis Among Patients With Hepatitis C. *Clinical gastroenterology and hepatology* **13** 1521.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Morales DR, Thomas K, Presser A, Bernstein BE, Van Oudenaarden A, Regev A, Lander ES & Rinn JL 2009 Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106** 11667-11672.

Kienhuis AS, Bessems JG, Pennings JL, Driessen M, Luijten M, van Delft JH, Peijnenburg AA & van der Ven LT 2011 Application of toxicogenomics in hepatic systems toxicology for risk assessment: acetaminophen as a case study. *Toxicology and Applied Pharmacology* **250** 96-107.

Kienhuis AS, van de Poll MC, Dejong CH, Gottschalk R, van Herwijnen M, Boorsma A, Kleinjans JC, Stierum RH & van Delft JH 2009a A toxicogenomics-based parallelogram approach to evaluate the relevance of coumarin-induced responses in primary human hepatocytes in vitro for humans in vivo. *Toxicology In Vitro* **23** 1163-1169.

Kienhuis AS, van de Poll MC, Wortelboer H, van Herwijnen M, Gottschalk R, Dejong CH, Boorsma A, Paules RS, Kleinjans JC, Stierum RH & van Delft JH 2009b Parallelogram approach using rat-human in vitro and rat in vivo toxicogenomics predicts acetaminophen-induced hepatotoxicity in humans. *Toxicological Sciences* **107** 544-552.

Kim SK, Suh MR, Yoon HS, Lee JB, Oh SK, Moon SY, Moon S-, Lee JY, Hwang JH, Cho WJ & Kim K- 2005 Identification of developmental pluripotency associated 5 expression in human pluripotent stem cells. *Stem Cells* **23** 458-462.

Kleinstreuer NC, Yang J, Berg EL, Knudsen TB, Richard AM, Martin MT, Reif DM, Judson RS, Polokoff M, Dix DJ, Kavlock RJ & Houck KA 2014 Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nature Biotechnology* **32** 583-591.

Klopfleisch R, Weiss AT & Gruber AD 2011 Excavation of a buried treasure--DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histology and Histopathology* **26** 797-810.

Knapen D, Vergauwen L, Villeneuve DL & Ankley GT 2015 The potential of AOP networks for reproductive and developmental toxicity assay development. *Reproductive Toxicology* **56** 52-55.

Knight TR, Choudhuri S & Klaassen CD 2007 Constitutive mRNA expression of various glutathione S-transferase isoforms in different tissues of mice. *Toxicological Sciences* **100** 513-524.

Knudsen TB, Houck KA, Sipes NS, Singh AV, Judson RS, Martin MT, Weissman A, Kleinstreuer NC, Mortensen HM, Reif DM, Rabinowitz JR, Setzer RW, Richard AM, Dix DJ & Kavlock RJ 2011 Activity profiles of 309 ToxCast chemicals evaluated across 292 biochemical targets. *Toxicology* **282** 1-15.

Konno Y, Kamino H, Moore R, Lih F, Tomer KB, Zeldin DC, Goldstein JA & Negishi M 2010 The nuclear receptors constitutive active/androstane receptor and pregnane x receptor activate the Cyp2c55 gene in mouse liver. *Drug Metabolism and Disposition* **38** 1177-1182.

Krewski D, D. A., Jr, Andersen M, Anderson H, Bailar JC, 3rd, Boekelheide K, Brent R, Charnley G, Cheung VG, S. G., Jr, Kelsey KT, Kerkvliet NI, Li AA, McCray L, Meyer O, Patterson RD, Pennie W, Scala RA, Solomon GM, Stephens M, Yager J & Zeise L 2010 Toxicity testing in the 21st century: a vision and a strategy. *Journal of Toxicology and Environmental Health. Part B, Critical reviews* **13** 51-138.

Kupersmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, Park J, Cui W, Wall GD, Wisotzkey R, Alag S, Akhtari S & Ronaghi M 2010 Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS ONE* **5**.

Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, Odom DT & Marques AC 2012 Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genetics* **8**.

Labib S, Yauk C, Williams A, Arlt VM, Phillips DH, White PA & Halappanavar S 2012 Subchronic oral exposure to benzo(a)pyrene leads to distinct transcriptomic changes in the lungs that are related to carcinogenesis. *Toxicological Sciences* **129** 213-224.

Landesmann B, Mennecozi M, Berggren E & Whelan M 2013 Adverse outcome pathway-based screening strategies for an animal-free safety assessment of chemicals. *Alternatives to Laboratory Animals* **41** 461-471.

Larson JL, Wolf DC & Butterworth BE 1994 Induced cytolethality and regenerative cell proliferation in the livers and kidneys of male B6C3F1 mice given chloroform by gavage. *Fundamentals of Applied Toxicology* **23** 537-43.

Lasken RS & Goodman MF 1984 The biochemical basis of 5-bromouracil-induced mutagenesis. Heteroduplex base mispairs involving bromouracil in G x C----A x T and A x T----G x C mutational pathways. *The Journal of biological Chemistry* **259** 11491-11495.

Lassmann S, Kreutz C, Schoepflin A, Hopt U, Timmer J & Werner M 2009 A novel approach for reliable microarray analysis of microdissected tumor cells from formalin-fixed and paraffin-embedded colorectal cancer resection specimens. *Journal of Molecular Medicine* **87** 211-24.

Lau A, Villeneuve NF, Sun Z, Wong PK & Zhang DD 2008 Dual roles of Nrf2 in cancer. *Pharmacological Research* **58** 262-270.

Law CW, Chen Y, Shi W & Smyth GK 2014 voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15** R29-2014-15-2-r29.

Lempiäinen H, Couttet P, Bolognani F, Müller A, Dubost V, Luisier R, Espinola ADR, Vitry V, Unterberger EB, Thomson JP, Treindl F, Metzger U, Wrzodek C, Hahne F, Zollinger T, Brasa S, Kalteis M, Marcellin M, Giudicelli F, Braeuning A, Morawiec L, Zamurovic N, Längle U, Scheer N, Schübeler D, Goodman J, Chibout S, Marlowe J, Theil D, Heard DJ, Grenet O, Zell A, Templin MF, Meehan RR, Wolf RC, Elcombe CR, Schwarz M, Moulin P, Terranova R & Moggs JG 2013 Identification of Dlk1-Dio3 imprinted gene cluster noncoding RNAs as novel candidate biomarkers for liver tumor promotion. *Toxicological Sciences* **131** 375-386.

Leopardi P, Cordelli E, Villani P, Cremona TP, Conti L, De Luca G & Crebelli R 2010 Assessment of in vivo genotoxicity of the rodent carcinogen furan: evaluation of DNA damage and induction of micronuclei in mouse splenocytes. *Mutagenesis* **25** 57-62.

Li Q, Van Antwerp D, Mercurio F, Lee KF & Verma IM 1999 Severe liver degeneration in mice lacking the IkappaB kinase 2 gene. *Science* **284** 321-325.

Li S, Labaj PP, Zumbo P, Sykacek P, Shi W, Shi L, Phan J, Wu PY, Wang M, Wang C, Thierry-Mieg D, Thierry-Mieg J, Kreil DP & Mason CE 2014a Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nature Biotechnology* **32** 888-895.

Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, Viale A, Wright C, Schweitzer PA, Gao Y, Kim D, Boland J, Hicks B, Kim R, Chhangawala S, Jafari N, Raghavachari N, Gandara J, Garcia-Reyero N, Hendrickson C, Roberson D, Rosenfeld J, Smith T, Underwood JG, Wang M, Zumbo P, Baldwin DA, Grills GS & Mason CE 2014b Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nature biotechnology* **32** 915-925.

Liao Y, Smyth GK & Shi W 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30** 923-930.

Ligges U & Mächler M 2003 Scatterplot3d - An R package for visualizing multivariate data. *Journal of Statistical Software* **8** 1-20.

Lin ZY, Dechesne CA, Eldridge J & Paterson BM 1989 An avian muscle factor related to MyoD1 activates muscle-specific promoters in nonmuscle cells of different germ-layer origin and in BrdU-treated myoblasts. *Genes and Development* **3** 986-996.

Linton K, Hey Y, Dibben S, Miller C, Freemont A, Radford J & Pepper S 2009 Methods comparison for high-resolution transcriptional analysis of archival material on Affymetrix Plus 2.0 and Exon 1.0 microarrays. *Biotechniques* **47** 587-96.

Linton K, Howarth C, Wappett M, Newton G, Lachel C, Iqbal J, Pepper S, Byers R, Chan WJ & Radford J 2012 Microarray gene expression analysis of fixed archival tissue permits molecular classification and identification of potential therapeutic targets in diffuse large B-cell lymphoma. *The Journal of Molecular Diagnostics* **14** 223-232.

Liu H, Jones BE, Bradham C & Czaja MJ 2002 Increased cytochrome P-450 2E1 expression sensitizes hepatocytes to c-Jun-mediated cell death from TNF-alpha. *American Journal of Physiology. Gastrointestinal and Liver Physiology* **282** G257-66.

Liu J, Mansouri K, Judson RS, Martin MT, Hong H, Chen M, Xu X, Thomas RS & Shah I 2015 Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. *Chemical Research in Toxicology* **28** 738-751.

Lu Y & Cederbaum AI 2008 CYP2E1 and oxidative liver injury by alcohol. *Free Radical Biology and Medicine* **44** 723-738.

Lu Y & Cederbaum AI 2010 CYP2E1 potentiation of LPS and TNFalpha-induced hepatotoxicity by mechanisms involving enhanced oxidative and nitrosative stress, activation of MAP kinases, and mitochondrial dysfunction. *Genes and Nutrition* **5** 149-167.

Luedde T & Schwabe RF 2011 NF-κB in the liver-linking injury, fibrosis and hepatocellular carcinoma. *Nature Reviews Gastroenterology and Hepatology* **8** 108-118.

Ma Q & He X 2012 Molecular basis of electrophilic and oxidative defense: promises and perils of Nrf2. *Pharmacological Reviews* **64** 1055-1081.

Ma X, Ke T, Mao P, Jin X, Ma L & He G 2008 The mutagenic properties of BrdUTP in a random mutagenesis process. *Molecular Biology Reports* **35** 663-667.

MacKay C, Davies M, Summerfield V & Maxwell G 2013 From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. *Altex* **30** 473-486.

Maeda S, Kamata H, Luo JL, Leffert H & Karin M 2005 IKK β couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. *Cell* **121** 977-990.

Maertens RM, White PA, Williams A & Yauk CL 2013 A global toxicogenomic analysis investigating the mechanistic differences between tobacco and marijuana smoke condensates in vitro. *Toxicology* **308** 60-73.

Maga JA & Katz I 1979 Furans in foods. *CRC Critical Reviews in Food Science and Nutrition* **11** 355-400.

Malik AI, Williams A, Lemieux CL, White PA & Yauk CL 2012 Hepatic mRNA, microRNA, and miR-34a-target responses in mice after 28 days exposure to doses of benzo(a)pyrene that elicit DNA damage and mutation. *Environmental and molecular mutagenesis* **53** 10-21.

Malone JH & Oliver B 2011 Microarrays, deep sequencing and the true measure of the transcriptome. *BMC biology* **9** 34.

Manibusan MK, Odin M & Eastmond DA 2007 Postulated carbon tetrachloride mode of action: A review. *Journal of Environmental Science and Health - Part C Environmental Carcinogenesis and Ecotoxicology Reviews* **25** 185-209.

Mardia KV, Kent JT & Bibby JM 1979 Multivariate Analysis.

Marioni JC, Mason CE, Mane SM, Stephens M & Gilad Y 2008 RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* **18** 1509-1517.

McDaniel LP, Ding W, Dobrovolsky VN, G. SJ, Jr, Mittelstaedt RA, Doerge DR & Heflich RH 2012 Genotoxicity of furan in Big Blue rats. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **742** 72-8.

Meleady P & Clynes M 2001 Bromodeoxyuridine induces integrin expression at transcriptional ($\alpha 2$ subunit) and post-transcriptional ($\beta 1$ subunit) levels, and alters the adhesive properties of two human lung tumour cell lines. *Cell Adhesion and Communication* **8** 45-59.

Meloche S & Pouyssegur J 2007 The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. *Oncogene* **26** 3227-3239.

Meng W, McElroy JP, Volinia S, Palatini J, Warner S, Ayers LW, Palanichamy K, Chakravarti A & Lautenschlaeger T 2013 Comparison of microRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. *PloS ONE* **8**.

Merrick BA, Auerbach SS, Stockton PS, Foley JF, Malarkey DE, Sills RC, Irwin RD & Tice RR 2012 Testing an aflatoxin B1 gene signature in rat archival tissues. *Chemical Research in Toxicology* **25** 1132-1144.

Min L, He B & Hui L 2011 Mitogen-activated protein kinases in hepatocellular carcinoma development. *Seminars in cancer biology* **21** 10-20.

Min L, Ji Y, Bakiri L, Qiu Z, Cen J, Chen X, Chen L, Scheuch H, Zheng H, Qin L, Zatloukal K, Hui L & Wagner EF 2012 Liver cancer initiation is controlled by AP-1 through SIRT6-dependent inhibition of survivin. *Nature cell biology* **14** 1203-1211.

Moffat I, Chepelev NL, Labib S, Bourdon-Lacombe J, Kuo B, Buick JK, Lemieux F, Williams A, Halappanavar S, Malik AI, Luijten M, Aubrecht J, Hyduke DR, Fornace AJ, Jr., Swartz CD, Recio L & Yauk CL 2015 Comparison of toxicogenomics and traditional approaches to inform mode of action and points of departure in human health risk assessment of benzo[a]pyrene in drinking water. *Critical Reviews in Toxicology* **45** 1-43.

Mohammad F, Pandey GK, Mondal T, Enroth S, Redrup L, Gyllensten U & Kanduri C 2012 Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development* **139** 2792-2803.

Moran VA, Perera RJ & Khalil AM 2012 Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic acids research* **40** 6391-6400.

Morisco F, Lembo V, Mazzone G, Camera S & Caporaso N 2014 Coffee and liver health. *Journal of clinical gastroenterology* **48** Suppl 1 S87-90.

Moro S, Chipman JK, Wegener JW, Hamberger C, Dekant W & Mally A 2012 Furan in heat-treated foods: Formation, exposure, toxicity, and aspects of risk assessment. *Molecular Nutrition and Food Research* **56** 1197-1211.

Moser GJ, Foley J, Burnett M, Goldsworthy TL & Maronpot R 2009 Furan-induced dose-response relationships for liver cytotoxicity, cell proliferation, and tumorigenicity (furan-induced liver tumorigenicity). *Experimental and Toxicologic Pathology* **61** 101-111.

Nakagawa H & Maeda S 2012 Molecular mechanisms of liver injury and hepatocarcinogenesis: focusing on the role of stress-activated MAPK. *Pathology Research International*.

Nakano H, Nakajima A, Sakon-Komazawa S, Piao JH, Xue X & Okumura K 2006 Reactive oxygen species mediate crosstalk between NF-kappaB and JNK. *Cell death and differentiation* **13** 730-737.

- Nesnow S, Ward W, Moore T, Ren H & Hester SD** 2009 Discrimination of tumorigenic triazole conazoles from phenobarbital by transcriptional analyses of mouse liver gene expression. *Toxicological Sciences* **110** 68-83.
- Nie L, Wu H-, Hsu J-, Chang S-, LaBaff AM, Li C-, Wang Y, Hsu JL & Hung M-** 2012 Long non-coding RNAs: Versatile master regulators of gene expression and crucial players in cancer. *American Journal of Translational Research* **4** 127-150.
- NTP** 1993 NTP Technical Report on the Toxicology and Carcinogenesis Studies of Furan in F344/N Rats and B6C3F1 Mice (Gavage Studies). **402**.
- Nuwaysir EF, Bittner M, Trent J, Barrett JC & Afshari CA** 1999 Microarrays and toxicology: the advent of toxicogenomics. *Molecular carcinogenesis* **24** 153-159.
- OECD** 2009 OECD Guideline for the Testing of Chemicals: Carcinogenicity Studies (Test Guideline 451).
- Oishi N, Yamashita T & Kaneko S** 2014 Molecular biology of liver cancer stem cells. *Liver cancer* **3** 71-84.
- Okazaki T, Himi T, Peterson C & Mori N** 1993 Induction of stathmin mRNA during liver regeneration. *FEBS letters* **336** 8-12.
- Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, Sipes N, Kleinstreuer N, Knudsen T, Dix DJ & Reif DM** 2012 Zebrafish developmental screening of the ToxCast Phase I chemical library. *Reproductive Toxicology* **33** 174-187.
- Pakozdy A, Sarchahi AA, Leschnik M, Tichy AG, Halasz P & Thalhammer JG** 2012 Treatment and long-term follow-up of cats with suspected primary epilepsy. *Journal of Feline Medicine and Surgery* **15** 267-273.
- Papa S, Zazzeroni F, Pham CG, Bubici C & Franzoso G** 2004 Linking JNK signaling to NF-kappaB: a key to survival. *Journal of cell science* **117** 5197-5208.
- Pastorino JG & Hoek JB** 2000 Ethanol potentiates tumor necrosis factor-alpha cytotoxicity in hepatoma cells and primary rat hepatocytes by promoting induction of the mitochondrial permeability transition. *Hepatology* **31** 1141-1152.
- Patlewicz G, Simon TW, Rowlands JC, Budinsky RA & Becker RA** 2015 Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regulatory Toxicology and Pharmacology* **71** 463-477.
- Patterson KI, Brummer T, O'Brien PM & Daly RJ** 2009 Dual-specificity phosphatases: critical regulators with diverse cellular targets. *The Biochemical journal* **418** 475-489.

Pfaffl MW, Horgan GW & Dempfle L 2002 Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research* **30** e36.

Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, Gutkovich-Pyest E, Urieli-Shoval S, Galun E & Ben-Neriah Y 2004 NF-kappaB functions as a tumour promoter in inflammation-associated cancer. *Nature* **431** 461-466.

Powell CL, Kosyk O, Ross PK, Schoonhoven R, Boysen G, Swenberg JA, Heinloth AN, Boorman GA, Cunningham ML, Paules RS & Rusyn I 2006 Phenotypic anchoring of acetaminophen-induced oxidative stress with gene expression profiles in rat liver. *Toxicological Sciences* **93** 213-222.

Pye C & Crews C 2014 Furan in canned sardines and other fish. *Food additives & contaminants. Part B, Surveillance* **7** 43-45.

Rauth S & Davidson RL 1993 Suppression of tyrosinase gene expression by bromodeoxyuridine in Syrian hamster melanoma cells is not due to its incorporation into upstream or coding sequences of the tyrosinase gene. *Somatic cell and molecular genetics* **19** 285-293.

Rauth S, Hoganson GE & Davidson RL 1990 Bromodeoxyuridine- and cyclic AMP-mediated regulation of tyrosinase in Syrian hamster melanoma cells. *Somatic cell and molecular genetics* **16** 583-592.

Ravo M, Mutarelli M, Ferraro L, Grober OM, Paris O, Tarallo R, Vigilante A, Cimino D, De Bortoli M, Nola E, Cicatiello L & Weisz A 2008 Quantitative expression profiling of highly degraded RNA from formalin-fixed, paraffin-embedded breast tumor biopsies by oligonucleotide microarrays. *Laboratory Investigation* **88** 430-440.

R-Core-Development-Team 2012 R: A Language and Environment for Statistical Computing. . **Version 2.11.1**.

Recio L, Phillips SL, Maynor T, Waters M, Jackson AF & Yauk CL 2013 Differential Expression of Long Non-Coding RNAs in the Livers of Female B6C3F1 Mice Exposed to the Carcinogen Furan. *Toxicological Sciences* **135** 369-379.

Rentoft M, Coates PJ, Laurell G & Nylander K 2012 Transcriptional profiling of formalin fixed paraffin embedded tissue: pitfalls and recommendations for identifying biologically relevant changes. *PLoS ONE* **7** e35276.

Rinn JL & Chang HY 2012 Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry* **81** 145-166.

- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W & Smyth GK** 2015 limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43** e47.
- Roat R, Rao V, Doliba NM, Matschinsky FM, Tobias JW, Garcia E, Ahima RS & Imai Y** 2014 Alterations of Pancreatic Islet Structure, Metabolism and Gene Expression in Diet-Induced Obese C57BL/6J Mice. *PloS one* **9** e86815.
- Roberts L, Bowers J, Sensinger K, Lisowski A, Getts R & Anderson MG** 2009 Identification of methods for use of formalin-fixed, paraffin-embedded tissue samples in RNA expression profiling. *Genomics* **94** 341-8.
- Romer M, Eichner J, Metzger U, Templin MF, Plummer S, Ellinger-Ziegelbauer H & Zell A** 2014 Cross-platform toxicogenomics for the prediction of non-genotoxic hepatocarcinogenesis in rat. *PloS one* **9** e97640.
- Ross J, Plummer SM, Rode A, Scheer N, Bower CC, Vogel O, Henderson CJ, Wolf CR & Elcombe CR** 2010 Human constitutive androstane receptor (CAR) and pregnane X receptor (PXR) support the hypertrophic but not the hyperplastic response to the murine nongenotoxic hepatocarcinogens phenobarbital and chlordane in vivo. *Toxicological Sciences* **116** 452-66.
- Ross PK, Woods CG, Bradford BU, Kosyk O, Gatti DM, Cunningham ML & Rusyn I** 2009 Time-course comparison of xenobiotic activators of CAR and PPARalpha in mouse liver. *Toxicology and Applied Pharmacology* **235** 199-207.
- Rowlands DC, Harrison RF, Jones NA, Williams A, Hubscher SG & Brown G** 1995 Stathmin is expressed by the proliferating hepatocytes during liver regeneration. *Journal of Clinical Pathology - Clinical Molecular Pathology* **48** M88-M92.
- Roy NC, Altermann E, Park ZA & McNabb WC** 2011 A comparison of analog and Next-Generation transcriptomic tools for mammalian studies. *Briefings in functional genomics* **10** 135-150.
- Rudolph D, Yeh W-, Wakeham A, Rudolph B, Nallainathan D, Potter J, Elia AJ & Mak TW** 2000 Severe liver degeneration and lack of NF- κ B activation in NEMO/IKK γ -deficient mice. *Genes and Development* **14** 854-862.
- Russell W & Burch R** 1959 The Principles of Human Experimental Technique.
- Russom CL, LaLone CA, Villeneuve DL & Ankley GT** 2014 Development of an adverse outcome pathway for acetylcholinesterase inhibition leading to acute mortality. *Environmental toxicology and chemistry / SETAC* **33** 2157-2169.

Sadi AM, Wang DY, Youngson BJ, Miller N, Boerner S, Done SJ & Leong WL 2011 Clinical relevance of DNA microarray analyses using archival formalin-fixed paraffin-embedded breast cancer specimens. *BMC Cancer* **11** 253:1-13.

Saitoh M, Nishitoh H, Fujii M, Takeda K, Tobiume K, Sawada Y, Kawabata M, Miyazono K & Ichijo H 1998 Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO Journal* **17** 2596-2606.

Schrum LW, Black D, Iimuro Y, Rippe RA, Brenner DA & Behrns KE 2000 c-Jun does not mediate hepatocyte apoptosis following NF κ B inhibition and partial hepatectomy. *Journal of Surgical Research* **88** 142-149.

Searle SR, Speed FM & Milliken GA 1980 The population marginal means in the linear model: An alternative to least squares means. *The American Statistician* **34** 216-216-221.

Seki E, Brenner DA & Karin M 2012 A liver full of JNK: Signaling in regulation of cell function and disease pathogenesis, and clinical approaches. *Gastroenterology* **143** 307-320.

SEQC/MAQC-III Consortium 2014 A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature biotechnology* **32** 903-914.

Setiawan VW, Wilkens LR, Lu SC, Hernandez BY, Le Marchand L & Henderson BE 2015 Association of coffee intake with reduced incidence of liver cancer and death from chronic liver disease in the US multiethnic cohort. *Gastroenterology* **148** 118-25; quiz e15.

Settivari RS, Ball N, Murphy L, Rasoulpour R, Boverhof DR & Carney EW 2015 Predicting the future: opportunities and challenges for the chemical industry to apply 21st-century toxicity testing. *Journal of the American Association for Laboratory Animal Science* **54** 214-223.

Singh R & Czaja MJ 2007 Regulation of hepatocyte apoptosis by oxidative stress. *Journal of gastroenterology and hepatology* **22** Suppl 1 S45-8.

Singh R, Wang Y, Schattenberg JM, Xiang Y & Czaja MJ 2009 Chronic oxidative stress sensitizes hepatocytes to death from 4-hydroxynonenal by JNK/c-Jun overactivation. *American journal of physiology. Gastrointestinal and liver physiology* **297** G907-17.

Smyth GK 2005 limma: Linear Models for Microarray Data. In pp 397-420. Eds R Gentleman, V Carey, W Huber, R Irizarry and S Dudoit Springer New York.

- Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD & Duncavage EJ** 2013 Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *Journal of Molecular Diagnostics* **15** 623-633.
- Stahl S, Ittrich C, Marx-Stoelting P, Kohle C, Ott T, Buchmann A & Schwarz M** 2005 Effect of the tumor promoter phenobarbital on the pattern of global gene expression in liver of connexin32-wild-type and connexin32-deficient mice. *International Journal of Cancer* **115** 861-869.
- Stanger BZ** 2015 Cellular homeostasis and repair in the mammalian liver. *Annual Review of Physiology* **77** 179-200.
- Tamasi V, Juvan P, Beer M, Rozman D & Meyer UA** 2009 Transcriptional activation of PPARalpha by phenobarbital in the absence of CAR and PXR. *Molecular pharmacology* **6** 1573-1581.
- Templin MV, Jamison KC, Sprankle CS, Wolf DC, Wong BA & Butterworth BE** 1996 Chloroform-induced cytotoxicity and regenerative cell proliferation in the kidneys and liver of BDF1 mice. *Cancer Letters* **108** 225-231.
- Thomas R, Thomas RS, Auerbach SS & Portier CJ** 2013a Biological Networks for Predicting Chemical Hepatocarcinogenicity Using Gene Expression Data from Treated Mice and Relevance across Human and Rat Species. *PLoS ONE* **8**.
- Thomas RS, Clewell III HJ, Allen BC, Wesselkamper SC, Wang NCY, Lambert JC, Hess-Wilson JK, Zhao QJ & Andersen ME** 2011a Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicological Sciences* **120** 194-205.
- Thomas RS, Clewell III HJ, Allen BC, Wesselkamper SC, Wang NCY, Lambert JC, Hess-Wilson JK, Zhao QJ & Andersen ME** 2011b Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicological Sciences* **120** 194-205.
- Thomas RS, Clewell HJ,3rd, Allen BC, Wesselkamper SC, Wang NC, Lambert JC, Hess-Wilson JK, Zhao QJ & Andersen ME** 2011c Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicological sciences : an official journal of the Society of Toxicology* **120** 194-205.
- Thomas RS, Clewell HJ, Allen BC, Yang L, Healy E & Andersen ME** 2012 Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: A five chemical case study. *Mutation Research - Genetic Toxicology and Environmental Mutagenesis* **746** 135-143.

- Thomas RS, O'Connell TM, Pluta L, Wolfinger RD, Yang L & Page TJ** 2007 A comparison of transcriptomic and metabonomic technologies for identifying biomarkers predictive of two-year rodent cancer bioassays. *Toxicological Sciences* **96** 40-46.
- Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME, Meek ME, Reiter LW, Lambert JC, Clewell HJ,3rd, Stephens ML, Zhao QJ, Wesselkamper SC, Flowers L, Carney EW, Pastoor TP, Petersen DD, Yauk CL & Nong A** 2013b Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. *Toxicological Sciences* **136** 4-18.
- Thomas SR, Wesselkamper SC, Wang NCY, Zhao QJ, Petersen DD, Lambert JC, Cote I, Yang L, Healy E, Black MB, Clewell III HJ, Allen BC & Andersen ME** 2013c Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicological Sciences* **134** 180-194.
- Thompson MR, Xu D & Williams BR** 2009 ATF3 transcription factor and its emerging roles in immunity and cancer. *Journal of Molecular Medicine* **87** 1053-1060.
- Thomson EM, Williams A, Yauk CL & Vincent R** 2012 Overexpression of tumor necrosis factor- α in the lungs alters immune response, matrix remodeling, and repair and maintenance pathways. *American Journal of Pathology* **180** 1413-1430.
- Thybaud V, Le Fevre AC & Boitier E** 2007 Application of toxicogenomics to genetic toxicology risk assessment. *Environmental and molecular mutagenesis* **48** 369-379.
- Tien ES & Negishi M** 2006 Nuclear receptors CAR and PXR in the regulation of hepatic metabolism. *Xenobiotica* **36** 1152-63.
- Tonk EC, Pennings JL & Piersma AH** 2014 An adverse outcome pathway framework for neural tube and axial defects mediated by modulation of retinoic acid homeostasis. *Reproductive Toxicology* **55** 104-113.
- Tsuchishima M, George J, Shiroeda H, Arisawa T, Takegami T & Tsutsumi M** 2013 Chronic Ingestion of Ethanol Induces Hepatocellular Carcinoma in Mice Without Additional Hepatic Insult. *Digestive diseases and sciences* 1-11.
- Tulloch JK, Carr RR & Ensom MH** 2012 A systematic review of the pharmacokinetics of antiepileptic drugs in neonates with refractory seizures. *Journal of Pediatric Pharmacology and Therapeutics* **17** 31-44.
- US-EPA** 2012 Benchmark Dose Technical Guidance. **EPA/100/R-12/001**.
- Venables WN & Ripley BD** 1999 Modern Applied Statistics with S.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, Ottinger MA, Vergauwen L & Whelan M 2014a Adverse outcome pathway (AOP) development I: strategies and principles. *Toxicological Sciences* **142** 312-320.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M, Ottinger MA, Vergauwen L & Whelan M 2014b Adverse outcome pathway development II: best practices. *Toxicological Sciences* **142** 321-330.

Vinken M, Landesmann B, Goumenou M, Vinken S, Shah I, Jaeschke H, Willett C, Whelan M & Rogiers V 2013 Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. *Toxicological Sciences* **136** 97-106.

Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, Fang H, Hong H, Shen J, Su Z, Meehan J, Li X, Yang L, Li H, Labaj PP, Kreil DP, Megherbi D, Gaj S, Caiment F, van Delft J, Kleinjans J, Scherer A, Devanarayan V, Wang J, Yang Y, Qian HR, Lancashire LJ, Bessarabova M, Nikolsky Y, Furlanello C, Chierici M, Albanese D, Jurman G, Riccadonna S, Filosi M, Visintainer R, Zhang KK, Li J, Hsieh JH, Svoboda DL, Fuscoe JC, Deng Y, Shi L, Paules RS, Auerbach SS & Tong W 2014 The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nature biotechnology* **32** 926-932.

Wang J, Huang Q & Chen M 2003 The role of NF-kappaB in hepatocellular carcinoma cell. *Chinese medical journal* **116** 747-752.

Wang L, Wang S & Li W 2012 RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28** 2184-2185.

Wapinski O & Chang HY 2011 Long noncoding RNAs and human disease. *Trends in Cell Biology* **21** 354-361.

Waters MD, Jackson M & Lea I 2010 Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. *Mutation Research - Reviews in Mutation Research* **705** 184-200.

Webster AF, Williams A, Recio L & Yauk CL 2014 Bromodeoxyuridine (BrdU) treatment to measure hepatocellular proliferation does not mask furan-induced gene expression changes in mouse liver. *Toxicology* **323** 26-31.

Wilson DM, Goldsworthy TL, Popp JA & Buttersorth BE 1992 Evaluation of genotoxicity, pathological lesions, and cell proliferation in livers of rats and mice treated with furan. *Environmental and molecular mutagenesis* **19** 209-222.

- Wu H, Kerr KK, Cui X & Churchill GA** 2003 MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments.
- Wu J, Li J, Salcedo R, Mivechi NF, Trinchieri G & Horuzsko A** 2012 The proinflammatory myeloid cell receptor TREM-1 controls Kupffer cell activation and development of hepatocellular carcinoma. *Cancer research* **72** 3977-3986.
- Xu C, Li CY & Kong AN** 2005 Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research* **28** 249-68.
- Xu D, Yang F, Yuan J-, Zhang L, Bi H-, Zhou C-, Liu F, Wang F & Sun S-** 2013 Long noncoding RNAs associated with liver regeneration 1 accelerates hepatocyte proliferation during liver regeneration by activating Wnt/ β -Catenin signaling. *Hepatology* **58** 739-751.
- Yamada Y, Kirillova I, Peschon JJ & Fausto N** 1997 Initiation of liver growth by tumor necrosis factor: Deficient liver regeneration in mice lacking type I tumor necrosis factor receptor. *Proceedings of the National Academy of Sciences of the United States of America* **94** 1441-1446.
- Yang L, Allen BC & Thomas RS** 2007 BMDEExpress: a software tool for the benchmark dose analyses of genomic data. *BMC Genomics* **8** 387.
- Yang L, Lin C & Rosenfeld MG** 2011 A lincRNA switch for embryonic stem cell fate. *Cell research* **21** 1646-1648.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J & Speed TP** 2002 Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic acids research* **30**.
- Yanger K & Stanger BZ** 2011 Facultative stem cells in liver and pancreas: fact and fancy. *Developmental dynamics* **240** 521-529.
- Yao CQ, Prokopec SD, Watson JD, Pang R, P'ng C, Chong LC, Harding NJ, Pohjanvirta R, Okey AB & Boutros PC** 2012 Inter-strain heterogeneity in rat hepatic transcriptomic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Toxicology and applied pharmacology* **260** 135-145.
- Yauk CL, Lambert IB, Meek ME, Douglas GR & Marchetti F** 2015 Development of the adverse outcome pathway "alkylation of DNA in male premeiotic germ cells leading to heritable mutations" using the OECD's users' handbook supplement. *Environmental and molecular mutagenesis*.
- Yoon J-, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG & Gorospe M** 2012 LincRNA-p21 Suppresses Target mRNA Translation. *Molecular cell* **47** 648-655.

- Yoon M, Blauboer BJ & Clewell HJ** 2015 Quantitative in vitro to in vivo extrapolation (QIVIVE): An essential element for in vitro-based risk assessment. *Toxicology* **332** 1-3.
- Young MR, Yang H- & Colburn NH** 2003 Promising molecular targets for cancer prevention: AP-1, NF- κ B and Pcd4. *Trends in molecular medicine* **9** 36-41.
- Yuan L & Kaplowitz N** 2009 Glutathione in liver diseases and hepatotoxicity. *Molecular aspects of medicine* **30** 29-41.
- Yuan R-, Jeng Y-, Chen H-, Lai P-, Pan H-, Hsieh F-, Lin C-, Lee PH & Hsu H-** 2006 Stathmin overexpression cooperates with p53 mutation and osteopontin overexpression, and is associated with tumour progression, early recurrence, and poor prognosis in hepatocellular carcinoma. *Journal of Pathology* **209** 549-558.
- Zhang J, Ma C, Liu Y, Yang G, Jiang Y & Xu C** 2014 Interleukin 18 accelerates the hepatic cell proliferation in rat liver regeneration after partial hepatectomy. *Gene* **537** 230-237.
- Zhang Y & Chen F** 2004 Reactive oxygen species (ROS), troublemakers between nuclear factor-kappaB (NF-kappaB) and c-Jun NH(2)-terminal kinase (JNK). *Cancer research* **64** 1902-1905.
- Zhao S, Fung-Leung WP, Bittner A, Ngo K & Liu X** 2014a Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PloS one* **9** e78644.
- Zhao W, He X, Hoadley KA, Parker JS, Hayes DN & Perou CM** 2014b Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC genomics* **15** 419-2164-15-419.
- Zheng JL, Parfett C, Williams A, Yagminas A, Zhou G, Douglas GR & Yauk CL** 2011 Assessment of subclinical, toxicant-induced hepatic gene expression profiles after low-dose, short-term exposures in mice. *Regulatory Toxicology and Pharmacology* **60** 54-72.

Appendices

APPENDIX A GENE EXPRESSION ANALYSIS OF LIVERS FROM FEMALE B6C3F1 MICE EXPOSED TO CARCINOGENIC AND NONCARCINOGENIC DOSES OF FURAN, WITH OR WITHOUT 5 DAYS BROMODEOXYURIDINE (BRDU) TREATMENT.

Modified from: A.F. Webster, Williams, A., Recio, R., Yauk, C.L. 2014. *Genomics Data*, 2:117-122.

Preamble: This manuscript describes the method for two-colour microarray analysis of frozen tissues in greater detail. This was the method used for RNA extracted from FRO samples in Chapters 2-4, 6, and Appendix B.

A.1 Abstract

Standard methodology for identifying chemical carcinogens is both time-consuming and resource intensive. Researchers are actively investigating how new technologies can be used to identify chemical carcinogens in a more rapid and cost-effective manner (Mahadevan et al., 2011). Here we performed a toxicogenomic case study of the liver carcinogen furan. Full study and MoA details were previously published in the *Journal of Toxicology and Applied Pharmacology* (Jackson et al., 2014). Female B6C3F1 mice were sub-chronically treated with two non-carcinogenic (1 and 2 mkd) and two carcinogenic (4 and 8 mkd) doses of furan for 21 days. Half of the mice in each dose group were also treated with 0.02% bromodeoxyuridine (BrdU) for five days prior to sacrifice (unpublished data, submitted manuscript). Agilent gene expression microarrays were used to measure changes in liver gene and long non-coding RNA expression (Recio et al., 2013). Here we describe the experimental and quality control details for the microarray data. We also provide the R code used to analyze the raw data files, produce fold change and false discovery rate (FDR) adjusted p values for each gene, and construct hierarchical clustering between datasets

A.2 Specifications

Organism	B6C3F1 mice
Sex	Female
Array type	Agilent SurePrint G3 Mouse GE 8x60K Microarray
Data format (in GEO)	Raw data: TXT files; normalized data: TXT files
Experimental factors	Furan exposed vs. un-exposed control
Experimental features	Female B6C3F1 mice were sub-chronically exposed for 21 days to control (0 mkd), non-carcinogenic (1, 2 mkd), and carcinogenic (4, 8 mkd) doses of furan. Half of the mice in each group were also given 0.02% BrdU for five days prior to sacrifice (days 16-21). All non-BrdU mice as well as the 0, 1, and 8 mkd bw furan +BrdU mice were used for gene expression analysis. Necropsy occurred four hours after the final furan dosing. RNA was extracted from livers and changes in gene expression were analyzed using Agilent microarrays.
Consent	All procedures were conducted in compliance with the Animal Welfare Act Regulations (9CFR1–4). Mice were handled and treated according to the guidelines provided in the National Institutes of Health (NIH) Guide for the Care and Use of Laboratory Animals (ILAR, 1996; dels.nas.edu/ilar/).
Sample source location	5–6 week old female specific pathogen free B6C3F1 mice were purchased from Charles River Breeding Laboratories (Portage, ME). Experiments were conducted at ILS, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA.

A.3 Publically Available Data

The complete dataset is available through the Gene Expression Omnibus.

- Non-BrdU dataset: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48644>
- BrdU dataset : <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54078>

A.4 Experimental Design, Materials and Methods

Animal Study

As described in Chapter 2.

RNA Extraction

As described in Chapter 2.

Microarray: sample labeling and hybridization

Sample RNA (200 ng) was used together with a mouse universal reference RNA (Stratagene by Agilent Technologies Inc.) to synthesize, amplify and label cRNA using the Low Input Quick Amp Labeling Kit (Agilent Technologies Inc.). Labeled cRNA was purified using the RNeasy Mini Kit (Qiagen). Amplification and labeling efficiency of cRNA were quantified using a NanoDrop spectrophotometer. Hybridization mixes were prepared using the Hi-RPM Gene Expression Hybridization Kit (Agilent Technologies Inc.). 300 ng of Cy3-labeled reference RNA and 300 ng Cy5-labeled sample cRNA were hybridized on SurePrint G3 Mouse GE 8 × 60 K microarrays (Agilent Technologies Inc.) at 65 °C for 17 h at 10 rpm. Samples were arranged on arrays according to a randomized block design (RBD). An RBD is used to control for a source of variability that is not related to the experimental question. The microarray slides used in this experiment each

had eight arrays, which means that a total of eight samples could be hybridized to each slide. Since this experiment had more than eight samples, there were two blocking factors that could impact the results: (1) which slide the samples appeared on, and (2) which array the sample was assigned to (i.e. the sample's location on a microarray slide, of which there are eight options). All samples (control, treated, \pm BrdU) were randomized across five slides. Slides were washed according to the manufacturer's specifications with Gene Expression Wash Buffers 1 and 2 (Agilent Technologies Inc.), scanned using an Agilent G2505B scanner at 5 μ m resolution. Data were pre-processed using the Agilent Feature Extraction Software, version 11. All arrays met the minimum Agilent QA/QC standards. All kits were carried out according to the manufacturer's protocol.

Normalization of microarray data

Boxplots of the relative ratio and signal intensities (MA plots) were constructed and inspected. MA plots are used to identify systematic variations (including dye biases and poor hybridization efficiency) within arrays. 'M' is the log red/green intensity ratio ($\log(R/G)$, a measure of differential gene expression) and 'A' is the average log of the product of the two intensities ($\log(R*G)$, a measure of overall fluorescence intensity). Since the majority of probes are not expected to be differentially expressed in response to the treatment, the majority of points on the y-axis (M) should fall at zero. Data where the majority of points on the y-axis are not at zero must be LOWESS (LOcally WEighted Scatterplot Smoothing) normalized (Yang *et al.* 2002) (Fig. A1). Non-background-subtracted median signal intensities were normalized in R (R-Core-Development-Team

2012) using the `transform.madata` function in the MAANOVA library (Wu *et al.* 2003). Probes with technical replicates were then averaged.

Quality assessment and control of microarray data

The background for each array was assessed using the negative control 3xSLv1 probe, which forms a hair pin and does not hybridize well with labeled samples. There are 182 technical replicates of this probe distributed across the microarray. The trimmed mean (trim = 0.05) and standard deviation was used to measure the background fluorescence and background variability for each array. Using trimmed statistics reduces the effects of outliers that may be present. This is done by removing a percentage of the largest and smallest values before calculating the statistics. Figure A2 provides a visual representation of these estimates. Arrays with background fluorescence that fell outside of these ranges were repeated.

To identify microarrays with poor data quality, hierarchical clustering analysis with average linkage (a method for calculating distance between clusters that averages over all pairs of objects between two clusters) was performed in R software (`hclust` function in the `stats` library) using the normalized log ratios (M values) for all probes (Fig. A3). Since the majority of probes will not be differentially expressed in response to the treatment, the expectation is that there will be no differences between samples. To obtain a fair inter-array comparison, signal intensities were adjusted to control for slide effect (i.e. to control for slide-to-slide variation in signal intensity that occurs as an artifact of experimental conditions, as opposed to a real treatment-related effect) using a linear model, which subtracted the estimated slide effect from the log ratios. One minus

the Pearson correlation was used as the measure of dissimilarity. The original dendrogram was cut at 0.3, which has 6 clusters. The 5 samples that clustered separately (at the top) were considered outliers and were then repeated.

Upon repetition and/or removal of outlying arrays (arrays with high background, systematic dye variation, or poor hybridization efficiency), the final sample sizes used for gene expression analysis were $n = 5, 4, 5, 4,$ and 5 for the non-BrdU 0, 1, 2, 4, and 8 mkd furan dose groups, respectively; and $n = 4, 5, 5$ for the 0, 1, and 8 mkd furan +BrdU groups, respectively. Gene expression analysis was not conducted for the 2 and 4 mkd furan +BrdU animals.

Differential gene expression analysis

Differential gene expression was determined in R using the MicroArray ANalysis Of VAriance (MAANOVA) library (Wu *et al.* 2003) . Included in the MAANOVA model are the block (slide) and the treatment effects. A typical ANOVA generates an F statistic, which is the ratio of inter- to intra-group variance. The MAANOVA uses a modified F statistic, called an F_s statistic (Cui *et al.* 2005), to determine gene-specific treatment effects. Using the James-Stein shrinkage concept (James & Stein 1961), the F_s statistic has an improved estimation of error variance because it estimates intra-group variance based on global gene expression (as opposed to expression of individual genes, which have far fewer data points associated with them). The associated p-values were then estimated using the permutation method (i.e. bootstrapping) with 30,000 permutations and residual shuffling. In this instance, bootstrapping increases statistical power of gene expression associated p-values because the typical sample size for gene

expression studies is quite low ($n = \sim 3-6$ per group). The p-values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) approach (Benjamini & Hochberg 1995). This approach sets a more stringent significance threshold of $\alpha = 0.05$ for all comparisons (as opposed to $\alpha = 0.05$ for individual comparisons, which would result in a higher false positive rate). Individual gene fold changes were estimated using least square means of each pairwise comparison. Genes having an FDR-adjusted $P \leq 0.05$ and a fold change $\geq \pm 1.5$ were deemed differentially expressed.

Gene expression meta-analysis in R (Figure 2.2)

Data were obtained from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and the European Bioinformatics Institute (EBI, <http://www.ebi.ac.uk/arrayexpress/>). All data generated using the one-colour Affymetrix platform (E-MEXP-82; GSE13149; GSE18858; GSE20427; GSE26538) were background-subtracted and normalized using Robust Multi-array Average, RMA (Irizarry *et al.* 2003), using the `ReadAffy()` function in the `affy` library in R (Gautier *et al.* 2004). Since the RMA requires background subtraction, data generated by all other platforms were also background subtracted for consistency. Signal intensities generated on the Agilent two-colour platform (GSE48644; BaP) were normalized using the LOWESS approach. For two-colour experiments using other, non-Agilent platforms (GSE35934; GSE4874), signal intensities (with a dye adjustment) were normalized using cyclic LOWESS (Bolstad *et al.* 2003). Normalized probe intensities for probes with common

gene symbols were then averaged using the median normalized signal intensity. Normalized data from each data set were then merged together yielding 3190 common gene symbols. Hierarchical clustering was conducted using the `hclust` function with average linkage using one minus the Pearson correlation as the distance metric in R.

A.5 Discussion

Here we have described the steps taken to analyze changes in gene expression in the livers of furan-exposed female B6C3F1 mice using Agilent microarrays and R (see Fig. A4 for summary). We have demonstrated that the data are of high quality and have provided all of the tools required for reproducibility. The biological significance of the study was previously reported: Chapter 2, differential gene expression (Jackson *et al.* 2014); Appendix B, differential long non-coding RNA expression (Recio *et al.* 2013); and Chapter 3, effect of BrdU on gene expression (Webster *et al.* 2014). We anticipate that the importance of toxicogenomics studies in chemical carcinogen assessment will continue to increase in the coming years and believe that the rate at which this occurs will be highly dependent upon ensuring public availability of these very powerful datasets.

A.6 Tables and Figures

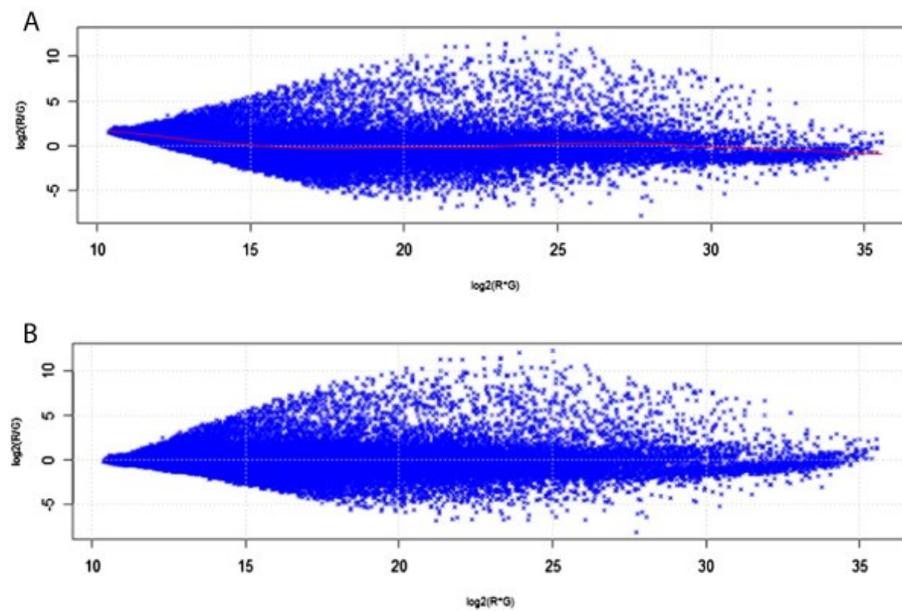


Figure A1 A representative MA plot for (A) before and (B) after LOWESS normalization. Each dot represents a probe on the microarray. The red line is the line of best fit through all data points.

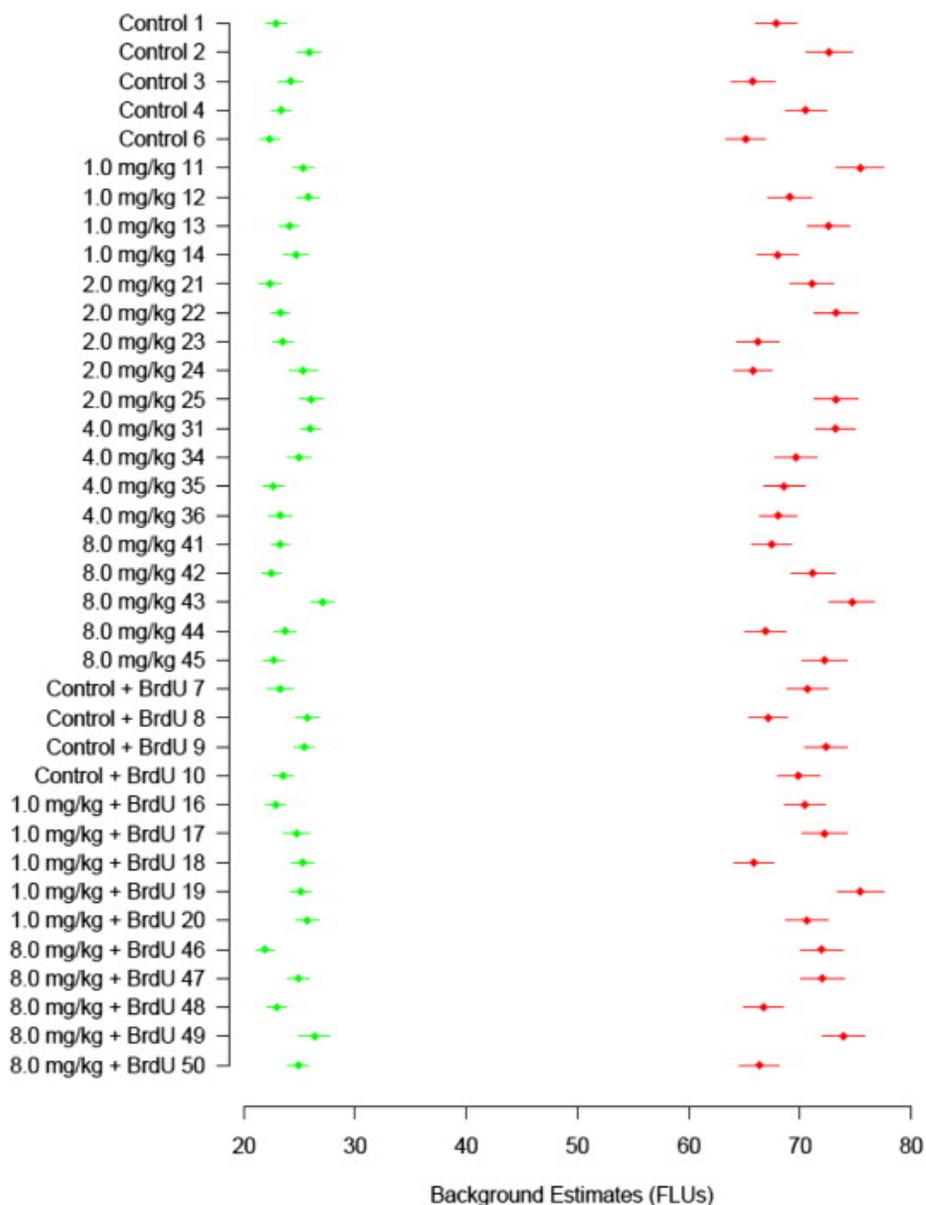


Figure A2 Background estimates of each array for reference (Cy3, green) and sample (Cy5, red) mean fluorescent signal of the 3xSLv1 negative control probe (n = 182 probes). Error bars = standard deviation; FLU = fluorescent units.

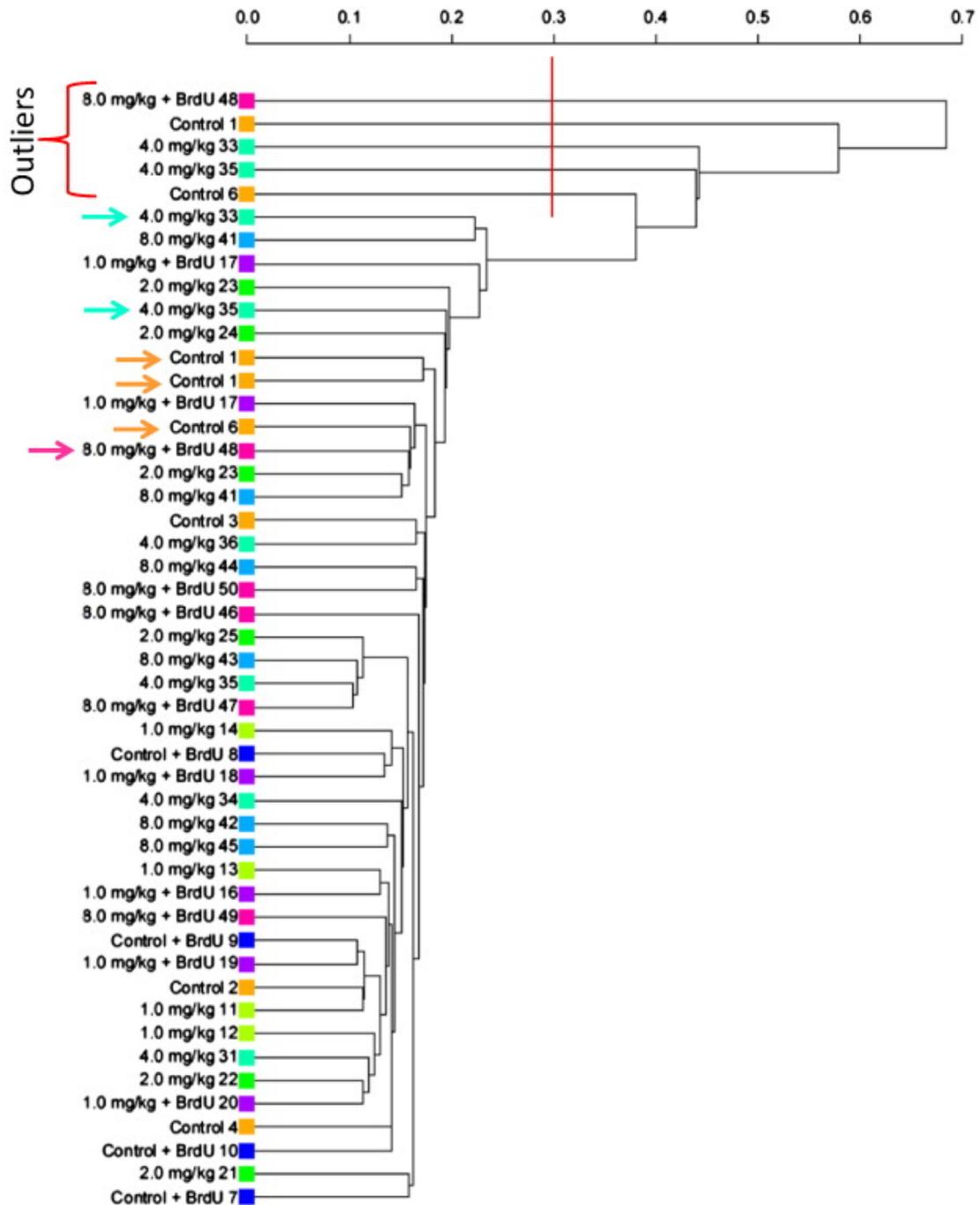


Figure A3 Hierarchical clustering of all probes based on normalized signal intensity ratios. The red line and bracket delineate outlier arrays, which were repeated (repeated arrays are indicated by arrows). Coloured boxes represent dose groups where orange, lime, green, aqua-green, and light blue are the 0, 1, 2, 4, and 8 mkd groups; and, royal blue, purple, and pink are the 0, 1, and 8 mkd + BrdU groups.

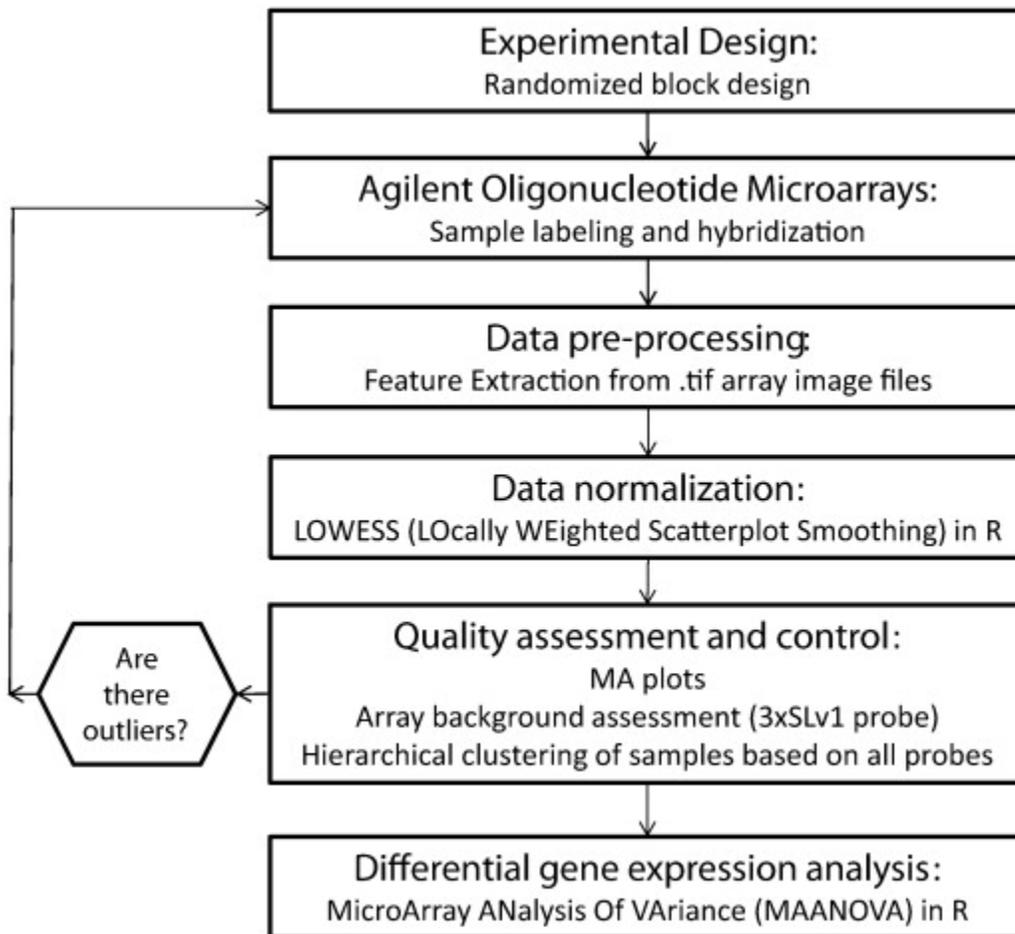


Figure A4 Summary of steps taken to generate, normalize and analyze two-colour Agilent gene expression microarray data.

**APPENDIX B DIFFERENTIAL EXPRESSION OF LONG NON-CODING RNAS
IN THE LIVERS OF FEMALE B6C3F1 MICE EXPOSED TO THE
CARCINOGEN FURAN.**

Modified from: L. Recio, Phillips, S.L., Maynor, T., Waters, M., Jackson, A.F., Yauk, C.L. 2013. *Toxicological Sciences*, 135(2):369-79.

Preamble: This manuscript is a continuation of Chapter 2. However, instead of looking at furan-dependent changes in mRNA expression, here we looked at furan-dependent changes in long non-coding RNA (lncRNA) expression.

B.1 Abstract

The mammalian genome is transcribed into mRNAs that code for protein and a broad spectrum of other noncoding (nc) RNA products. Long ncRNAs (lncRNA), defined as ncRNA species > 200 nucleotides long, are emerging as important epigenetic regulators of gene expression that are involved in a spectrum of biological processes of relevance to toxicology. We conducted a gene expression profiling study in the livers of female B6C3F1 mice exposed to the carcinogen furan at 0, 1, and 2 mkd (noncarcinogenic doses) and at 4 and 8 mkd (carcinogenic doses) for 3 weeks. LncRNA differential expression showed a nonlinear dose response with none differentially expressed at 1 or 2 mkd, 2 lncRNAs at 4 mkd furan, and 83 at 8 mkd, representing 13.3% (83/632) of the total number of differentially expressed transcripts. Among the lncRNAs observed, two lncRNAs examined showed transcriptional clustering with nearby protein-coding genes. LincRNA-p21 is an antisense transcript that is 15 kb downstream from Cdkn1a locus and appears to be cotranscribed with the protein coding gene Cdkn1a at 8 mkd furan. In a separate study, RNA samples from the livers of mice administered benzo(a)pyrene also demonstrated increased levels of Cdkn1a and the antisense lincRNAp21 transcript. These data demonstrate that lncRNAs are transcriptional targets of furan exposures associated with levels of furan that are cytotoxic and induce cell proliferation. In addition, certain lncRNA transcripts are associated with the expression of nearby coding protein genes. We hypothesize that lncRNAs have potential as epigenetic biomarkers of carcinogenic exposures.

B.2 Introduction

Although more than 60% of the content of the mammalian genome are transcribed, less than 2% of this is transcribed into mRNAs that code for proteins (Djebali *et al.* 2012). The transcriptional landscape of the mammalian genome was described in the study by Carninci *et al.* (2005) to consist of “a forest of transcription with overlapping transcription of the genome in both directions separated by deserts in which few transcripts are observed.” Interrogation of the RNA content of a variety of human tissues and cell lines has revealed that, collectively, at least 75% of genome is transcribed in one cell type or another (Clark *et al.* 2011, Djebali *et al.* 2012). Clark *et al.* (2011) analyzed the transcriptome of many mouse tissues and developmental stages and concluded that at least 63% of the genome is transcribed into thousands of protein-coding transcripts and over 30,000 long noncoding intronic, intergenic, and antisense transcripts. Although the pervasive transcription of the genome was once considered transcriptional noise, and the functions of most ncRNAs were unknown, it is apparent that certain ncRNA play critical roles as epigenetic regulators of gene expression for key biological processes of concern to the field of toxicology (Rinn & Chang 2012).

Long ncRNAs (lncRNAs) are defined as RNA species > 200 nucleotides with no protein product. These RNA molecules were initially discovered by their association with transcriptionally active chromatin complexes (Guttman *et al.* 2009, Khalil *et al.* 2009, Rinn & Chang 2012). The transcription of lncRNAs appears to be associated with their coorganization in the genome with transcriptional regulatory elements of protein coding genes. LncRNAs are frequently cotranscribed with nearby protein-coding genes from intergenic regions (referred to as long intergenic noncoding, lincRNAs), whereas others

are cotranscribed from within key protein-coding genes in the sense (lncRNAs) and antisense direction (referred to as natural antisense transcripts (NAT)) (Rinn & Chang 2012).

To date, few lncRNAs have been functionally characterized (especially in rodents); however, those that have been characterized in humans are emerging as regulators of key cellular processes including stem cell pluripotency and differentiation (Guttman *et al.* 2009, Yang *et al.* 2011, Guttman *et al.* 2011), cell cycle (Hung *et al.* 2011), molecular scaffolding (Guttman & Rinn 2012), maintenance of DNA methylation/demethylation and transcriptional gene silencing (Mohammad *et al.* 2012), and global p53-mediated DNA damage response (Huarte *et al.* 2010, Yoon *et al.* 2012). Furthermore, certain lncRNAs are genetically imprinted, and some lncRNAs are processed into functional microRNAs (Moran *et al.* 2012, Jalali *et al.* 2012). Functional studies have also demonstrated that lncRNAs fulfill critical roles as transcriptional and posttranscriptional regulators of gene expression by interacting with chromatin complexes (Guttman & Rinn 2012, Nie *et al.* 2012). Once referred to as the “dark matter” or “junk” DNA of the genome, lncRNAs are emerging as novel epigenomic gene regulators that contribute to disease processes including cancer, neurodegenerative disorders, and diabetes (Wapinski & Chang 2011, Gutschner & Diederichs 2012). These data taken together suggest that lncRNAs can serve as integrated biomarkers of disease-related epigenetic effects.

In a furan dose-response study (0, 1, 2, 4 and 8 mkd) using female B6C3F1 mice (Moser *et al.* 2009), tumours were observed at 4 and 8 mkd and were associated with hepatotoxicity and compensatory cell proliferation in a dose-related manner. Furan is

considered to cause tumours in rodents by a non-genotoxic MoA involving oxidative stress, accompanied by inflammation, cell proliferation, and toxicity (Ding *et al.* 2012, McDaniel *et al.* 2012). In this study, we conducted a short-term 3-week exposure of female mice to doses of 0, 1, 2, 4 and 8 mkd furan and analyzed global expression profiles derived from liver tissue (Jackson *et al.* 2014). This article is focused on the spectrum of lncRNAs that were induced by furan exposure in mice and the analysis by qRT-PCR of two specific lncRNAs: lincRNA-p21 and lncRNA Chromosome (Chr.) 9: 78107225–78118850 and nearby protein-coding genes. These two lincRNAs are co-expressed with nearby protein-coding mRNAs: lincRNA-p21 with *Cdkn1a* (cell cycle control gene, alias p21), and lncRNA Chr. 9: 78107225–78118850 with *Gsta1* (encoding an enzyme that involved in detoxification), and *Dppa5a* (involved pluripotency/differentiation).

B.3 Materials and Methods

Test article

Described above (section 2.4.1).

Animals

Described above (section 2.4.2). Female B6C3F1 mice were chosen for these studies because they were used in previous short-term studies with furan and were used in a furan cancer bioassay (Wilson *et al.* 1992, Moser *et al.* 2009). Male and female B6C3F1 mice appear to respond similarly to furan-induced liver biochemical, proliferative, and cytotoxic effects (Wilson *et al.* 1992, Fransson-Steen *et al.* 1997).

Female mice and the dose range of 1-8 mkd daily were chosen for this study because they represent no observed effect levels (1-2 mkd) to dose levels (4-8 mkd) known to induce a broad spectrum of biological effects including hepatocellular carcinomas at a dose of 8 mkd.

RNA extraction from liver samples

Described above (section 2.4.4).

Microarray analysis of furan liver samples

Described above (section 2.4.5). The SurePrint G3 Mouse GE 8x60K microarrays are printed with 39,430 Entrez Gene RNAs and 16,251 lncRNAs.

Statistical analysis of microarray data

Described above (section 2.4.5 and in Appendix A).

Significant changes in gene expression were defined as p values < 0.05 and ± 1.5 -fold change. The ANOVA analysis and associated contrasts were performed using Partek Genomics Suite (<http://www.partek.com/>). Pathway enrichment analysis was performed using both Partek and Ingenuity Systems (IPA) (<http://www.ingenuity.com/>). Significant enrichment p values were defined as a false discovery corrected p value < 0.05 .

Quantitative RT-PCR

Complementary DNA (cDNA) was generated from total RNA samples using the High Capacity cDNA Reverse Transcription Kit according to the manufacturer's

recommended protocol (Life Technologies, Carlsbad, CA). Quantitative real-time PCR (Q-PCR) was performed using TaqMan Fast Advanced Master Mix and custom TaqMan Gene Expression assays (Life Technologies) on the ViiA 7 Real-Time PCR System (Life Technologies): 50°C for 2 min, 95°C for 20 s followed by 40 cycles of 95°C for 1 s and 60°C for 20 s. All Q-PCR primers used in the study were designed using Primer Express software (Life Technologies). The primer efficiency (E) of each custom TaqMan Gene Expression assay primer set was determined using the CT Slope Method on serial dilutions of cDNA. Q-PCR analysis was performed using the ViiA 7 Real-Time PCR System software using the $2^{-(\Delta\Delta C T)}$ method.

Statistical analysis

Data are presented as means \pm standard error of the mean (SEM). For all endpoints, statistical analyses were done using Analyse-it statistical software (Analyse-it Software, Ltd, Leeds, UK). A one-way ANOVA was performed, and significant differences by ANOVA were further analyzed by Dunnett's multiple comparison methods with $p < 0.05$.

qRT-PCR of lncRNA-p21 from livers of Benzo(a)pyrene-exposed mice

The samples used in this analysis were from a previous experiment by Malik *et al.* (2012), and full experimental details are described therein. Briefly, adult MutaMouseTM males were exposed for 28 days to 0, 25, 50, or 75 mkd benzo(a)pyrene (BaP) dissolved in olive oil (n=5 per group) by oral gavage. Three days after the last exposure, mice were euthanized by cardiac puncture under isoflurane anaesthesia. Mice were maintained

under conditions approved by the Health Canada Animal Care Committee. Livers were removed, flash frozen, and stored at -80°C until use. RNA was extracted and handled as described above. qRT-PCR was conducted and analyzed as described above.

B.4 Results

Differential Expression of lincRNAs in Furan-Exposed Mice

Global expression profiling from the livers of female mice exposed to furan for 3 weeks showed that 2/11 (18.2%) at 4 mkd and 83/632 (13.1%) at 8 mkd of the differentially expressed probe sets on the SurePrint G3 Mouse GE 8x60K microarrays comprised lincRNAs (Table B1). No differentially expressed lincRNAs were observed at 1 or 2 mkd furan in the liver. The lincRNAs in Table B1 are annotated by the Agilent probe set identification number and Chr. position number. The differentially expressed lincRNAs in mouse liver show lengths ranging from 199 to 80,751bp and are transcribed from the sense and antisense directions. With the exception of Chr. 16, each Chr. has 1–8 differentially expressed lincRNAs (Fig. B1). Only two lincRNAs among the 83 in Table B1 are annotated in the lincRNA database (<http://www.lincrnadb.org/>): lincRNA-p21 (Chr. 17: 29183003–29217681) and linc1257 (Chr. 11: 6421951–6429173). The chromosomal positions and lengths of other lincRNAs, as shown in Table B1, are from the Agilent database and the NCBI.

lincRNAs and Clusters of Transcription

A number of lincRNAs show “clusters of transcription” with multiple transcripts originating from relative short segments of the genome: on mouse Chr. 1, three lincRNAs

(199bp, 521bp, and 78kb) are transcribed from a 78kb chromosomal region; on mouse Chr. 2, four lncRNAs (567bp, 23kb, 62kb, and 39kb) are transcribed from a 1356kb chromosomal region; on mouse Chr. 4, four lncRNAs (702bp, 5.6kb, 7.3kb, and 25kb) are transcribed from a 21,665kb chromosomal region; and on Chr. 9, four lncRNAs (1.6kb, 12.0kb, 13.6kb, and 28.3kb) are transcribed from a 5,752kb chromosomal region (Fig. B2). Mouse Chr. 17 has three clusters of lncRNAs transcription (Fig. B2). The “clusters of transcription” shown on Chr. 9 and 17 are nearby the protein-coding genes *Gst1a* (Chr. 9), *Dppa5a* (Chr. 9), and *Cdkn1a* (Chr. 17); these genes were also differentially expressed in furan-exposed mice (Chapter 2). These three genes represent (1) *Cdkn1a*—a cell cycle gene known to be altered by a number of carcinogenic exposures in mouse liver (Harrill *et al.* 2009), (2) *Gst1a*—a member of the alpha class glutathione-S-transferases expressed in liver (Knight *et al.* 2007), and (3) *Dppa5a* (developmental pluripotency associated 5A)—a gene associated with stem cell pluripotency and differentiation (Kim *et al.* 2005).

qRT-PCR of lncRNAs and Protein-Encoding Genes Cdkn1a, Gst1a, and Dppa5a

To validate the co-transcriptional regulation of specific clusters of lncRNAs/mRNAs by furan, we conducted qRT-PCR of a lncRNA (lncRNA Chr. 9:78107225–78118850) nearby the protein-coding genes *Gst1a*, *Dppa5a*, (Fig. B3); and lncRNA-p21, an antisense lncRNA located 15kb downstream of *Cdkn1a* (Fig. B4). We examined the degree of correlation of these RNAs in furan-treated livers. These three genes were among the top 10 genes impacted by furan exposure (Jackson *et al.* 2014). qRT-PCR analysis revealed that the lncRNA on Chr. 9, in addition to the associated

genes *Gsta1* and *Dppa5a*, was not affected at 1, 2, or 4 mkd furan relative to control mice. However, all three transcripts (*lncRNA* Chr. 9: 78107225–78118850, *Gsta1*, and *Dppa5a*) were significantly increased at 8 mkd ($p < 0.05$) (Fig. B 3). *lncRNA-p21* and *Cdkn1a* mRNA each showed a dose response and were significantly elevated at 4 and 8mkd furan ($p < 0.05$) (Fig. B4). These data also suggest that exposure to furan induces bidirectional transcription at the protein-coding locus *Cdkn1a* and downstream in the antisense direction for *lncRNA-p21* with both transcripts as potential independent targets for either p53-mediated transcription or a number of other transcription factors known to act at this site (Fig. B5) (Abbas & Dutta 2009).

qRT-PCR of lncRNA-p21 in Livers of Mice Administered BaP

In order to determine whether the results obtained were specific to furan or reflected a broader response to hepatotoxicants and hepatocarcinogens, we performed qRT-PCR for *lncRNA-p21* on RNA isolated from the livers of mice exposed to the genotoxic carcinogen BaP from a previously reported study (Malik *et al.* 2012). Under the condition used in Malik *et al.*, BaP induced a >10-fold increase in *Cdkn1a* mRNA (Fig. B6). RNA samples from this study analyzed by qRT-PCR also showed a >5-fold induction of *lncRNA-p21* (Fig. B6), again suggesting bidirectional transcription of *Cdkn1a* and *lncRNA-p21* in BaP-exposed mice.

B.5 Discussion

The demonstration that a major fraction of the eukaryotic genome is transcribed into tens of thousands *lncRNAs* has prompted the functional characterization of many of

the ncRNA species. Although the functions of a large proportion of lncRNAs remain elusive, certain lncRNAs play key roles in biological processes and genome response networks that are known to be impacted by hepatotoxicants and hepatocarcinogens (Cui & Paules 2010, Rinn & Chang 2012). LncRNAs exert their biological effects through a number of mechanisms including direct binding to target mRNAs and interacting with chromatin regulators to control gene expression (Rinn & Chang 2012). These current data suggest that lncRNAs are part of the cellular circuitry regulating differentiation, stress responses, and biochemical defense networks that can respond to cellular stressors (Rinn & Chang 2012).

The observation that lncRNAs comprise a large part of the hepatocellular response to the liver carcinogen furan is novel. We hypothesized that it is not furan specific but rather reflects the transcriptome response to the biological impact of liver carcinogen-induced cytotoxicity and cellular proliferation. This global cellular defense response includes alterations in mRNA levels of protein-coding genes within cellular response pathways and as part of the epigenomic response network miRNA and lncRNAs. A recent article has shown the association with increased expression of an imprinted Dlk1-Dio3lncRNA cluster with hepatocyte hypertrophy in phenobarbital-exposed mice (Lempiäinen *et al.* 2013). The induction of the imprinted Dlk1-Dio3n lncRNA cluster was genetically dependent on both CAR and β -catenin signaling pathways, well-known pathways that impact mouse liver tumour development. A second expression profiling and mechanistic study using siRNA knockdown in livers of mice that had undergone 2/3 partial hepatectomy has also shown that a specific lncRNA (referred to as lncRNA associated with liver regeneration 1 or LALR1) has a key role in

accelerating liver cell proliferation by activating the Wnt/ β -Catenin signaling pathways (Xu *et al.* 2013). A previously published study from our group demonstrates dose-dependent induction of miR-34a and Cdkn1a mRNA in the liver of mice exposed to BaP (Malik *et al.* 2012). Reanalysis of RNA samples from this study by qRT-PCR also showed a dose-dependent induction of lincRNA-p21 (Fig. B6). These data taken together demonstrate that the transcriptional response to liver carcinogens includes not only protein-coding mRNAs and noncoding miRNAs but also includes lncRNAs. Moreover, the apparent pervasive transcription of the lncRNA epigenome in response to furan appears to be more robust than the limited miRNA epigenomic response of the liver to carcinogenic exposures such as BaP (Malik *et al.* 2012). These data demonstrate that the MoA for the non-genotoxic carcinogen furan has an epigenetic component, and BaP has both a mutagenic and an epigenetic component to its MoA.

Among the 83 lncRNAs showing differential expression in furan-exposed mice from this study, only 2 have been explored for functional properties in mouse and human cells: linc1257 and lincRNA-p21. Linc1257 is expressed in mouse embryonic stem cells and interacts with a number of chromatin-binding proteins/complexes in stem cells leading to repression of gene expression (Guttman *et al.* 2011). LincRNA-p21 is a repressor of p53-dependent transcriptional responses, affecting the expression of hundreds of p53 gene targets (Huarte *et al.* 2010). Based on studies using siRNA knockdowns, lncRNA-p21 appears to be necessary for p53-dependent induction of apoptosis but not cell cycle arrest (Huarte *et al.* 2010). This is mediated by physical association of lincRNA-p21 with heterogeneous nuclear ribonucleoprotein K (hnRNP-K), directing the genomic localization of hnRNP-K to repressed genes that trigger apoptosis

(Yoon *et al.* 2012). These authors also demonstrated that lincRNA-p21 directly associates with Junb and Ctnnb1 mRNAs and selectively lowers their translation into proteins. These data indicate that lincRNA-p21 has two roles in regulating gene expression: (1) directly hybridizing to target mRNAs in the nucleus and (2) affecting the translation of p53 target genes in the cytoplasm.

Some of the lncRNAs observed in this study reside in the genome near protein-coding sequences and are apparently co-expressed with these genes. Furan resulted in the induction of a number of lncRNAs that demonstrated “clustering of transcription” with multiple lncRNA and protein-coding genes expressed within short chromosomal segments. We used qRT-PCR to validate two cases of apparent lncRNA co-expression with nearby protein-coding genes induced by furan exposure in liver. LincRNA-p21 is 15kb downstream from the start site of the p53 DNA damage-response gene Cdkn1a and appears to be co-expressed with Cdkn1a in the antisense direction (Fig. B5). In keeping with the response observed for furan, BaP exposure resulted in a highly similar dose-dependent increase in lincRNA-p21 (Fig. B6), mir-34a, and Cdkn1a expression in mouse liver (Malik *et al.* 2012). Studies using lincRNA-p21 siRNA knockdowns in hepatocytes cultures could be used to begin understanding impact of lincRNA-p21 on cancer hallmark biomarkers and the role of lincRNA-p21 in hepatotoxicity of carcinogenic exposures (Hanahan & Weinberg 2011).

Numerous transcription factor binding sites are downstream of Cdkn1a and upstream of lncRNA including p53 (Abbas & Dutta 2009), and in humans, this transcription factor-rich binding region is suggested to drive transcription of up to five lncRNAs from the Cdkn1a promoter in both the sense and antisense directions (Hung *et*

al. 2011). In the human genome, two lncRNAs are downstream from the CDKN1 locus and are transcribed in the antisense direction: one of these is referred to as PANDA, whereas the second transcript is lincRNA-p21 (Hung *et al.* 2011). Both PANDA and lincRNA-p21 are part of the human cell p53 DNA damage response. In contrast, we have not shown increased levels for any other lncRNAs immediately downstream from Cdkn1a in the mouse. Bioinformatic examination of the downstream sequence of Cdkn1a also did not show a PANDA ortholog. The apparent species difference in the lincRNA-p21, PANDA, and Cdkn1a transcriptional cluster suggests a potential species difference in the p53 DNA damage response (Hung *et al.* 2011).

Recently, an experimental and computational survey of lncRNAs from three closely related rodent species, *Mus musculus domesticus* (Mmus: C57BL/6J), *Mus musculus castaneus* (Mcas), and *Rattus norvegicus* (Rnor), was conducted (Kutter *et al.* 2012). The authors used both RNA (RNAseq) and H3K4me3-bound (ChIPseq) DNA data (to identify putative active transcription sites) and combined both to construct catalogues of transcripts expressed in the adult liver of Mmus, Mcas, and Rnor. The authors found that a majority of the lncRNAs identified using this approach have no overlap with intergenic lncRNAs annotated in the mouse genome by Ensembl (build 64), demonstrating that current mouse lncRNA catalogues are largely incomplete. In addition, the lncRNAs observed showed mouse and rat strain differences in lncRNA transcripts; lineage-specific lncRNAs were associated with genomically neighboring protein-coding genes, and liver lncRNAs show differential expression when examined at five different developmental stages (e10, e12, e14, e18, and adult) of Mmus liver development. This study in mice and a previous study in humans have demonstrated the striking tissue

specificity of lncRNA expression. For example, a survey of 4273 human lncRNAs revealed that 78% were tissue specific compared with only 19% of the 28,803 protein-coding genes (Cabili *et al.* 2011). Overall, these studies suggest that differential expression of lncRNAs may contribute to strain-, species-, and tissue-specific biological differences. Based on these findings, we also speculate that differential expression of lncRNAs may contribute to the spectrum of genomic and cellular responses to carcinogen exposure.

The purpose of this report was not to provide a comprehensive analysis of a toxicogenomic response of female mouse liver to furan, rather it was to demonstrate that lncRNAs are part of the transcriptional response to carcinogenic exposures, in this case furan. The robust increase of lncRNAs across the mouse genome in livers of female B6C3F1 mice exposed to a cytotoxic carcinogen like furan is a novel observation, suggesting that lncRNAs may be epigenetic targets for mouse liver carcinogens. In addition, there is a clear need to assess if furan alters expression of lncRNAs in non-target tissues.

The cellular processes influenced by lncRNAs can contribute to the six hallmarks of cancer that underlie malignant transformation (Hanahan & Weinberg 2011, Gutschner & Diederichs 2012). A survey of lncRNA expression within a panel of solid cancers has identified a number of novel transcribed regions differentially expressed across distinct cancer types that represent candidate biomarkers (Brunner *et al.* 2012). We propose that lncRNAs can serve as integrated epigenomic biomarkers of carcinogenic exposures since they can reflect a spectrum of biological processes of fundamental relevance to toxicology. Construction of an atlas of mRNA, miRNA, lncRNAs, and other mobile

genetic elements from rodent and human tissues for use in toxicology studies is needed to assess the integrated transcriptional response of the genome to xenobiotic exposures as a tool to interrogate MoA needed for biologically based risk assessments.

B.6 Tables and Figures

Table B1 Differentially Expressed lncRNAs at 4 and 8 mkd furan (lncRNAs with $P \leq 0.05$ and ± 1.5 -fold change are indicated in bold).

Agilent Probe ID	Chromosome number	lncRNA/ chromosome genomic coordinates	lncRNA length (bp)	Strand	Fold change, 4 mkd furan	Fold change, 8 mkd furan
4 and 8 mkd Furan						
A_30_P01025511	4	lncRNA: 149445201–149445902	702	Forward	3.56	7.19
A_30_P01025159	6	lncRNA:129151512–129208737	57226	Reverse	-1.60	-2.07
8 mkd Furan: 83 Unique lncRNAs						
A_30_P01027205	1	lncRNA:138442536–138520766	78231	Forward	1.86	2.07
A_30_P01029956	1	lncRNA:163528200–163528398	199	Forward	1.66	1.89
A_30_P01026536	1	lncRNA:163508244–163586072	77829	Forward	1.3	1.64
A_30_P01026752	1	lncRNA:163528792–163529312	521	Forward	1.81	2.35
A_30_P01025359	1	lncRNA: 23340743–23341144	402	Forward	1.67	3.48
A_30_P01025912	1	lncRNA: 58443698–58449898	6201	Reverse	-1.43	-2.86
A_30_P01024134	2	lncRNA: 60038543–60046318	7776	Reverse	-1.16	-1.55
A_55_P2164075	2	lncRNA: 93292976–93310028	17053	Forward	-1.75	-5.72
A_30_P01026482	2	lncRNA:174945941–174946507	567	Reverse	-1.35	-1.55
A_30_P01018973	2	lncRNA:174943325–174966675	23351	Reverse	-1.37	-1.59

A_30_P01028205	2	lncRNA:175640275–175701875	61601	Forward	-1.30	-1.52
A_30_P01027279	2	lncRNA:176264400–176302925	38526	Forward	-1.26	-1.61
A_30_P01024689	3	lncRNA: 121615836–121616109	274	Forward	-1.62	-4.30
A_30_P01019901	3	lncRNA: 121627545–121639470	11926	Reverse	-1.05	-2.07
A_30_P01030305	3	lncRNA: 21964899–21974524	9626	Reverse	-1.41	-1.66
A_30_P01032887	3	lncRNA: 21973306–21974266	961	Reverse	-1.14	-1.58
A_30_P01027028	3	lncRNA: 51004595–51005002	408	Reverse	1.48	3.16
A_30_P01021547	4	lncRNA: 82156098–82196298	40201	Forward	-2.00	-3.95
A_30_P01025368	4	lncRNA: 116791819–116798263	6445	Reverse	1.13	1.59
A_30_P01021813	4	lncRNA: 116790373–116797823	7451	Reverse	1.57	3.67
A_30_P01018287	4	lncRNA: 131734714–131741964	7251	Reverse	-1.10	1.56
A_30_P01023606	4	lncRNA: 146507925–146532650	24726	Reverse	1.23	1.68
A_30_P01025511	4	lncRNA: 149445201–149445902	702	Forward	3.56	7.19
A_52_P551011	4	lncRNA: 153394697–153400261	5565	Reverse	-1.48	-1.89
A_30_P01029305	5	lncRNA: 22887983–22939658	51676	Reverse	1.96	3.92
A_30_P01028374	5	lncRNA: 22890322–22939119	48798	Reverse	1.87	3.5
A_55_P2075258	5	lncRNA: 24607276–24615927	8652	Forward	-1.60	-2.18
A_55_P2361647	5	lncRNA: 25004885–25007359	2475	Forward	-1.35	-1.73
A_30_P01028938	5	lncRNA: 37051650–37110575	58926	Reverse	1.4	1.93

A_30_P01032429	5	lncRNA: 52461266–52541416	80151	Reverse	1.47	2.82
A_30_P01024631	5	lncRNA: 74487385–74497428	10044	Forward	1.25	1.53
A_30_P01025226	5	lncRNA: 121759042–121789992	30951	Reverse	1.61	2.29
A_51_P416613	6	lncRNA: 35202654–35213496	10843	Forward	–1.26	–1.71
A_30_P01017502	6	lncRNA: 108162092–108162886	795	Reverse	–1.27	–1.84
A_30_P01025159	6	lncRNA: 129151512–129208737	57226	Reverse	–1.60	–2.07
A_55_P2167955	7	lncRNA: 68427625–68456607	28983	Reverse	1.22	1.99
A_30_P01026270	7	lncRNA: 134620707–134626739	6033	Reverse	–1.97	–1.97
A_30_P01018981	7	lncRNA: 13533925–13544525	10601	Forward	2.58	3.77
A_30_P01030178	8	lncRNA: 27912000–27992750	80751	Forward	1.28	1.56
A_30_P01029190	8	lncRNA: 41475940–41476460	521	Forward	1.62	2.36
A_30_P01022473	8	lncRNA: 108120629–108123201	2573	Forward	–2.80	–2.21
A_30_P01029662	8	lncRNA: 10886200–10905925	19726	Forward	–1.70	–2.87
A_30_P01032012	8	lncRNA: 119739165–119739905	741	Forward	1.72	2.32
A_55_P2183513	8	lncRNA: 123113431–123117701	4271	Forward	–2.02	–3.36
A_30_P01031402	8	lncRNA: 124364048–124365705	1658	Forward	3.39	6.31
A_51_P185292	9	lncRNA: 34924313–34938507	14195	Forward	–1.54	–2.07
A_66_P104624	9	lncRNA: 72366069–72379665	13597	Reverse	1.81	2.47
A_30_P01021117	9	lncRNA: 75316023–75344401	28379	Forward	1.58	2.18

A_30_P01020518	9	lncRNA: 78104935–78116974	12040	Forward	1.4	4
A_30_P01018914	9	lncRNA: 78107225–78118850	11626	Forward	2.75	23.43
A_30_P01022249	10	lncRNA: 60473030–60535340	62311	Reverse	1.15	2.13
A_30_P01032909	11	lncRNA: 6421951–6429173	7223	Reverse	2.68	4.54
A_30_P01028002	11	lncRNA: 69189587–69190674	1088	Reverse	-2.40	-2.97
A_30_P01031674	11	lncRNA: 53596919–53672731	75813	Reverse	-1.69	-2.23
A_30_P01030240	11	lncRNA: 53592946–53637922	44977	Reverse	-1.45	-1.60
A_30_P01017718	11	lncRNA: 119797116–119804012	6897	Reverse	-1.22	-1.82
A_30_P01018231	12	lncRNA: 16833551–16901804	68254	Reverse	1.18	1.64
A_30_P01030946	13	lncRNA: 3497296–3500441	3146	Forward	1.54	2.62
A_30_P01024071	13	lncRNA: 65341237–65342006	770	Reverse	-2.59	-3.05
A_30_P01024541	13	lncRNA: 108259901–108272775	12875	Forward	1.34	2.15
A_30_P01018679	14	lncRNA: 73729977–73751877	21901	Reverse	1.65	2.5
A_30_P01031079	14	lncRNA: 73746324–73749119	2796	Reverse	1.72	2.28
A_30_P01020522	14	lncRNA: 65720392–65769859	49468	Reverse	-1.01	-1.96
A_30_P01021252	15	lncRNA: 77164450–77201325	36876	Forward	-1.26	-1.79
A_30_P01033325	15	lncRNA: 83266264–83294664	28401	Reverse	-1.50	-1.54
A_30_P01024684	15	lncRNA: 101084597–101095097	10501	Reverse	-2.19	-2.98
A_30_P01027464	17	lncRNA: 15100982–15115630	14649	Reverse	1.58	2.17

A_30_P01025984	17	lncRNA: 15226181–15250333	24153	Forward	2.24	2.41
A_30_P01021704	17	lncRNA: 17342991–17343357	367	Forward	–1.23	–1.53
A_30_P01020960	17	lncRNA: 29183003–29217681	34679	Reverse	2.21	5.43
A_30_P01023554	17	lncRNA: 34018371–34019394	1024	Reverse	1.24	1.73
A_55_P2075248	17	lncRNA: 35113844–35118733	4890	Reverse	–1.41	–1.64
A_30_P01032529	17	lncRNA: 85422370–85423168	799	Reverse	1.18	1.5
A_30_P01017543	17	lncRNA: 85418489–85427364	8876	Reverse	1.21	1.52
A_55_P2412319	18	lncRNA: 32518711–32537938	19228	Reverse	–1.25	–1.74
A_30_P01017688	18	lncRNA: 84303291–84303890	600	Forward	–1.37	–1.99
A_30_P01030779	18	lncRNA: 84757475–84757969	495	Reverse	–1.44	–1.60
A_30_P01024790	19	lncRNA: 23068397–23190256	121860	Reverse	1.13	1.99
A_30_P01018128	19	lncRNA: 36699712–36763966	64255	Reverse	–1.83	–2.80
A_30_P01026306	19	lncRNA: 57426515–57435165	8651	Reverse	–1.63	–2.07
A_30_P01021226	X	lncRNA:12811766–12813409	1644	Reverse	2.41	2.83
A_52_P223508	X	lncRNA: 47908921–47988498	79578	Reverse	–1.38	–1.64
A_55_P2310898	X	lncRNA: 148733043–148762038	28996	Reverse	–1.45	–2.49

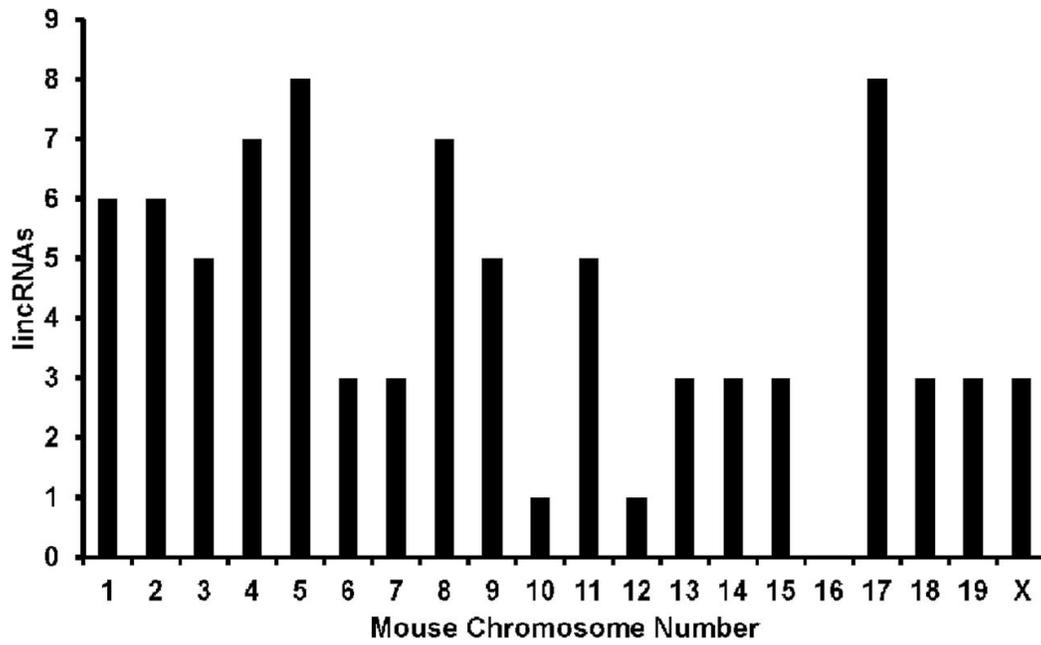


Figure B1 Chromosome distribution of differentially expressed lincRNAs based on microarray data in liver of furan-exposed mice.

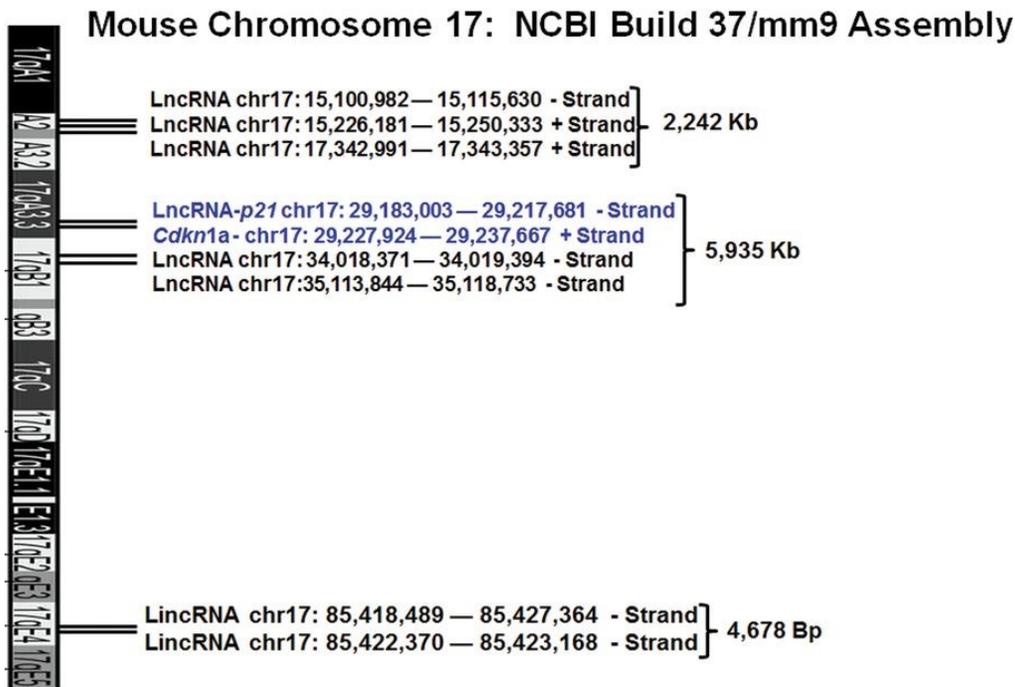
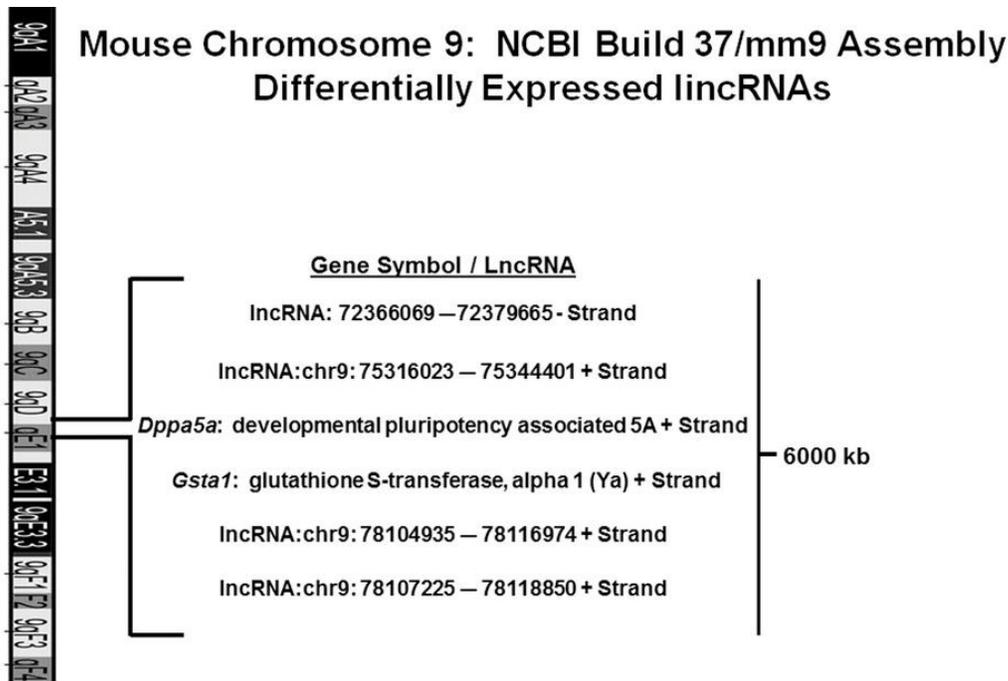


Figure B2 Chromosome localization of differentially expressed lincRNAs and nearby protein-coding genes based on microarray data in liver of furan-exposed mice. (Upper) Chromosome 9; (Lower) Chromosome 17.

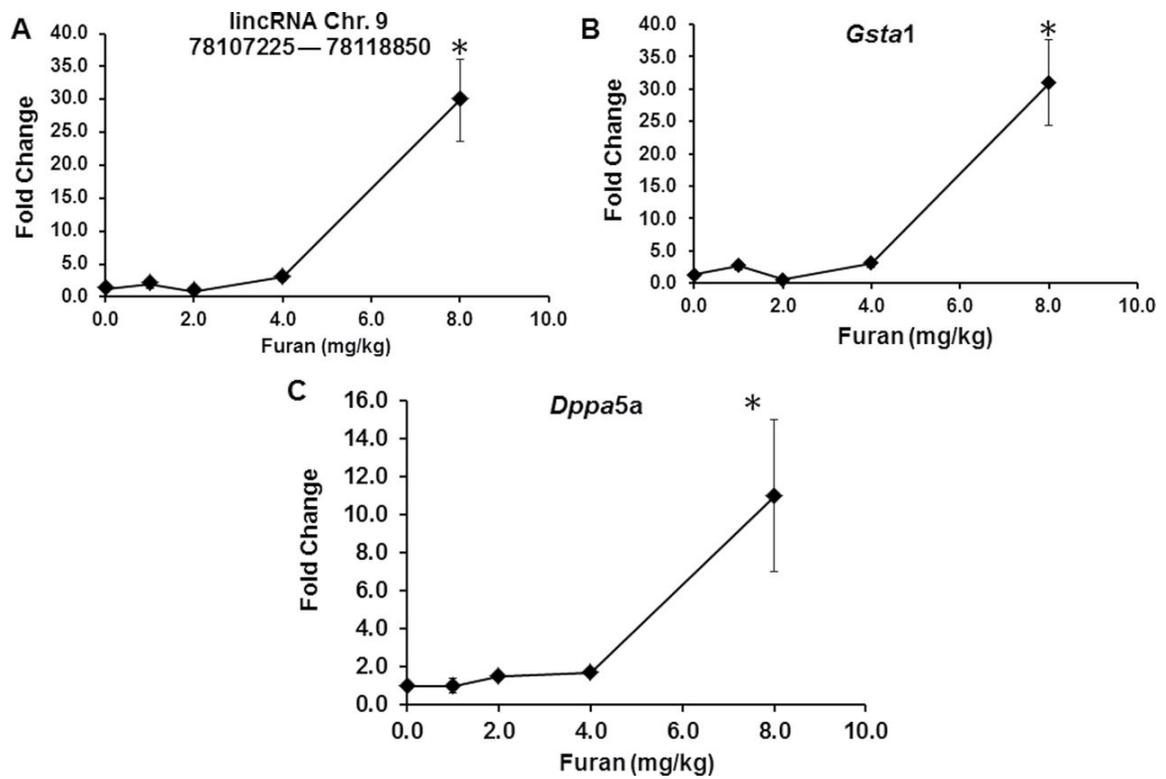


Figure B3 qPCR of lincRNA Chr. 9 78107225–78118850 (A) and *Gsta1* (B) and *Dppa5a* (C). Error bars indicate SE and * $p < 0.05$; 4 and 8mkd are carcinogenic exposure levels (Moser et al., 2009).

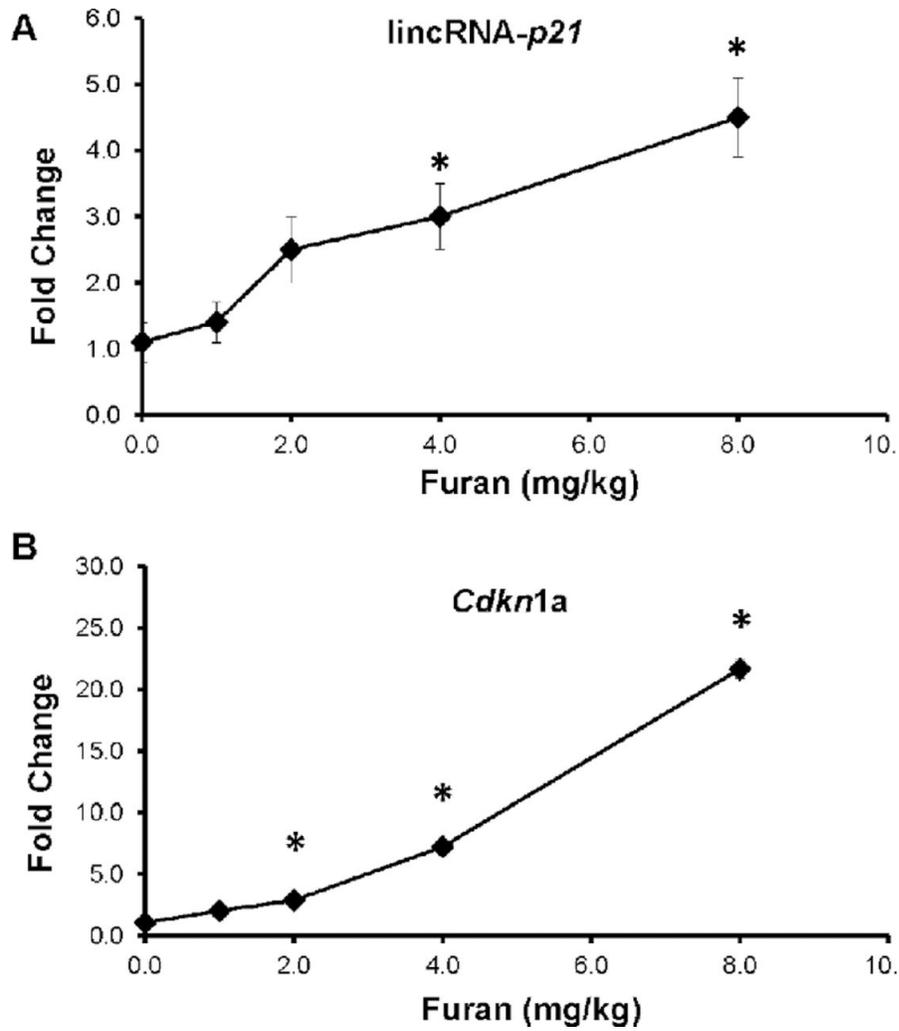


Figure B4 qPCR of *lincRNA-p21* (A) and Chr. 9 78107225–78118850 (A) and *Cdkn1a* (B). Error bars indicate SE and * $p < 0.05$; 4 and 8mkd are carcinogenic exposure levels (Moser et al., 2009).

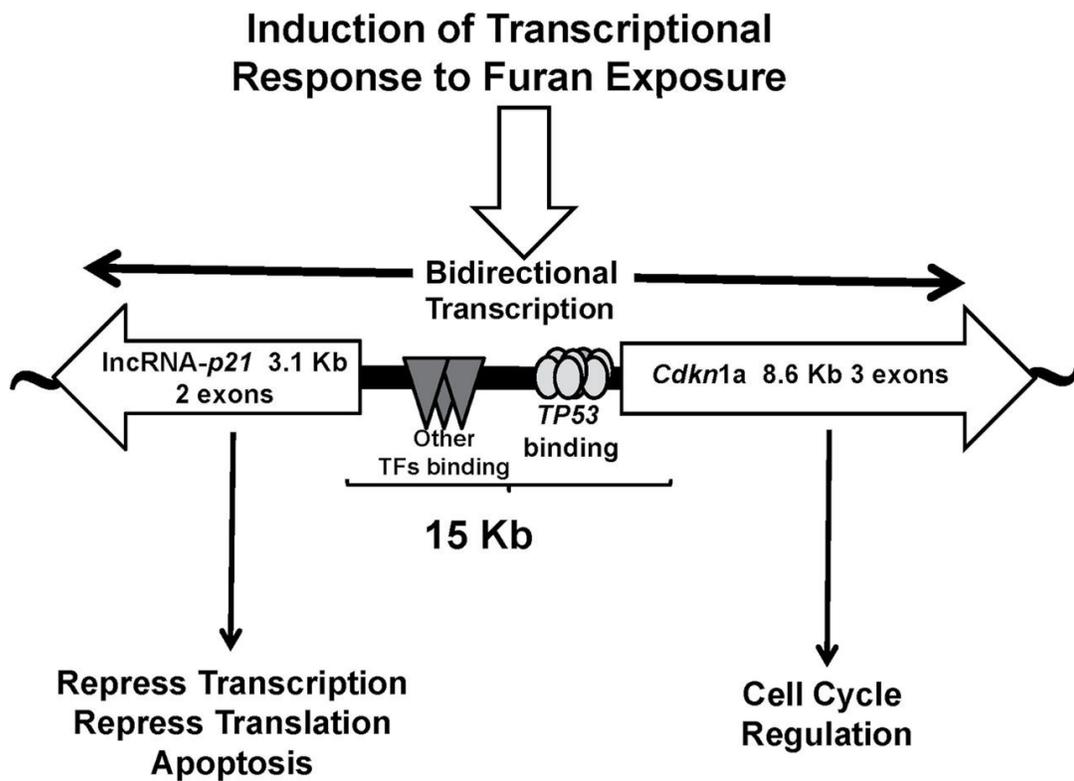


Figure B5 The induction of a transcriptome response in furan-exposed mice induces the apparent bidirectional transcription of *Cdkn1a* and *lincRNA-p21*.

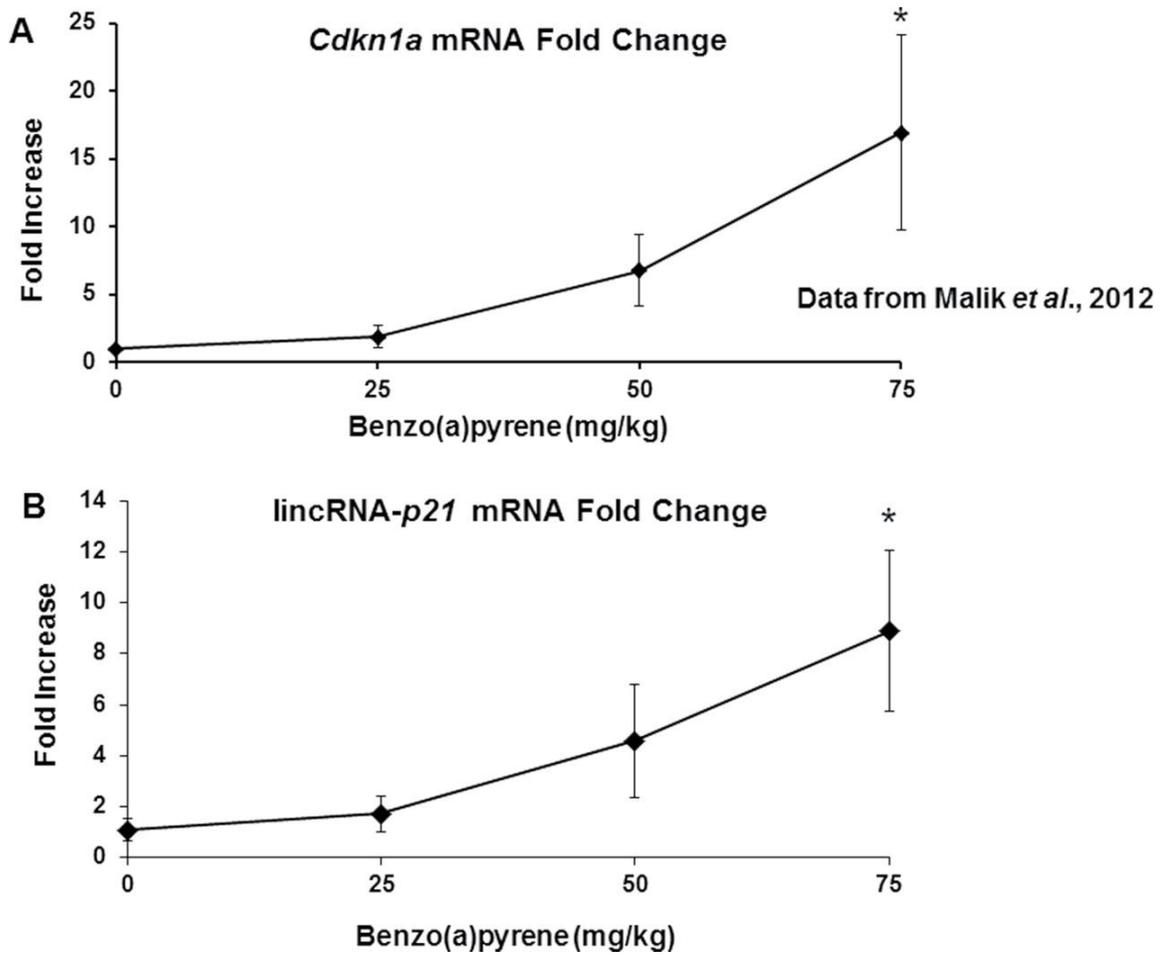


Figure B6 qPCR of *Cdkn1a* (A) (data reproduced from data in Malik et al., 2012) and *lincRNA-p21* (B). Error bars indicate SE and * $p < 0.05$.

APPENDIX C SUPPLEMENTAL MATERIAL

C.1 Chapter 2

Summary of the PCA Analysis

A principle component is a linear combination of the genes, such that the individual components are orthogonal to each other (i.e. independent). These components are then ordered based on the explained variation. For example, the first PC explains 49.6% of the variability for the 463 significant genes ((FDR $p < 0.05$ & $\text{abs}(\text{FC}) > 1.5$).

These components or new variables are calculated for each sample (or observation). For the j^{th} observation and i^{th} principle component, the components are calculation as follows:

$$PC_{j,i} = \sum_{g=1}^p PCLoading_{i,g} \frac{(Ratio_{j,g} - mean_g)}{Std.Dev_g}$$

Here, the vector of PC loadings represents the eigenvector from the spectral decomposition of the correlation matrix. The variance for this eigenvector is the eigenvalue associated with the eigenvector.

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PC10									
Standard Dev.	15.1573	8.4580	5.3275	4.52080	3.96282	3.63338	3.59324	3.17378	2.78212
	2.64731								
Prop. of Var.	0.4962	0.1545	0.0613	0.04414	0.03392	0.02851	0.02789	0.02176	0.01672
	0.01514								
Cum. Proportion	0.4962	0.6507	0.7120	0.75615	0.79007	0.81858	0.84647	0.86823	0.88494
	0.90008								

The first 3 PCs explain 71.2% of the variance.

Genes with large loadings in absolute value, contribute more to the component than genes with loadings small loadings in absolute value.

PC 1 - Top 20 Genes:

A_55_P2094896 A_51_P421140 A_55_P1959703 A_55_P2153291 A_55_P2153292
A_52_P550843 A_55_P2007470 A_55_P2310898 A_55_P2100715 A_55_P2158990
A_51_P320614 A_51_P329928 A_51_P369252 A_55_P2125588 A_55_P2145804
A_55_P2141860 A_51_P415220 A_66_P103792 A_51_P363947 A_51_P510891

Annotation Cluster 1	Enrichment Score: 2.27	G		Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_FAT	cellular response to stress	RT		5	3.2E-4	7.1E-2
<input type="checkbox"/> SP_PIR_KEYWORDS	Apoptosis	RT		4	1.3E-3	8.4E-2
<input type="checkbox"/> GOTERM_BP_FAT	response to DNA damage stimulus	RT		4	1.8E-3	1.9E-1
<input type="checkbox"/> GOTERM_BP_FAT	apoptosis	RT		4	7.0E-3	4.2E-1
<input type="checkbox"/> GOTERM_BP_FAT	programmed cell death	RT		4	7.3E-3	3.5E-1
<input type="checkbox"/> GOTERM_BP_FAT	cell death	RT		4	8.8E-3	3.4E-1
<input type="checkbox"/> GOTERM_BP_FAT	death	RT		4	9.4E-3	3.1E-1
<input type="checkbox"/> SP_PIR_KEYWORDS	nucleus	RT		5	2.4E-1	1.0E0

PC 2 - Top 20 Genes:

A_55_P2048478 A_30_P01024339 A_55_P1976127 A_55_P2085955 A_30_P01021547
A_55_P1976744 A_55_P2332194 A_51_P374137 A_55_P1978825 A_51_P440092
A_55_P2251974 A_55_P2077048 A_55_P1975973 A_55_P1971938 A_55_P2010066
A_30_P01032333 A_55_P2142789 A_55_P2105685 A_30_P01030419 A_51_P406165

Annotation Cluster 1	Enrichment Score: 0.74	G		Count	P_Value	Benjamini
<input type="checkbox"/> SP_PIR_KEYWORDS	calcium	RT		3	4.0E-2	8.4E-1
<input type="checkbox"/> GOTERM_MF_FAT	calcium ion binding	RT		3	6.8E-2	1.0E0
<input type="checkbox"/> GOTERM_MF_FAT	metal ion binding	RT		4	3.3E-1	1.0E0
<input type="checkbox"/> GOTERM_MF_FAT	cation binding	RT		4	3.4E-1	1.0E0
<input type="checkbox"/> SP_PIR_KEYWORDS	metal-binding	RT		3	3.4E-1	1.0E0
<input type="checkbox"/> GOTERM_MF_FAT	ion binding	RT		4	3.4E-1	1.0E0

PC 3 - Top 20 Genes:

A_30_P01028123 A_55_P2142251 A_55_P1979295 A_55_P2140107 A_52_P224358
A_55_P2069969 A_55_P1954718 A_51_P219505 A_30_P01020518 A_51_P193490
A_55_P2387665 A_55_P2167269 A_51_P172054 A_55_P2070869 A_55_P2198983
A_52_P585652 A_55_P1969882 A_51_P243755 A_52_P318673 A_55_P1975185

Annotation Cluster 1	Enrichment Score: 3.26	G	Count	P_Value	Benjamini
<input type="checkbox"/> INTERPRO	Glutathione S-transferase, alpha class	RT	3	1.1E-5	6.5E-4
<input type="checkbox"/> PIR_SUPERFAMILY	PIRSF000503:glutathione transferase	RT	3	1.4E-4	1.3E-3
<input type="checkbox"/> GOTERM_MF_FAT	glutathione transferase activity	RT	3	1.6E-4	8.0E-3
<input type="checkbox"/> INTERPRO	Glutathione S-transferase, N-terminal	RT	3	2.6E-4	7.5E-3
<input type="checkbox"/> INTERPRO	Glutathione S-transferase, C-terminal	RT	3	2.8E-4	5.4E-3
<input type="checkbox"/> INTERPRO	Glutathione S-transferase, C-terminal-like	RT	3	3.3E-4	4.6E-3
<input type="checkbox"/> INTERPRO	Glutathione S-transferase/chloride channel, C-terminal	RT	3	4.2E-4	4.8E-3
<input type="checkbox"/> GOTERM_MF_FAT	transferase activity, transferring alkyl or aryl (other than methyl) groups	RT	3	6.9E-4	1.7E-2
<input type="checkbox"/> INTERPRO	Thioredoxin fold	RT	3	4.0E-3	3.7E-2
<input type="checkbox"/> SP_PIR_KEYWORDS	transferase	RT	3	3.6E-1	1.0E0

PC Loadings

	Probes	PC1	PC2	PC3
1	A_52_P476431	-0.04528762	5.583179e-02	-0.0597389877
2	A_30_P01025243	-0.03434705	-8.125802e-02	-0.0569082347
3	A_55_P1973543	-0.05384122	-3.068907e-02	-0.0070028884
4	A_52_P258338	0.03895328	-6.742680e-02	-0.0054231538
5	A_30_P01033650	-0.02707184	1.913624e-02	-0.0764933207
6	A_55_P2049095	-0.04904902	-5.434188e-03	-0.0683781270
7	A_55_P2149500	-0.05446022	2.514847e-02	-0.0600152317
8	A_51_P334789	-0.05558848	-5.795805e-03	-0.0437355187
9	A_55_P2002849	-0.04984913	-3.237870e-02	-0.0531782862
10	A_30_P01029691	-0.02086555	2.890524e-02	-0.0094357252
11	A_66_P110769	0.04928890	-2.219134e-02	-0.0379325814
12	A_55_P2033272	0.05858591	-3.578278e-02	0.0154753173
13	A_55_P2064547	0.05310014	-1.701291e-02	0.0212526305
14	A_52_P653825	0.04475672	-5.737844e-02	-0.0052163214
15	A_30_P01026972	-0.02227645	7.276431e-02	-0.0115512050
16	A_55_P2029687	0.03352399	5.599928e-02	-0.0437494572
17	A_52_P11402	0.03995709	-7.891773e-03	-0.0097716918
18	A_51_P424338	0.04930828	2.870123e-02	-0.0778770476
19	A_55_P2165560	-0.04662606	5.668375e-02	-0.0224198838
20	A_55_P1965467	-0.04938535	2.608430e-02	-0.0652120467
21	A_55_P2056493	-0.04556470	6.202952e-02	-0.0550056135
22	A_55_P2380132	0.05438719	-3.126299e-02	-0.0652963011
23	A_30_P01024344	0.05476380	4.131172e-02	-0.0325337500
24	A_30_P01021631	-0.03626494	4.343281e-02	-0.0136509840
25	A_66_P119350	-0.04636976	-4.372779e-02	-0.0326775742
26	A_51_P155873	-0.02519181	-3.399461e-02	-0.0805817593
27	A_55_P2122504	0.04727055	3.486068e-02	-0.0577382798
28	A_55_P1964752	0.05412697	4.166556e-02	-0.0420657277
29	A_52_P449629	0.04963005	1.715138e-02	-0.0785620607
30	A_52_P614777	-0.02735213	-7.414542e-02	-0.0333510241
31	A_51_P263246	0.05212360	3.442563e-02	-0.0739239698
32	A_55_P2170454	0.04884463	-3.002980e-02	-0.0577522527
33	A_55_P2059931	0.03543591	-5.381103e-02	0.0722594767
34	A_30_P01026021	0.04238955	-6.141005e-03	-0.0111977946
35	A_55_P2065899	0.03834964	-5.501877e-02	0.0680931280
36	A_30_P01028123	0.04358730	-3.186757e-02	-0.0959672654
37	A_52_P366772	0.05528984	-2.922412e-02	-0.0508201203
38	A_55_P2017116	0.05573400	-1.292967e-02	-0.0666655681
39	A_30_P01030951	-0.04655478	-6.593120e-02	-0.0228972435
40	A_55_P1979639	0.04132323	-6.381691e-02	0.0132429615
41	A_51_P365019	0.04868167	-1.692388e-02	-0.0755683480
42	A_55_P2361647	-0.04705398	-4.092838e-02	-0.0688752985
43	A_30_P01029788	-0.02520259	-8.694527e-02	-0.0412093443
44	A_30_P01019873	0.04785145	-2.779291e-02	-0.0627621280
45	A_66_P114333	-0.04414267	-3.224942e-02	-0.0563392812
46	A_55_P2007326	0.03351804	-8.004432e-02	0.0438838232
47	A_52_P24439	0.05385076	-4.503464e-02	-0.0180101754
48	A_55_P2116496	0.04755832	3.466704e-02	0.0189412343
49	A_52_P561205	0.04021256	-2.922153e-02	-0.0743584664
50	A_55_P2048478	-0.02604343	-9.432245e-02	-0.0179366325

51	A_55_P2181665	0.03681245	-7.393896e-02	-0.0413701211
52	A_30_P01033583	0.04027225	-6.953305e-02	-0.0663883370
53	A_55_P1985433	0.02607973	2.493189e-02	-0.0589518482
54	A_55_P1960735	0.05597127	3.917516e-02	-0.0209724485
55	A_55_P2094896	-0.06058895	-1.872300e-03	-0.0038258769
56	A_30_P01031112	-0.04481110	-2.389403e-02	-0.0632707614
57	A_51_P377526	-0.05079995	-4.165155e-03	-0.0058881536
58	A_65_P06446	0.04560321	2.339040e-02	0.0187530198
59	A_51_P208121	0.05835656	9.373051e-04	0.0164250532
60	A_55_P2184009	0.04440943	5.204455e-02	-0.0067848623
61	A_51_P343350	0.02805136	-7.522900e-02	-0.0155530055
62	A_55_P2142251	0.02890839	-3.545044e-02	0.1078003070
63	A_51_P502614	0.05225514	-2.420633e-03	-0.0439772832
64	A_30_P01028766	0.05544466	9.509649e-04	-0.0566397319
65	A_55_P2108837	0.05398930	-2.256953e-02	0.0318206617
66	A_55_P1979295	-0.03668577	-6.511874e-02	-0.0958799263
67	A_30_P01021311	0.05481677	-2.487221e-02	-0.0710047488
68	A_52_P190763	0.05428914	-1.782528e-02	-0.0084581974
69	A_51_P516133	0.04406466	3.750779e-02	0.0069352825
70	A_55_P1969506	0.03003542	-7.949431e-02	0.0077438543
71	A_55_P2351193	-0.04425558	-5.676263e-02	-0.0030940995
72	A_52_P337427	-0.05770426	1.797945e-02	-0.0269407048
73	A_51_P175580	0.05458751	2.490452e-02	-0.0696209574
74	A_30_P01025676	-0.03217418	-7.809230e-02	-0.0103454745
75	A_55_P1967350	0.04780560	-6.721084e-02	-0.0209950589
76	A_55_P2140107	0.02861402	-6.045484e-02	0.0974583301
77	A_51_P153063	0.02997891	-8.305738e-02	0.0454633460
78	A_55_P2006236	0.05531510	-1.630881e-03	-0.0481765321
79	A_52_P224358	0.04493647	-3.877998e-02	-0.0933632175
80	A_30_P01028297	-0.03605453	-8.402917e-02	-0.0287389750
81	A_52_P627816	0.03760716	-5.193081e-02	0.0766279379
82	A_66_P128931	0.02760045	-4.830812e-02	0.0294694958
83	A_51_P421140	0.05975004	2.194595e-02	0.0172299716
84	A_55_P2360266	-0.04853527	-5.468767e-02	-0.0270692061
85	A_52_P664506	0.03274695	-7.210239e-02	0.0274312987
86	A_51_P324633	-0.04542671	-5.255171e-02	-0.0169855174
87	A_55_P2084332	0.05126847	-1.657693e-02	-0.0773079848
88	A_51_P311904	-0.04768661	3.196078e-02	-0.0726641705
89	A_55_P2069969	0.04458200	-3.370663e-02	-0.0972760408
90	A_51_P241995	0.03613004	5.298031e-02	-0.0812492390
91	A_55_P1954718	0.03472607	1.921373e-02	0.1115126657
92	A_55_P2177233	0.05874183	-1.907054e-02	0.0027743362
93	A_52_P211185	0.03049877	-6.979298e-02	0.0783564043
94	A_30_P01032002	0.05732469	3.372216e-02	-0.0315567564
95	A_55_P2079430	0.05695052	1.022132e-02	0.0029546274
96	A_66_P124179	0.03947115	-2.024030e-02	-0.0390179036
97	A_52_P612803	0.05471492	3.118753e-02	-0.0423709724
98	A_55_P1972541	-0.04646810	-1.061375e-02	-0.0115092049
99	A_55_P1959703	0.06035181	-2.188515e-02	0.0124321003
100	A_55_P2153291	0.06148778	-7.342749e-03	0.0040834963
101	A_30_P01028938	0.05516941	1.453249e-02	-0.0516313055
102	A_51_P515684	0.03828733	-7.626728e-02	-0.0403831750
103	A_51_P133137	0.03598724	7.406767e-02	-0.0127879465
104	A_55_P2059010	0.05430512	1.162252e-04	0.0241654467
105	A_55_P2129449	-0.04182691	-5.538413e-02	0.0334343447
106	A_51_P163578	0.04858356	-5.016860e-02	-0.0260507293
107	A_51_P185292	-0.05334916	-3.598913e-02	-0.0282690140
108	A_51_P219505	0.02601581	-7.385521e-03	0.1371573234
109	A_55_P2075258	-0.04918883	-5.629304e-02	0.0154777757
110	A_30_P01024339	0.02821320	-8.837273e-02	0.0023645164
111	A_55_P2144597	-0.05679568	7.694031e-04	-0.0445395423
112	A_55_P2078246	0.02883707	8.143065e-02	-0.0308109365
113	A_55_P2029106	-0.04867545	-1.460138e-02	0.0485825083
114	A_51_P353252	0.04091198	-7.390847e-02	0.0293349307
115	A_55_P2148171	-0.04744096	-3.400226e-02	-0.0276232083
116	A_51_P290207	0.01782468	-3.456896e-02	0.0450790746
117	A_55_P2015892	-0.04305742	-1.115988e-02	-0.0878401147
118	A_55_P2027836	0.05382592	4.512281e-02	-0.0134188694
119	A_51_P260300	0.04895463	-3.799508e-02	-0.0234555104
120	A_52_P286520	-0.04451041	-1.976812e-02	-0.0623325878
121	A_55_P2153292	0.06126838	1.346640e-02	0.0156273967

122	A_51_P162955	0.03996103	-4.459755e-02	0.0064075779
123	A_55_P2077783	0.05002896	1.088198e-02	0.0008548452
124	A_52_P550843	0.06071182	2.526322e-02	-0.0058905960
125	A_52_P176619	-0.05447293	-1.388964e-02	-0.0507457989
126	A_51_P283473	-0.05750639	1.080089e-03	0.0123610406
127	A_30_P01020518	0.04762918	1.609133e-04	-0.0962499449
128	A_55_P2033120	0.05549157	2.930660e-02	-0.0526876095
129	A_51_P439085	0.04338341	4.474387e-02	-0.0695717657
130	A_55_P1987196	0.04311092	5.724929e-02	-0.0261597108
131	A_55_P2066429	0.05708381	-1.752152e-02	0.0218374325
132	A_30_P01023273	0.05183772	-2.963267e-02	-0.0291952737
133	A_55_P1954698	-0.03973954	-3.980209e-02	0.0532490152
134	A_55_P2113466	0.02970203	3.655786e-03	-0.0805787611
135	A_55_P2016014	-0.04326957	-7.467055e-02	-0.0066528777
136	A_55_P2063654	0.04132036	-5.784277e-02	0.0671614341
137	A_51_P193490	-0.04237350	-1.079471e-02	-0.1137139514
138	A_55_P2131428	-0.04156016	-6.660496e-02	-0.0328462836
139	A_30_P01033086	-0.05175606	-6.758318e-02	0.0133720367
140	A_55_P1978666	0.04530374	-1.799432e-02	0.0817334131
141	A_51_P353895	-0.05121025	-3.979562e-02	-0.0520473242
142	A_55_P2165544	0.05494035	3.973797e-02	-0.0132211819
143	A_55_P2050652	0.05639413	-2.611965e-02	-0.0085159254
144	A_52_P308465	0.04915431	-6.630448e-03	0.0313714444
145	A_55_P1991841	0.03204346	-5.289433e-02	-0.0784515295
146	A_51_P423825	-0.05497270	-3.227607e-02	-0.0246461323
147	A_55_P2129696	-0.03120261	-5.585484e-02	-0.0471184147
148	A_55_P2062589	0.03777114	-5.910793e-02	-0.0293939384
149	A_55_P2387665	0.04257929	-3.901937e-02	0.1036628683
150	A_55_P2161545	0.05708909	-2.071726e-02	-0.0357098538
151	A_66_P116678	0.05148874	5.308941e-02	-0.0560390658
152	A_52_P646979	0.04923987	1.313859e-02	-0.0846501888
153	A_55_P2007470	0.06008578	9.044316e-03	0.0005863565
154	A_51_P481159	0.05518770	4.937684e-03	-0.0262403137
155	A_55_P2057577	0.04368993	-5.829287e-02	-0.0127731911
156	A_52_P683306	0.05655644	-2.090542e-03	-0.0156183216
157	A_30_P01018679	0.04401899	-2.493511e-02	-0.0135793299
158	A_51_P402160	0.04203287	-8.394292e-02	-0.0069952786
159	A_66_P137556	0.03921719	-6.572822e-02	-0.0758200941
160	A_55_P2014511	0.04702407	-4.366963e-03	0.0081430997
161	A_55_P2062250	-0.03836195	-8.120270e-02	-0.0376963779
162	A_52_P434974	0.05354292	-2.710365e-02	0.0522709362
163	A_52_P296913	0.04474289	-5.431977e-02	-0.0712051272
164	A_55_P1956557	-0.05508930	-3.592648e-02	-0.0223426818
165	A_55_P2057946	0.05350894	-3.690540e-03	0.0907095275
166	A_55_P2164075	-0.04632724	-5.210375e-02	-0.0501136106
167	A_52_P621617	0.05160106	-3.219994e-02	-0.0250632311
168	A_55_P1976127	0.03365943	8.926666e-02	0.0046938907
169	A_55_P2310898	-0.06069321	5.119870e-03	0.0320266009
170	A_55_P2135967	0.05215849	8.664376e-03	0.0190399927
171	A_30_P01032980	-0.04170791	-4.353409e-02	-0.0638050461
172	A_55_P2167269	-0.04197467	-2.130477e-02	-0.1048649507
173	A_55_P2100715	0.06089561	-6.320504e-03	-0.0403047683
174	A_66_P115389	0.04915336	-3.530487e-02	-0.0356446592
175	A_52_P456750	0.04573411	-1.553860e-02	0.0479325057
176	A_51_P228658	0.03678095	2.961837e-02	0.0799135220
177	A_66_P108152	0.05273927	-3.897222e-02	-0.0562721834
178	A_55_P2129309	0.05362689	-1.197630e-02	-0.0355856079
179	A_55_P2158990	0.06104268	1.786527e-02	-0.0277256353
180	A_30_P01018036	0.04927330	-2.235253e-02	-0.0226017751
181	A_55_P2027879	0.04812361	9.399620e-03	-0.0589836226
182	A_55_P22141299	0.05263431	-2.323987e-03	-0.0380084981
183	A_55_P2068988	0.05363369	-3.257906e-02	-0.0673775695
184	A_66_P130813	0.05468956	-2.900819e-02	-0.0459538099
185	A_55_P1973213	0.05536490	1.415459e-02	-0.0432725718
186	A_55_P2114342	0.04576728	-7.077958e-02	0.0217265760
187	A_55_P2052563	0.04436207	-5.677339e-03	-0.0520207862
188	A_51_P172054	0.01759975	9.313080e-03	0.1245047968
189	A_55_P2154416	0.05596210	1.712703e-02	-0.0394662937
190	A_55_P2394308	0.05736890	4.862639e-03	0.0361187197
191	A_52_P141112	0.04956567	-3.638803e-02	-0.0522633219
192	A_51_P296036	-0.05850041	2.975229e-02	0.0064869296

193	A_52_P71686	0.02541956	-5.645741e-02	-0.0572098971
194	A_51_P179919	0.04127969	-8.050928e-02	-0.0332983336
195	A_51_P411345	0.04649920	6.166474e-02	-0.0433463640
196	A_55_P2108903	0.03532870	-9.167010e-03	0.0673580354
197	A_30_P01026079	0.05661603	2.927123e-02	-0.0296011917
198	A_30_P01029783	0.05047194	2.899016e-03	-0.0431246974
199	A_30_P01023269	0.04083708	-1.116877e-03	-0.0682977425
200	A_51_P356283	-0.05148213	9.865906e-03	-0.0148539634
201	A_66_P126485	-0.05550451	-2.964705e-02	0.0061954591
202	A_52_P137371	0.03709953	-6.079231e-02	0.0109470141
203	A_51_P189361	0.05116993	-3.732495e-02	0.0100612319
204	A_55_P2430221	0.04669321	-4.897833e-02	0.0015395598
205	A_55_P2002577	0.04671968	2.850351e-02	-0.0761615064
206	A_52_P468068	0.05071591	-2.513121e-02	0.0179512347
207	A_51_P514405	-0.03552512	-8.288451e-02	0.0082490860
208	A_51_P234044	-0.05496948	-3.387895e-02	-0.0518554817
209	A_55_P2085955	-0.02957002	-9.150528e-02	-0.0155341199
210	A_55_P2137840	-0.03804472	-1.101087e-02	0.0792640241
211	A_51_P404377	0.04927773	2.307052e-02	-0.0039522392
212	A_55_P1999022	-0.04241309	-4.027025e-02	0.0142761056
213	A_51_P159610	0.04581039	5.588356e-02	-0.0519615138
214	A_30_P01020716	0.03893310	-6.076238e-02	-0.0486668264
215	A_52_P164161	0.04050975	-1.307055e-02	0.0280419796
216	A_52_P452667	0.03206060	-3.310538e-02	-0.0167811536
217	A_55_P2002854	-0.04535253	2.536971e-05	-0.0862431911
218	A_30_P01021547	-0.03578565	-8.949667e-02	0.0196779408
219	A_55_P2072908	0.05863391	1.728000e-02	0.0341864225
220	A_66_P113662	0.05377179	-1.603995e-02	0.0425048670
221	A_55_P1969131	0.04796658	-5.562506e-02	-0.0267385967
222	A_55_P1961567	-0.05251540	1.168073e-02	-0.0855548918
223	A_55_P1982818	0.04764204	1.682127e-02	-0.0245106499
224	A_51_P416613	-0.04751583	-4.175307e-02	0.0095173781
225	A_55_P2083180	0.04693370	-6.324071e-02	-0.0119748692
226	A_52_P151320	-0.04703496	-1.581995e-02	-0.0395699458
227	A_51_P171200	0.05165695	3.892475e-02	0.0328817250
228	A_55_P2030899	0.03744626	7.635178e-02	-0.0024252429
229	A_55_P2053988	0.03787339	-7.714088e-02	0.0527459387
230	A_52_P205255	-0.02971189	-3.152946e-02	0.0214894864
231	A_30_P01032529	0.05063223	-5.974025e-02	-0.0069938495
232	A_55_P2080956	0.05593041	-3.787288e-03	-0.0260992189
233	A_51_P117794	-0.03780521	-8.196266e-02	-0.0053697257
234	A_51_P320614	0.06094676	-1.057724e-02	-0.0110300814
235	A_51_P505662	0.05590003	-1.580494e-02	-0.0339054668
236	A_55_P2070869	0.01230067	1.189314e-02	0.1247686557
237	A_51_P329928	0.05959643	2.453028e-02	-0.0157697352
238	A_52_P311853	0.05870347	-3.058528e-02	-0.0228094783
239	A_51_P473953	0.03962652	-6.093586e-02	0.0280635183
240	A_55_P2025514	0.03606601	-6.830809e-02	0.0314401684
241	A_52_P235347	0.04816053	-2.051035e-02	-0.0466363271
242	A_52_P452689	0.05426157	3.645761e-02	-0.0505395259
243	A_55_P2164977	0.03194577	-8.554525e-02	0.0118224837
244	A_52_P93910	-0.05017808	6.662295e-03	0.0566538292
245	A_51_P131408	0.05892054	2.378904e-02	-0.0421092905
246	A_55_P1976744	0.03093243	-9.299978e-02	0.0146525993
247	A_55_P2009988	-0.05203195	-7.546213e-03	-0.0467283537
248	A_55_P2137406	0.05250557	5.945019e-02	-0.0098831589
249	A_51_P437240	-0.05674964	2.888211e-02	-0.0218572676
250	A_55_P2198983	0.02012803	-5.426060e-03	-0.1262384592
251	A_51_P142923	0.04735104	1.792012e-02	-0.0310663985
252	A_30_P01030510	-0.04423983	-4.424898e-02	-0.0456090707
253	A_55_P2019585	-0.05591466	3.234970e-02	-0.0368350287
254	A_52_P585652	0.04077218	-2.923247e-03	0.0927508992
255	A_51_P419637	-0.04558665	-5.005871e-02	0.0317894773
256	A_55_P2016462	0.05726279	3.298865e-03	-0.0371428824
257	A_55_P2153621	-0.05464960	7.725134e-03	-0.0388745256
258	A_51_P215438	-0.05290148	-2.517747e-02	-0.0098017998
259	A_55_P2207342	-0.05038893	-6.539008e-02	-0.0087148786
260	A_51_P111164	-0.05032633	4.597605e-02	0.0036348026
261	A_51_P360492	0.04425347	3.083300e-02	0.0526510369
262	A_52_P763938	0.04321294	-6.865954e-02	0.0119446668
263	A_55_P2066036	-0.03963238	-6.308697e-02	-0.0234266928

264	A_55_P2072906	0.05234174	4.055383e-02	0.0493382943
265	A_55_P2173737	0.05082589	-3.367886e-03	0.0609568657
266	A_55_P2032946	0.05565179	-2.108694e-03	-0.0496961560
267	A_52_P58059	0.04349416	-4.425935e-02	0.0212725532
268	A_55_P2102065	0.05429150	-6.815863e-04	-0.0670062543
269	A_55_P2163098	0.04690089	-6.361760e-02	0.0032379277
270	A_55_P2040860	-0.03843709	-8.218701e-02	-0.0294407418
271	A_55_P2019113	-0.05457951	-4.373945e-02	-0.0131088804
272	A_55_P2039284	0.05534335	-3.468243e-03	-0.0109160809
273	A_51_P505521	0.05327899	-7.802365e-03	0.0343960912
274	A_55_P2332731	0.03657549	-4.194750e-02	0.0003949764
275	A_30_P01024607	0.03983507	-4.490771e-02	-0.0507449623
276	A_55_P2001780	0.03218693	-6.710546e-02	0.0737566804
277	A_55_P2003513	0.05421241	-4.471996e-04	0.0447348675
278	A_55_P2067947	0.04721430	-2.652738e-02	-0.0673273575
279	A_51_P335480	0.04618211	3.092575e-05	-0.0796818375
280	A_55_P1974487	0.04130732	-7.918239e-02	-0.0005112731
281	A_55_P1977776	-0.04517869	-1.669810e-02	-0.0395194564
282	A_55_P2332194	-0.03002379	-8.820754e-02	0.0255232757
283	A_51_P372550	0.05085210	-7.081378e-03	0.0047313768
284	A_55_P2028171	0.05350686	-1.495241e-02	-0.0660266995
285	A_55_P2062981	0.04660596	-5.803574e-02	0.0129839288
286	A_55_P2151956	-0.02375338	-6.866575e-02	-0.0403117223
287	A_30_P01024929	0.05226113	-3.046269e-02	-0.0031125731
288	A_51_P369252	0.05986186	5.003553e-03	0.0052064285
289	A_52_P676271	0.03439050	-3.124749e-02	0.0654855764
290	A_55_P2080603	0.05299902	-1.805611e-02	0.0518860370
291	A_52_P647919	-0.05694880	-2.081583e-02	-0.0374419761
292	A_51_P479818	0.05100035	4.076589e-02	-0.0352489690
293	A_55_P2408588	0.03767557	1.546576e-02	-0.0099505045
294	A_51_P144531	0.04923240	-3.895301e-02	-0.0753536925
295	A_55_P2068663	0.04055686	4.054863e-02	0.0497038008
296	A_51_P482051	0.03628376	-4.535282e-02	0.0445180428
297	A_55_P2057528	-0.03890669	-5.277253e-02	-0.0590958393
298	A_51_P148612	-0.03250527	-3.904650e-03	-0.0301604167
299	A_30_P01032927	-0.04264574	-6.951336e-02	-0.0133975868
300	A_55_P2148071	0.02819371	2.206401e-04	-0.0688457421
301	A_51_P374137	0.02479548	-9.570080e-02	0.0509456495
302	A_52_P551011	-0.05715023	-2.686052e-02	-0.0020588222
303	A_52_P78373	-0.04934500	2.573453e-02	-0.0313472241
304	A_51_P212038	0.04790210	3.709623e-02	-0.0005136008
305	A_30_P01031812	0.04153570	-3.697322e-02	-0.0366861100
306	A_52_P326664	-0.03894398	-6.876849e-02	-0.0812300927
307	A_55_P2031045	-0.05227166	9.731279e-03	0.0112323741
308	A_55_P2177410	-0.03094496	-7.282760e-02	-0.0115729338
309	A_55_P2105632	-0.04415326	-7.350115e-02	-0.0060703264
310	A_55_P1958678	0.05737983	-1.745566e-02	-0.0275063696
311	A_30_P01032251	-0.03199713	3.310610e-02	0.0217802114
312	A_30_P01031026	0.05930346	-5.268748e-03	0.0022613246
313	A_55_P2274378	-0.03562565	-7.161619e-02	-0.0428628171
314	A_55_P1978825	-0.02463602	-9.749924e-02	0.0307823208
315	A_55_P2125588	0.06146443	6.591418e-03	0.0098730920
316	A_55_P2042297	0.05235348	-1.596681e-02	-0.0644759658
317	A_55_P2075263	0.03359987	-7.310378e-02	0.0701211713
318	A_55_P2150757	0.05126064	3.271251e-02	-0.0904348166
319	A_55_P2002578	0.05674182	-2.031801e-02	-0.0431390659
320	A_51_P213334	-0.03514935	-7.552083e-02	-0.0343520618
321	A_52_P489295	0.05607236	3.011387e-03	-0.0111823695
322	A_55_P2087984	0.05061153	-3.183731e-03	0.0393069958
323	A_51_P134812	-0.04211200	-4.739497e-03	0.0381439105
324	A_55_P2020248	0.05649342	-2.040424e-02	0.0561489942
325	A_30_P01025798	-0.02940313	2.697397e-02	0.0243437405
326	A_55_P1979728	0.05649622	-3.749486e-03	0.0286218971
327	A_51_P440092	0.03268465	-8.967959e-02	0.0632185718
328	A_52_P193194	0.04204135	-5.402599e-02	-0.0572689707
329	A_52_P484838	-0.05540410	8.304198e-03	0.0287402101
330	A_55_P2251974	-0.03031174	-9.036913e-02	0.0425296246
331	A_30_P01025159	-0.05159381	-6.010594e-02	0.0218353750
332	A_55_P2238059	-0.03123161	-3.179698e-02	-0.0549466854
333	A_55_P2157620	0.04178482	-7.878093e-02	0.0134864475
334	A_52_P273821	0.05236872	1.080439e-02	0.0020569489

335	A_55_P2026275	0.03600401	-7.111014e-02	-0.0081640471
336	A_51_P263965	0.04919446	8.997403e-03	-0.0385713043
337	A_52_P627269	0.05525204	-3.171634e-02	-0.0525060811
338	A_55_P2006008	-0.04628643	2.086142e-02	-0.0628821483
339	A_55_P2145804	0.06345368	7.203129e-03	-0.0047357090
340	A_55_P2077048	0.02960723	-9.005628e-02	-0.0024148913
341	A_55_P2062190	0.04715854	-4.246415e-02	-0.0820549403
342	A_51_P343517	0.04692768	2.080936e-02	0.0304745415
343	A_30_P01025511	0.05744600	3.031173e-02	-0.0181386576
344	A_51_P496751	0.05610236	1.906627e-02	0.0214898496
345	A_30_P01032430	0.03946476	-4.906325e-02	-0.0492327010
346	A_55_P2420694	0.05235983	-2.801721e-02	-0.0336872927
347	A_51_P219483	0.05427639	-7.426708e-03	-0.0741820226
348	A_55_P1975973	-0.02835304	-8.893589e-02	-0.0182793501
349	A_55_P1998656	0.05281888	-1.426609e-02	-0.0724227253
350	A_55_P1956812	0.04721402	-5.249342e-02	-0.0418034294
351	A_51_P457664	-0.03740432	-6.156877e-02	-0.0800265948
352	A_55_P1983754	-0.03351181	-8.353949e-02	0.0120039542
353	A_55_P2092219	0.04518634	-5.910487e-02	0.0339802015
354	A_55_P1976351	-0.03008328	-4.378654e-02	0.0670589106
355	A_66_P106760	0.05443099	-2.877974e-02	-0.0213721357
356	A_52_P318361	0.05362690	-5.132038e-02	-0.0395701915
357	A_55_P2399499	-0.05256774	-5.504699e-02	-0.0413342177
358	A_55_P1971938	0.02785647	-9.179051e-02	0.0067222330
359	A_51_P447545	0.04033573	-7.437077e-02	0.0019926984
360	A_55_P2069721	0.05081772	5.356083e-03	0.0663687592
361	A_55_P2141860	0.06361823	8.581611e-03	-0.0038232300
362	A_51_P456208	0.04126868	-5.858924e-02	0.0172812026
363	A_51_P382928	0.05566286	-1.661534e-02	-0.0042092005
364	A_55_P1957038	0.04267672	-5.684963e-02	0.0091106844
365	A_51_P256682	0.05795716	-3.520811e-02	0.0262250608
366	A_55_P1969665	0.04995373	-3.626250e-02	0.0131894140
367	A_55_P2255449	-0.04016244	1.429310e-02	-0.0297723046
368	A_55_P2114498	0.03918144	6.849948e-02	-0.0036203487
369	A_51_P415220	0.06083191	1.199406e-02	0.0069976335
370	A_55_P2126950	-0.05790567	-2.449666e-02	-0.0360511531
371	A_30_P01029662	-0.04567949	-6.930581e-02	-0.0169481704
372	A_55_P2010066	-0.03224536	-8.918752e-02	-0.0097817759
373	A_55_P2038183	0.03378350	-8.483328e-02	0.0416669120
374	A_30_P01024055	-0.04237981	-4.066309e-02	-0.0598897214
375	A_55_P2126951	-0.05160402	-3.102883e-02	-0.0598191919
376	A_30_P01022593	-0.05403974	-4.760462e-02	0.0412020512
377	A_51_P234692	-0.03731400	1.475171e-02	0.0226416598
378	A_51_P357735	0.04342995	-5.614698e-02	0.0136296088
379	A_55_P1969882	0.04111286	-4.478587e-02	0.1052245997
380	A_55_P2131238	0.05496830	3.448934e-03	0.0642143563
381	A_30_P01020135	-0.03296983	3.294009e-02	0.0261480368
382	A_55_P1992849	-0.03789642	-8.348080e-02	-0.0218882290
383	A_55_P1972322	0.05257307	3.274643e-02	0.0121006569
384	A_55_P1961302	0.05675543	3.693612e-02	-0.0343265892
385	A_30_P01031079	0.04777362	-4.136304e-02	-0.0148579958
386	A_51_P256384	0.02932282	-8.085745e-02	0.0239772982
387	A_51_P336385	0.05269588	5.657411e-03	-0.0174578073
388	A_30_P01032333	-0.01841989	-1.033667e-01	-0.0108073977
389	A_51_P137452	-0.02857229	-8.768357e-02	-0.0506872671
390	A_51_P269792	-0.04643030	-1.698431e-02	0.0206831164
391	A_55_P1956223	0.04834677	-1.419654e-02	-0.0173029631
392	A_55_P2083629	-0.05646259	1.296872e-02	-0.0481296935
393	A_55_P1976898	0.03569552	-8.607991e-02	-0.0049656629
394	A_55_P1968355	0.04818467	1.166782e-03	0.0172297170
395	A_51_P213691	-0.04376321	-7.229468e-02	-0.0083075494
396	A_51_P220806	0.05766941	-3.540377e-02	-0.0528858804
397	A_30_P01029720	-0.03288525	1.359946e-02	0.0274233340
398	A_51_P161354	0.05642407	1.921329e-02	-0.0095365029
399	A_30_P01018128	-0.05572183	-4.431160e-02	-0.0309136246
400	A_55_P2092909	0.05768257	2.305598e-02	-0.0193609414
401	A_52_P14456	0.04623345	-3.033972e-02	-0.0113055893
402	A_51_P233160	0.05419145	3.089235e-02	0.0276715933
403	A_51_P490023	0.05359329	-2.442672e-02	0.0072079620
404	A_52_P363216	0.04023225	-6.740847e-02	0.0378674211
405	A_55_P2019054	-0.03210549	-5.445373e-02	0.0865392556

406	A_55_P2047614	0.05323088	-3.644079e-02	0.0318647526
407	A_52_P617327	0.05647644	2.144710e-02	0.0047483596
408	A_55_P2156625	0.05577499	-2.637275e-03	0.0252991283
409	A_55_P2142789	-0.02412687	-1.000368e-01	0.0158211755
410	A_55_P2124712	0.05370222	-4.317670e-02	-0.0403113054
411	A_55_P2107070	0.05504217	-1.946084e-02	-0.0301213990
412	A_51_P109449	0.03984656	-6.884692e-02	0.0586713957
413	A_51_P488554	-0.04756411	-3.381889e-02	0.0158685016
414	A_55_P2088750	-0.04668963	-7.246924e-02	-0.0008319486
415	A_51_P464146	0.04378449	-7.543782e-02	0.0178034235
416	A_55_P2321889	0.03070447	-6.563893e-02	-0.0084706566
417	A_55_P2030155	-0.02767758	-5.635896e-02	-0.0752826520
418	A_55_P1988148	-0.05522419	-1.622434e-02	-0.0336063260
419	A_55_P2018017	-0.04121847	-7.240415e-02	-0.0132460349
420	A_55_P2273656	-0.03823916	-6.479508e-02	0.0355979808
421	A_51_P243755	0.04585913	-4.090454e-02	0.1022818904
422	A_52_P318673	0.01104849	-2.292234e-02	0.1502678303
423	A_55_P2124976	0.05314264	-2.036059e-02	-0.0238211262
424	A_51_P246653	0.03724273	-4.823954e-03	-0.0129204983
425	A_30_P01019037	0.03014990	-3.112103e-02	0.0462582773
426	A_30_P01018914	0.05706635	-6.293171e-03	-0.0441693123
427	A_55_P2161347	0.04990842	-4.751454e-02	0.0499644759
428	A_51_P492742	-0.05517742	-4.913854e-02	0.0286961265
429	A_55_P2159485	-0.04790185	-6.960020e-03	0.0457670601
430	A_55_P1964228	-0.05563212	-3.018607e-02	-0.0231004272
431	A_55_P1982817	0.05514472	-1.016348e-02	0.0284459868
432	A_55_P2095039	-0.05193567	-4.498779e-02	-0.0168118335
433	A_51_P164835	0.04814098	1.213890e-02	0.0140466186
434	A_55_P2105685	-0.02630430	-9.019272e-02	0.0201224756
435	A_51_P496309	-0.02464830	-8.069294e-02	0.0098953721
436	A_30_P01030419	-0.02414191	-9.080543e-02	0.0081921440
437	A_51_P455647	0.05429275	-5.528985e-02	-0.0268869053
438	A_66_P103792	0.06126898	1.812247e-02	0.0048672391
439	A_55_P2035804	-0.05070139	-3.975912e-02	-0.0347176534
440	A_30_P01029006	0.03224337	-1.438396e-02	0.0411684337
441	A_55_P2002903	0.04750840	-2.284762e-02	-0.0103462554
442	A_55_P2162970	-0.02804775	-8.062155e-02	0.0408779167
443	A_55_P2032363	0.05211400	-3.250058e-02	0.0529073573
444	A_55_P2015182	-0.03365782	-8.672392e-02	-0.0222634626
445	A_55_P1985544	-0.03584759	-8.201987e-02	-0.0058703255
446	A_52_P334670	0.04479466	-5.444752e-02	0.0272104008
447	A_30_P01017543	0.04981283	-5.748091e-02	0.0141101153
448	A_55_P2031436	0.04404087	-2.624025e-02	0.0431787263
449	A_55_P2017347	0.01304247	-3.283773e-02	-0.0269167200
450	A_52_P137331	0.04755747	-5.387583e-02	-0.0053884066
451	A_66_P138004	0.04890001	-4.269114e-02	-0.0351532012
452	A_51_P323521	-0.05676691	-2.910820e-03	-0.0152421299
453	A_30_P01027568	-0.04918833	-6.017344e-02	0.0274406054
454	A_51_P406165	-0.02596291	-9.980437e-02	-0.0071595118
455	A_55_P2032966	0.02891024	-2.835226e-02	0.0764601625
456	A_55_P1975185	0.02740492	-1.749091e-02	0.1388205438
457	A_51_P126437	0.05421245	2.136281e-03	-0.0503242879
458	A_51_P306017	0.04016656	-7.724887e-02	-0.0524115643
459	A_51_P363947	0.06014630	2.962054e-02	-0.0395750634
460	A_51_P510891	0.05935061	3.607306e-03	-0.0272545149
461	A_52_P108346	0.04402219	2.256885e-02	-0.0210524347
462	A_52_P228236	0.04400262	7.135341e-02	0.0215480784
463	A_52_P237077	0.02747765	-7.563440e-02	0.0628476414

Table C1 List of relevant correlated disease as determined by a meta-analysis of published gene expression studies using the NextBio Human Disease Atlas.

Disease Group	Disease Phenotype	Number of studies	Direction of correlation	Score	
Liver and Biliary Diseases	Injury of liver	5	positive	97.4	
	Liver regeneration	5	positive	92.0	
	Disorder of iron metabolism	8	positive	83.4	
	Hepatic fibrosis	9	positive	82.2	
	Liver transplant disorder	1		80.9	
	Hepatocellular dysplasia	2	positive	76.6	
	Steatosis of liver	4	negative	74.5	
	Non-alcoholic fatty liver	3	negative	70.3	
	Inflammatory disease of liver	27	positive	68.5	
	Congenital biliary atresia	2		68.5	
	Cirrhosis of liver	8	positive	52.1	
	Primary sclerosing cholangitis	1		41.2	
	Cancer	Breast cancer	333	positive	79.4
		Malignant tumour of hypopharynx	1	positive	76.6
Liver cancer		111	positive	73.0	
Adrenal cancer		10	positive	69.3	
Kidney cancer		49	positive	67.8	
Lung cancer		158	positive	64.9	
Prostate cancer		125	positive	64.0	
Brain cancer		149	positive	62.9	
Gastric cancer		32	negative	61.4	
Testicular cancer		10	positive	61.3	
Lymphoma		170	positive	60.6	
Other cancer		133	negative	60.6	
Malignant tumour of muscle		37	positive	60.2	
Cancer of head and neck		38	positive	59.6	
Secondary malignant neoplastic disease		164	positive	58.7	
Esophageal cancer		24	positive	56.7	
Leukemia		225	negative	56.1	
Skin cancer		98	positive	55.8	
Ovarian cancer		68	negative	55.4	
Lymphoid leukemia		107	positive	55.3	
Uterine cancer	67	positive	55.1		

	Primary malignant neoplasm of bone	58	positive	55.0
	Bladder cancer	24	positive	53.8
	Malignant tumour of intestine	161	positive	51.5
	Myeloid leukemia	122	negative	50.9
	Retinoblastoma	6		47.0
	Neuroendocrine tumour	59	positive	46.6
	Thyroid cancer	13	positive	46.1
	Pancreatic cancer	40	positive	41.9
	Cancer of thymus	2	positive	30.0
<hr/>				
Autoimmune Diseases	Psoriasis	21	positive	61.0
	Sjögren's syndrome	7		56.5
	Rheumatoid arthritis	32	negative	49.9
	Polymyositis	8	positive	49.7
	Systemic sclerosis	9	positive	38.2
	Lupus erythematosus	15	negative	37.6
	Crohn's disease	10	negative	35.3
	Graves' disease	3		19.0
	Diabetes mellitus type 1	18	negative	17.9
	Autoimmune endocrine disease	3		16.7
<hr/>				
Hypersensitivity Disorders	Inflammatory disorder	1		68.0
	Allergic disorder	82	positive	58.2
	Delayed hypersensitivity disorder	30	negative	52.8
<hr/>				
Nutritional and Metabolic Diseases	High fat diet	43	positive	91.5
	Ischemic reperfusion injury	10	positive	89.9
	Obesity	29	positive	87.1
	Ketogenic diet	4	positive	84.4
	Proline dehydrogenase deficiency	2		77.3
	alpha, alpha-Trehalase deficiency	1		76.4
	Deficiency of oxygenase	1		74.8
	Iron overload	4	negative	73.5
	Hypervolemia	2	positive	71.4
	Cytochrome-c oxidase deficiency	1		71.4
	Triglyceride storage disease with ichthyosis	1		69.5

Nonpersistence of intestinal lactase	1		68.3
Deficiency state	53	negative	68.0
Hypercholesterolemia	3		64.9
Mucopolysaccharidosis	4	negative	64.3
Hypophosphatemia	2	negative	61.9
Zinc deficiency	3	negative	60.0
3-Methylglutaconic aciduria	1		58.1
Dehydration	9	positive	55.2
Complex V deficiency	1	positive	54.7
Hyperlipidemia	6		54.3
Metabolic syndrome X	1		54.1
Congenital glucose-galactose malabsorption	2	negative	53.3
Hyperoxia	3	positive	50.1
Pseudohypoaldosteronism	1		48.8
Adrenoleukodystrophy	1		45.9
Deficiency of multiple nutrient elements	1		38.1
Disorder of peroxisomal function	1		26.8
Diabetes mellitus type 1	18	negative	17.9
Goiter	3		14.7

Table C2 Full list of correlated compounds as determined by a meta-analysis of published gene expression studies using NextBio.

Compound Group	Compound	Number of studies	Direction of Correlation	Score
Air Pollutants	Ozone	3	positive	54.0
	Vehicle Emissions	4	positive	37.9
	bisphenol A	15	positive	35.9
	Particulate Matter	5	positive	30.0
	trichlorofluoromethane	1	negative	23.6
Alkylating Agents	Diethylnitrosamine	16	positive	56.7
	Carmustine	6	positive	44.7
	Tinidazole	3	positive	41.8
	Lomustine	10	positive	40.7
	Mitomycin	4	positive	37.4
	perfosfamide	1	positive	37.1
	Chlorambucil	11	positive	32.6
	Cyclophosphamide	17	positive	28.6
	Estramustine	1		27.2
	Altretamine	6	positive	26.2
	Methyl Methanesulfonate	5	positive	22.4
	esperamicin A1	1		21.3
	Busulfan	6	positive	19.5
	Ifosfamide	9	positive	18.7
	triptolide	1	negative	18.1
	Methylnitrosourea	2	positive	17.0
temozolomide	1	positive	3.1	
Melphalan	5	positive	0.0	
Antimetabolites	Ethionamide	4	positive	57.2
	Stavudine	6	positive	52.1
	compactin	3	positive	46.1
	Ethionine	6	positive	46.0
	Buthionine Sulfoximine	3	positive	44.4
	Lovastatin	11	positive	43.2
	fluvastatin	10	positive	42.8
	Azathioprine	13	positive	40.4
	Thiouridine	1		40.2
	Clofibrilic Acid	6	negative	40.0
	Pravastatin	5	positive	39.8
	Gemfibrozil	13	positive	38.7
	Choline	3	positive	38.7
	atorvastatin	16	negative	37.2

Propylthiouracil	10	positive	35.3
6-Mercaptopurine	6	positive	32.2
Cerivastatin	9	positive	31.9
pirinixic acid	12	positive	31.5
Allopurinol	6	positive	31.2
Simvastatin	17	positive	31.1
Thioguanine	6	negative	30.8
Torcetrapib	1		30.6
Puromycin			
Aminonucleoside	1	negative	30.5
Niacin	3	positive	30.1
Nafenopin	4	positive	30.0
Isoniazid	5	positive	30.0
Rosuvastatin	1		30.0
Bezafibrate	11	negative	29.7
Decitabine	21	positive	27.2
Doxifluridine	2	positive	26.9
Floxuridine	1	positive	26.7
Metyrapone	1	positive	26.4
Methotrexate	8	negative	26.3
Cytarabine	16	negative	24.9
Clofibrate	15	negative	23.8
Azaguanine	1	positive	21.9
LG 268	1	negative	21.9
Zalcitabine	4	positive	21.7
Capecitabine	4		21.1
Zidovudine	5	positive	17.9
Azacitidine	5	positive	16.5
Deoxyglucose	3	positive	16.0
Fluorouracil	11	positive	15.5
Azaauridine	3	positive	14.8
Benfluorex	1	positive	9.0
Trifluridine	1	positive	8.2
Flucytosine	1	positive	7.7
Puromycin	2	positive	7.1
Cycloserine	1	negative	6.9
6-thioguanosine	1	negative	6.8
Didanosine	1	positive	6.1
Ribavirin	6	positive	0.0
Carcinogens			
Thioacetamide	10	positive	70.9
1,5-naphthalenediamine	3	positive	67.8
N-nitrosomorpholine	2	positive	65.9
perfluorooctane sulfonic	4	positive	61.3

acid			
2-Acetylaminofluorene	7	positive	59.5
Dimethylnitrosamine	4	positive	56.6
perfluorooctanoic acid	4	positive	56.6
Benzo(a)pyrene	6	positive	53.8
Methylene Chloride	3	positive	50.8
Safrole	3	positive	48.3
nitrosobenzylmethylamine	2	positive	48.0
aristolochic acid I	3	positive	47.1
9,10-Dimethyl-1,2-benzanthracene	1		46.3
pristane	2	positive	45.7
Ethylene Dibromide	2	positive	44.5
2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine	1	positive	42.0
hydrazine	4	positive	41.5
2,4-diaminotoluene	1	positive	41.3
Diethylstilbestrol	16	positive	41.0
Methylcholanthrene	6	positive	40.5
Azoxymethane	4	positive	40.4
phenethyl isothiocyanate	2	positive	37.9
1-amino-2,4-dibromoanthraquinone	1	positive	32.4
pirinixic acid	12	positive	31.5
4-dichlorobenzene	4	positive	30.9
Nafenopin	4	positive	30.0
Dimethylformamide	8	positive	29.0
2,2-bis(bromomethyl)-1,3-propanediol	1	positive	27.1
Carcinogens	2	positive	26.4
methyleugenol	2	positive	26.3
Ultraviolet Rays	4	positive	25.4
4,4'-diaminodiphenylmethane	1	positive	25.3
bromodichloromethane	2	positive	23.8
7,8-Dihydro-7,8-dihydroxybenzo(a)pyrene			
9,10-oxide	3	negative	22.7
Urethane	4	positive	22.1
X-Rays	11	positive	21.2
ferric nitrilotriacetate	1	positive	20.7
ochratoxin A	4	negative	17.8
4-O-methyl-12-O-	1	positive	13.3

	tetradecanoylphorbol 13- acetate			
	Tetradecanoylphorbol Acetate	2	positive	10.5
	Benzene	2	positive	10.5
	pronethalol	1	negative	10.5
	phorbolol myristate acetate	5	positive	8.5
	mono-(2- ethylhexyl)phthalate	2	positive	6.3
	Asbestos	3	positive	0.0
	Phorbol Esters	1	negative	0.0
Cardiotoxins	Monocrotaline	11	positive	56.1
	Chloroform	4	positive	47.4
	acodazole	1	positive	40.0
	tripterine	2	positive	30.2
	Cardiotoxins	3	positive	29.8
	Diethylhexyl Phthalate	6	positive	22.5
	Strophanthidin	1	positive	15.7
Caustics	Potassium Dichromate	1	positive	46.9
	Hydrochloric Acid	2	negative	16.8
	Trichloroacetic Acid	2	positive	12.4
Endocrine Disruptors	perfluorooctane sulfonic acid	4	positive	61.3
	perfluorooctanoic acid	4	positive	56.6
	4-nonylphenol	5	negative	30.3
	Diethylhexyl Phthalate	6	positive	22.5
Insecticides	Malathion	3	positive	100.0
	2-dichlorobenzene	2	positive	63.6
	1,3-dichlorobenzene	1	positive	52.9
	closantel	3	positive	47.1
	Ethylene Dibromide	2	positive	44.5
	amitraz	3	positive	42.0
	4-dichlorobenzene	4	positive	30.9
	phosphine	1		27.7
	Dichlorvos	4	positive	21.9
	linalool	1	positive	21.0
	Anabasine	1	positive	18.3
	pentachlorobenzene	1	positive	16.5
	Ivermectin	3	negative	15.8
	Trichlorfon	1	negative	13.5
	Diazinon	2	positive	13.4
	Pyrethrins	1	positive	12.0

	Chlorpyrifos	4	negative	11.8
	Abamectin	3	negative	9.7
	Rotenone	3	positive	8.1
Mutagens	N-nitrosomorpholine	2	positive	65.9
	1,3-dichloro-2-propanol	1	positive	54.6
	Benzo(a)pyrene	6	positive	53.8
	aristolochic acid I	3	positive	47.1
	2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine	1	positive	42.0
	2,4-diaminotoluene	1	positive	41.3
	pirinixic acid	12	positive	31.5
	Cyclophosphamide	17	positive	28.6
	hydroxyhydroquinone	1	positive	27.9
	tris(2,3-dibromopropyl)phosphate	1	positive	27.4
	methyleugenol	2	positive	26.3
	7,8-Dihydro-7,8-dihydroxybenzo(a)pyrene 9,10-oxide	3	negative	22.7
	Methyl Methanesulfonate	5	positive	22.4
	ferric nitrilotriacetate	1	positive	20.7
	Patulin	2		10.6
	hydroquinone	1	negative	6.8
Neurotoxins	bromobenzene	8	positive	58.1
	Lead	3	positive	51.2
	Methylene Chloride	3	positive	50.8
	lead tetraacetate	8	positive	48.1
	Carbon Tetrachloride	14	positive	43.0
	transplatin	3	negative	33.7
	daboiatoxin	1	positive	33.4
	lead acetate	5	positive	32.5
	Clioquinol	1	positive	19.4
	N-acetyl-1-aspartylglutamic acid	1	positive	12.5
	1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine	4	positive	11.9
	Tetanus Toxin	1	negative	11.0
	Benzene	2	positive	10.5
Oxidants	Ozone	3	positive	54.0
	Hypochlorous Acid	1		40.9
	Hydrogen Peroxide	18	positive	40.8
	phenylhydrazine	2	positive	39.2
	cumene hydroperoxide	1	positive	8.0

Poisons	Phosgene	1	positive	61.4
	Sarin	2	negative	58.6
	Monocrotaline	11	positive	56.1
	Mustard Gas	2	positive	55.5
	Aflatoxin B1	7	positive	54.5
	Lead	3	positive	51.2
	lead tetraacetate	8	positive	48.1
	Cadmium	6	positive	44.7
	Ethylene Dibromide	2	positive	44.5
	Soman	1	positive	43.0
	Carbon Tetrachloride	14	positive	43.0
	Acrolein	5	negative	41.2
	Mycotoxins	1	positive	40.2
	vinylidene chloride	4	positive	40.1
	lead acetate	5	positive	32.5
	Alpha-Amanitin	2	positive	29.0
	naphthalene	2	positive	28.0
	4-nitrophenol	2	negative	26.9
	lycorine	1	positive	24.7
	Cholera Toxin	2	positive	23.3
	Phalloidine	1	negative	21.8
	Amanitins	2	negative	20.5
	Shiga Toxin	1	positive	19.8
	Strophanthidin	1	positive	15.7
	3-nitrophenol	2	negative	14.3
	Patulin	2		10.6
	solasodine	1	negative	9.7
	Ethylene Glycol	4	positive	7.9
	vanadium pentoxide	2	positive	0.0
	Pyrogens	Pyrogens	2	positive
Unclassified Toxicity	Methapyrilene	10	positive	63.5
	Nickel	2	positive	61.8
	methylformamide	1	positive	58.8
	Aminosalicylic Acid	1	positive	58.0
	Nitrofurazone	3	positive	57.8
	Diazepam	12	positive	57.4
	artemisinine	3	positive	57.0
	2-(4-morpholinoanilino)-6-cyclohexylaminopurine	1	negative	56.6
	Carbamazepine	7	positive	55.7
	oxidized-L-alpha-1-palmitoyl-2-arachidonoyl-	3	positive	55.6

sn-glycero-3-phosphorylcholine			
pralidoxime	5	positive	55.3
gedunin	2	positive	54.9
Clonazepam	4	positive	54.5
anastrozole	5	positive	54.2
ONO 2235	2	positive	54.2
15-deoxy-delta(12,14)-prostaglandin J2	1	positive	53.9
Cephalothin	5	positive	53.7
Estriol	6	positive	53.3
withaferin A	1	positive	52.9
Atropine	5	positive	52.8
Dipyron	3	positive	50.8
Disulfiram	6	positive	50.2
securinine	1	positive	50.1
sulforafan	3	positive	50.0
tosufloxacin	3	negative	50.0
eperisone	2	positive	49.8
Amphetamine	1	positive	49.0
fasudil	1	positive	48.8
pantoprazole	5	positive	48.6
Cinnarizine	4	positive	48.5
coumarin	5	positive	48.5
Bromhexine	2	positive	47.8
ferulic acid	2	positive	47.8
benzyloxycarbonylleucyl-leucyl-leucine aldehyde	2	positive	47.8
Cisplatin	28	positive	47.3
Cephaloridine	1		47.3
Aminoglutethimide	4	positive	47.3
Ticlopidine	6	positive	47.3
Iron	5		47.2
Praziquantel	4	positive	47.0
Ethambutol	7	positive	46.9
Mestranol	4	positive	46.9
8-aminohexylamino cAMP	1	positive	46.9
Methimazole	8	positive	46.6
Omeprazole	11	positive	46.5
Danazol	9	positive	46.3
Cyproterone Acetate	5	positive	46.3
bendazolic acid	1	positive	46.2

Ticrynafen	3	positive	45.9
Formaldehyde	6	positive	45.7
Methyltestosterone	8	positive	45.4
Ganciclovir	2	positive	45.3
Ethinyl Estradiol	13	positive	45.2
Oxazepam	1	positive	45.2
ipriflavone	2	positive	45.0
Estradiol	79	positive	44.6
salicylamide	4	positive	44.5
Hydrogel	1		44.4
rabeprazole	5	positive	44.3
lansoprazole	5	positive	44.3
Phosphorylcholine	1	positive	43.9
Chlormezanone	3	positive	43.9
Furosemide	6	positive	43.7
Vinblastine	10	positive	43.6
tranilast	4	positive	43.3
crotamiton	4	positive	43.2
Raloxifene	6	negative	43.1
Flutamide	8	positive	43.0
zileuton	4	positive	42.9
Phenobarbital	16	positive	42.8
Colchicine	9	positive	42.7
Thiostrepton	1	positive	42.7
N,N'-diphenyl-4-phenylenediamine	3	positive	42.7
systhane	5	positive	42.7
Pregnenolone Carbonitrile	2	positive	42.3
trilinolein	2	positive	42.1
Pyrazinamide	4	positive	42.1
Clotrimazole	8	positive	42.0
Benzalkonium Compounds	1	positive	41.9
Oxyphenbutazone	1	positive	41.8
Promegestone	1	negative	41.8
Hyaluronic Acid	1	positive	41.7
Dinoprostone	3	positive	41.6
piperlonguminine	1	positive	41.6
gefitinib	6	negative	41.5
flubendazole	3	positive	41.5
Fenofibrate	23	negative	41.4
ebesen	1	positive	41.3
Bleomycin	11	positive	41.2

Clomiphene	6	positive	41.0
erlotinib	5	negative	40.8
phenothiazine	4	positive	40.6
Lorazepam	4	positive	40.6
benazepril	2	positive	40.6
HI 6	1	positive	40.5
Tolazamide	6	positive	40.5
Enalapril	6	positive	40.4
marimastat	2	positive	40.3
balsalazide	3	positive	40.2
Phenacetin	8	positive	40.0
Levamisole	1	positive	40.0
Adenosine-5'-(N-ethylcarboxamide)	1	positive	40.0
zopiclone	2	positive	39.9
bacterial lysate	1	positive	39.8
Roxarsone	4	positive	39.8
shikonin	3	positive	39.7
letrozole	10	positive	39.5
Doxorubicin	33	positive	39.3
2-(1H-indazol-4-yl)-6-(4-methanesulfonylpiperazin-1-ylmethyl)-4-morpholin-4-ylthieno(3,2-d)pyrimidine	2	negative	39.2
Hypericum extract LI 160	1	negative	39.2
Acetaminophen	18	positive	39.1
Methocarbamol	3	positive	39.0
octylphenol	1		39.0
Trichloroethylene	3	positive	38.8
Gentamicins	13	positive	38.7
Tunicamycin	7	negative	38.6
Hydralazine	4	positive	38.6
carvedilol	7	positive	38.6
Medroxyprogesterone Acetate	3	negative	38.5
Auranofin	1	positive	38.4
terbinafine	5	positive	38.4
NOC 18	1	positive	38.4
methylatropine	2	positive	38.2
Sulindac	14	positive	38.1
Epirubicin	12	positive	38.1
Rifabutin	7	positive	38.0

Miconazole	5	positive	38.0
Guanethidine	4	negative	37.8
Melatonin	4	positive	37.8
Echinomycin	1	negative	37.8
cyanoginosin LR	2	positive	37.7
U 0126	10	positive	37.7
leflunomide	8	positive	37.6
Nitric Oxide	1	positive	37.6
Mercuric Chloride	1	positive	37.6
paeonol	1		37.5
parthenolide	3	positive	37.4
acadesine	1	positive	37.3
Thioctic Acid	2	positive	37.2
MF59 oil emulsion	1	positive	37.0
Suloctidil	1	positive	36.8
Calcitriol	6	positive	36.8
enrofloxacin	2	negative	36.8
1-hydroxycholecalciferol	4	positive	36.6
Vanadates	2	positive	36.6
Primaquine	3	negative	36.6
Dimercaprol	1		36.6
pioglitazone	11	negative	36.5
methyl salicylate	1	positive	36.5
Deoxycholic Acid	3	positive	36.5
desloratadine	2	positive	36.5
cetraxate	2	positive	36.4
Ethylestrenol	4	positive	36.4
sevoflurane	1	positive	36.4
Tacrine	11	negative	36.2
Nifedipine	6	positive	36.2
Fluoxetine	11	negative	36.1
nateglinide	3	positive	36.1
Dextran Sulfate	1	negative	36.0
Deferoxamine	3	positive	36.0
Albendazole	5	positive	36.0
lactacystin	1	positive	36.0
ibufenac	2	negative	35.9
Lisinopril	4	positive	35.9
Oxazolone	1		35.9
Bromisovalum	5	positive	35.8
halofuginone	2	negative	35.8
Mitoxantrone	3	positive	35.6
valdecoxib	6	negative	35.6

3-hydroxyacetanilide	4	positive	35.6
Naproxen	9	negative	35.5
Pyrogallol	3	positive	35.5
Isoproterenol	10	positive	35.5
Carboplatin	15	positive	35.4
Erythromycin	6	positive	35.3
fragment C, human serum albumin	2	positive	35.3
mometasone furoate	1	positive	35.3
Phenoxybenzamine	1	positive	35.2
calmidazolium	1	positive	35.2
Gentian Violet	4	negative	35.2
Caerulein	1	positive	35.2
rottlerin	1	positive	35.1
Dihydrotestosterone	11	positive	35.1
Oxymetazoline	5	positive	35.0
4-(4-fluorophenyl)-2-(4-hydroxyphenyl)-5-(4-pyridyl)imidazole	2	positive	35.0
Angiotensin-Converting Enzyme Inhibitors	1	positive	34.9
2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one	8	negative	34.9
oxcarbazepine	3	positive	34.9
Fluphenazine	9	positive	34.9
Ergocalciferols	9	positive	34.8
Sulfaphenazole	4	positive	34.8
Sotalol	7	positive	34.8
Cortisone	8	negative	34.7
Mifepristone	8	positive	34.7
resveratrol	6	positive	34.7
Trichloroepoxypropane	1	positive	34.6
Maprotiline	2	negative	34.5
Dexamethasone	56	negative	34.5
Trimethadione	8	negative	34.5
cis-9, trans-11-conjugated linoleic acid	1		34.4
Roxithromycin	6	negative	34.3
Chromium	1	positive	34.1
phenacetamide	2	positive	34.1
troglitazone	8	positive	34.1
Cholecalciferol	11	positive	34.1
Triazolam	2	positive	34.0

Gliclazide	3	positive	33.9
Daunorubicin	9	positive	33.7
beta-cyclodextrin- benzaldehyde	1	positive	33.6
Tetracycline	7	positive	33.6
Fluconazole	5	positive	33.6
Bacitracin	4	positive	33.5
naphthalan	1	positive	33.5
Anisomycin	1	positive	33.4
Quinacrine	1	positive	33.1
grepafloxacin	1	negative	33.1
Spironolactone	9	positive	33.0
Fenbendazole	4	positive	33.0
Digitoxigenin	1	positive	33.0
benoxaprofen	2	positive	33.0
Phenformin	4		33.0
Bithionol	6	positive	33.0
sulconazole	5	negative	32.9
idebenone	2	positive	32.9
Sirolimus	22	negative	32.8
Hydroxyzine	3	positive	32.7
zaleplon	2	positive	32.7
2-tert-butylhydroquinone	1	positive	32.4
Vecuronium Bromide	3	positive	32.3
acidocin CH5, Lactobacillus acidophilus	2		32.3
Gonadotropins	2	positive	32.2
sapphyrin	2	positive	32.1
nifuroxazide	1	positive	32.1
Capsaicin	3	positive	32.1
Bupropion	4	positive	32.0
Citric Acid	3	positive	32.0
Dexfenfluramine	3	negative	32.0
Megestrol Acetate	4	negative	32.0
Nitrofurantoin	7	positive	31.9
Acetazolamide	10	positive	31.9
Inosine Monophosphate	1	negative	31.8
Diltiazem	5	positive	31.8
Carbimazole	5	positive	31.8
venlafaxine	6	positive	31.7
lomefloxacin	3	negative	31.7
ubiquinol	1	positive	31.7
norethindrone acetate	3	positive	31.6

thrombin receptor peptide (42-55)	1	positive	31.6
Medroxyprogesterone	2	negative	31.6
rosiglitazone	26	positive	31.6

Table C3 Full list of enriched GO terms (GO level 5 and above, with at least five genes).

GO Accession	GO Level	GO Term Name	BMD Mean	BMDL Mean
<i>Molecular function</i>				
GO:0046872	5	metal ion binding	2.43	1.64
GO:0035639	6	purine ribonucleoside triphosphate binding	2.66	1.97
GO:0032555	5	purine ribonucleotide binding	2.66	1.97
GO:0005524	7	ATP binding	2.75	2.04
GO:0032559	6	adenyl ribonucleotide binding	2.75	2.04
GO:0030554	5	adenyl nucleotide binding	2.75	2.04
GO:0016301	5	kinase activity	2.63	1.89
GO:0016773	5	phosphotransferase activity, alcohol group as acceptor	2.63	1.89
GO:0008270	7	zinc ion binding	5.20	3.53
GO:0004672	6	protein kinase activity	2.37	1.62
GO:0046914	6	transition metal ion binding	5.20	3.53
<i>Biological Process</i>				
GO:0006355	7	regulation of transcription, DNA-dependent	2.94	2.14
GO:0031399	7	regulation of protein modification process	2.68	1.90
GO:0001932	7	regulation of protein phosphorylation	2.13	1.49
GO:0042325	7	regulation of phosphorylation	2.13	1.49
GO:0045859	7	regulation of protein kinase activity	2.13	1.49
GO:0016070	6	RNA metabolic process	2.41	1.75
GO:0006464	6	protein modification process	2.61	1.89
GO:0016310	6	phosphorylation	2.41	1.73
GO:2000112	6	regulation of cellular macromolecule biosynthetic process	2.85	2.04
GO:0042981	6	regulation of apoptosis	2.57	1.90
GO:0051252	6	regulation of RNA metabolic process	2.94	2.14
GO:0006468	6	protein phosphorylation	2.21	1.56
GO:0019220	6	regulation of phosphate metabolic process	2.15	1.52

GO:0032268	6	regulation of cellular protein metabolic process	2.68	1.90
GO:0006366	6	transcription from RNA polymerase II promoter	2.35	1.70
GO:0016071	6	mRNA metabolic process	1.86	1.36
GO:0019752	6	carboxylic acid metabolic process	2.72	1.92
GO:0043066	6	negative regulation of apoptosis	3.01	2.18
GO:0043549	6	regulation of kinase activity	2.13	1.49
GO:0044267	5	cellular protein metabolic process	2.49	1.81
GO:0090304	5	nucleic acid metabolic process	2.41	1.73
GO:0034645	5	cellular macromolecule biosynthetic process	2.66	1.92
GO:0006796	5	phosphate metabolic process	2.39	1.73
GO:0006915	5	apoptosis	2.64	1.93
GO:0010468	5	regulation of gene expression	2.85	2.04
GO:0010556	5	regulation of macromolecule biosynthetic process	2.85	2.04
GO:0019219	5	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	2.85	2.04
GO:0031326	5	regulation of cellular biosynthetic process	2.85	2.04
GO:0006351	5	transcription, DNA-dependent	2.94	2.14
GO:0032774	5	RNA biosynthetic process	2.94	2.14
GO:0043067	5	regulation of programmed cell death	2.57	1.90
GO:0009966	5	regulation of signal transduction	2.80	1.91
GO:0051174	5	regulation of phosphorus metabolic process	2.15	1.52
GO:0051246	5	regulation of protein metabolic process	2.68	1.90
GO:0006508	5	proteolysis	3.30	2.33
GO:0006974	5	response to DNA damage stimulus	2.27	1.66
GO:0045595	5	regulation of cell differentiation	2.84	2.04
GO:0006996	5	organelle organization	2.49	1.86
GO:0007167	5	enzyme linked receptor protein signaling pathway	2.38	1.79
GO:0007399	5	nervous system development	2.99	2.12
GO:0009968	5	negative regulation of signal transduction	2.80	1.91
GO:0010629	5	negative regulation of gene expression	3.25	2.31

GO:0043068	5	positive regulation of programmed cell death	1.99	1.45
GO:0043069	5	negative regulation of programmed cell death	3.01	2.18
GO:0043436	5	oxoacid metabolic process	2.72	1.92
GO:0051338	5	regulation of transferase activity	2.13	1.49
GO:0072358	5	cardiovascular system development	3.10	2.08
GO:0072359	5	circulatory system development	3.10	2.08
<i>Cellular component</i>				
GO:0005634	5	nucleus	2.24	1.64
GO:0031981	5	nuclear lumen	2.43	1.76
GO:0016021	5	integral to membrane	2.23	1.55
GO:0005654	5	nucleoplasm	2.60	1.89
GO:0005730	5	nucleolus	2.36	1.72
GO:0005739	5	mitochondrion	2.16	1.57
GO:0005829	5	cytosol	2.83	2.00

C.2 Chapter 3

Table C4 Enriched IPA Signaling Pathways for the high dose groups (\pm BrdU). A molecular pathway is a collection of cellular molecules that work together inside a cell to achieve some biological function. Here we show that many of the same pathways are enriched in both \pm BrdU conditions. Pathways are considered significantly enriched when $P \leq 0.05$ (or $-\log(p) \geq 1.3$), indicated in bold text, and there are at least four differentially expressed genes.

IPA Signaling Pathways	\pmBrdU	$-\log(p\text{-value})$	p-value	Genes that were differentially expressed by 8 mkd furan that also occur in the molecular pathway
NRF2-mediated Oxidative Stress Response	(-BrdU)	6.76	0.000	GSTM5,DNAJB4,GSTA5,NQO1,GCLC,DNAJA1,TXNRD1,HMOX1,JUN,DNAJB11,ATF4,ENC1,CBR1,GSTP1,EPHX1
	(+BrdU)	4.43	0.000	HMOX1,JUN,GSTM5,GSTA5,NQO1,GCLC,ATF4,NFE2L2,ENC1,EPHX1
Aryl Hydrocarbon Receptor Signaling	(-BrdU)	5.75	0.000	MYC,ALDH1L1,JUN,GSTM5,GSTA5,NQO1,CDKN1A,HSP90A1,BAX,ESR1,GSTP1,HSPB1
	(+BrdU)	5.47	0.000	MYC,JUN,GSTM5,GSTA5,NQO1,CDKN1A,MDM2,BAX,NFE2L2,HSPB1
Antioxidant Action of Vitamin C	(-BrdU)	3.80	0.000	SLC2A5,HMOX1,SLC23A2,IKBKG,PNPLA3,STAT5B,SLC23A3,TXNRD1
	(+BrdU)	1.01	0.098	SLC2A5,HMOX1,SLC23A3
Xenobiotic Metabolism Signaling	(-BrdU)	3.72	0.000	GSTM5,UGT2B7,UGT1A6,GSTA5,NQO1,GCLC,HMOX1,ALDH1L1,CYP3A5,SULT1C2,HSP90AA1,GSTP1,Ces2b/Ces2c
	(+BrdU)	3.24	0.001	HMOX1,GSTM5,UGT2B7,GSTA5,SULT1C2,NQO1,GCLC,NFE2L2,SULT2A1,Ces2b/Ces2c
14-3-3-mediated Signaling	(-BrdU)	3.20	0.001	JUN,TUBB6,TUBB4B,TUBB2A,TUBA4A,TUBA1C,BAX,TUBA1B
	(+BrdU)	1.94	0.011	JUN,TUBB6,TUBB4B,TUBB2A,BAX
Remodeling of Epithelial Adherens Junctions	(-BrdU)	3.18	0.001	TUBB6,TUBB4B,TUBB2A,TUBA4A,TUBA1C,TUBA1B

	(+BrdU)	1.39	0.041	TUBB6,TUBB4B,TUBB2A
Aldosterone Signaling in Epithelial Cells	(-BrdU)	3.10	0.001	SCNN1A,HSPA8,DNAJB4,DNAJB11,HSPH1,HSP90AA1,DNAJA1,HSPB1,HSPA4L
	(+BrdU)	0.60	0.254	SCNN1A,HSPA5,HSPB1
Breast Cancer Regulation by Stathmin1	(-BrdU)	2.45	0.004	STMN1,PPP1R10,TUBB6,TUBB4B,CDKN1A,TUBB2A,TUBA4A,TUBA1C,TUBA1B
	(+BrdU)	1.17	0.068	STMN1,TUBB6,TUBB4B,CDKN1A,TUBB2A
p53 Signaling	(-BrdU)	2.36	0.004	CCNG1,TP53INP1,JUN,CDKN1A,BAX,TNFRSF10A
	(+BrdU)	4.87	0.000	CCNG1,TP53INP1,JUN,CDKN1A,MDM2,BAX,TNFRSF10A,BIRC5
Estrogen-mediated S-phase Entry	(-BrdU)	2.15	0.007	MYC,CDKN1A,ESR1
	(+BrdU)	1.49	0.032	MYC,CDKN1A
Protein Ubiquitination Pathway	(-BrdU)	2.08	0.008	HSPA8,DNAJB4,DNAJB11,HSPH1,HSP90AA1,USP50,DNAJA1,USP34,HSPB1,HSPA4L
	(+BrdU)	0.48	0.334	CDC20,MDM2,HSPA5,HSPB1
VDR/RXR Activation	(-BrdU)	2.04	0.009	CXCL10,SERPINB1,PDGFA,CDKN1A,IGFBP1
	(+BrdU)	2.66	0.002	CXCL10,PDGFA,CDKN1A,IGFBP1,SULT2A1
Epithelial Adherens Junction Signaling	(-BrdU)	2.03	0.009	DLL1,TUBB6,TUBB4B,TUBB2A,TUBA4A,TUBA1C,TUBA1B
	(+BrdU)	1.06	0.087	TUBB6,TUBB4B,TUBB2A,FGF1
Axonal Guidance Signaling	(-BrdU)	1.99	0.010	RND1,ADAMTS1,NRP2,PDGFA,TUBB4B,TUBB2A,TUBA4A,ADAM11,TUBA1B,FZD8,TUBB6,GNAT1,PLXNB1,TUBA1C
	(+BrdU)	0.43	0.374	FZD8,TUBB6,ADAMTS1,PDGFA,TUBB4B,TUBB2A
ERK/MAPK Signaling	(-BrdU)	1.99	0.010	MYC,PPP1R10,PTK2B,DUSP6,HIST2H3C (includes others),ATF4,ESR1,HSPB1
	(+BrdU)	1.21	0.062	MYC,PTK2B,DUSP6,ATF4,HSPB1
Death Receptor Signaling	(-BrdU)	1.91	0.012	IKBKG,TNFSF10,TNFRSF10A,HSPB1
	(+BrdU)	1.58	0.026	TNFSF10,TNFRSF10A,HSPB1
Gap Junction Signaling	(-BrdU)	1.91	0.012	TUBB6,TUBB4B,CSNK1G3,TUBB2A,TUBA4A,TUBA1C,TUBA1B
	(+BrdU)	0.59	0.260	TUBB6,TUBB4B,TUBB2A

Germ Cell-Sertoli Cell Junction Signaling	(-BrdU)	1.88	0.013	RND2,TUBB6,TUBB4B,TUBB2A,TUBA4A,TUBA1C,TUBA1B
	(+BrdU)	0.98	0.105	RND2,TUBB6,TUBB4B,TUBB2A
ErbB2-ErbB3 Signaling	(-BrdU)	1.85	0.014	MYC,JUN,Nrg1,STAT5B
	(+BrdU)	0.85	0.143	MYC,JUN
PPAR Signaling	(-BrdU)	1.79	0.016	IKBKG,JUN,PDGFA,HSP90AA1,STAT5B
	(+BrdU)	1.05	0.089	JUN,PDGFA,IL1RAP
Chronic Myeloid Leukemia Signaling	(-BrdU)	1.73	0.019	MYC,IKBKG,HDAC11,CDKN1A,STAT5B
	(+BrdU)	1.02	0.095	MYC,CDKN1A,MDM2
LPS/IL-1 Mediated Inhibition of RXR Function	(-BrdU)	1.72	0.019	ALDH1L1,JUN,GSTM5,CYP2A13,CYP3A5,GSTA5,SULT1C2,GSTP1
	(+BrdU)	1.49	0.032	JUN,GSTM5,GSTA5,SULT1C2,IL1RAP,SULT2A1
Estrogen-Dependent Breast Cancer Signaling	(-BrdU)	1.71	0.019	JUN,ATF4,STAT5B,ESR1
	(+BrdU)	0.78	0.167	JUN,ATF4
Hypoxia Signaling in the Cardiovascular System	(-BrdU)	1.71	0.019	JUN,NQO1,HSP90AA1,ATF4
	(+BrdU)	2.20	0.006	JUN,NQO1,ATF4,MDM2
Sertoli Cell-Sertoli Cell Junction Signaling	(-BrdU)	1.61	0.025	JUN,TUBB6,TUBB4B,TUBB2A,TUBA4A,TUBA1C,TUBA1B
	(+BrdU)	0.84	0.146	JUN,TUBB6,TUBB4B,TUBB2A
Tec Kinase Signaling	(-BrdU)	1.44	0.036	RND2,GNAT1,PTK2B,TNFSF10,STAT5B,TNFRSF10A
	(+BrdU)	1.01	0.098	RND2,PTK2B,TNFSF10,TNFRSF10A
Prostate Cancer Signaling	(-BrdU)	1.35	0.045	CDKN1A,HSP90AA1,ATF4,GSTP1
	(+BrdU)	1.16	0.069	CDKN1A,ATF4,MDM2
PI3K/AKT Signaling	(-BrdU)	1.33	0.047	IKBKG,GDF15,CDKN1A,FOXO3,HSP90AA1
	(+BrdU)	1.28	0.052	GDF15,CDKN1A,FOXO3,MDM2
Bladder Cancer Signaling	(-BrdU)	1.29	0.051	MYC,FGF21,CDKN1A,RASSF1
	(+BrdU)	2.49	0.003	MYC,CDKN1A,MDM2,RASSF1,FGF1

ATM Signaling	(-BrdU)	1.14	0.072	JUN,CDKN1A,ATF4
	(+BrdU)	2.31	0.005	JUN,CDKN1A,ATF4,MDM2
Glioblastoma Multiforme Signaling	(-BrdU)	1.03	0.093	RND2,MYC,FZD8,PDGFA,CDKN1A
	(+BrdU)	2.13	0.007	RND2,MYC,FZD8,PDGFA,CDKN1A,MDM2
Molecular Mechanisms of Cancer	(-BrdU)	1.01	0.098	RND2,MYC,FZD8,JUN,APH1B,GNAT1,SMAD9,CDKN1A,BAX
	(+BrdU)	1.38	0.042	RND2,MYC,FZD8,JUN,APH1B,CDKN1A,MDM2,BAX
p38 MAPK Signaling	(+BrdU)	0.95	0.111	MYC,HIST2H3C (includes others),ATF4,HSPB1
	(-BrdU)	1.38	0.042	MYC,ATF4,IL1RAP,HSPB1
Sphingosine-1-phosphate Signaling	(-BrdU)	0.61	0.245	RND2,PTK2B,PDGFA
	(+BrdU)	1.47	0.034	RND2,PTK2B,S1PR5,PDGFA

C.3 Chapter 4

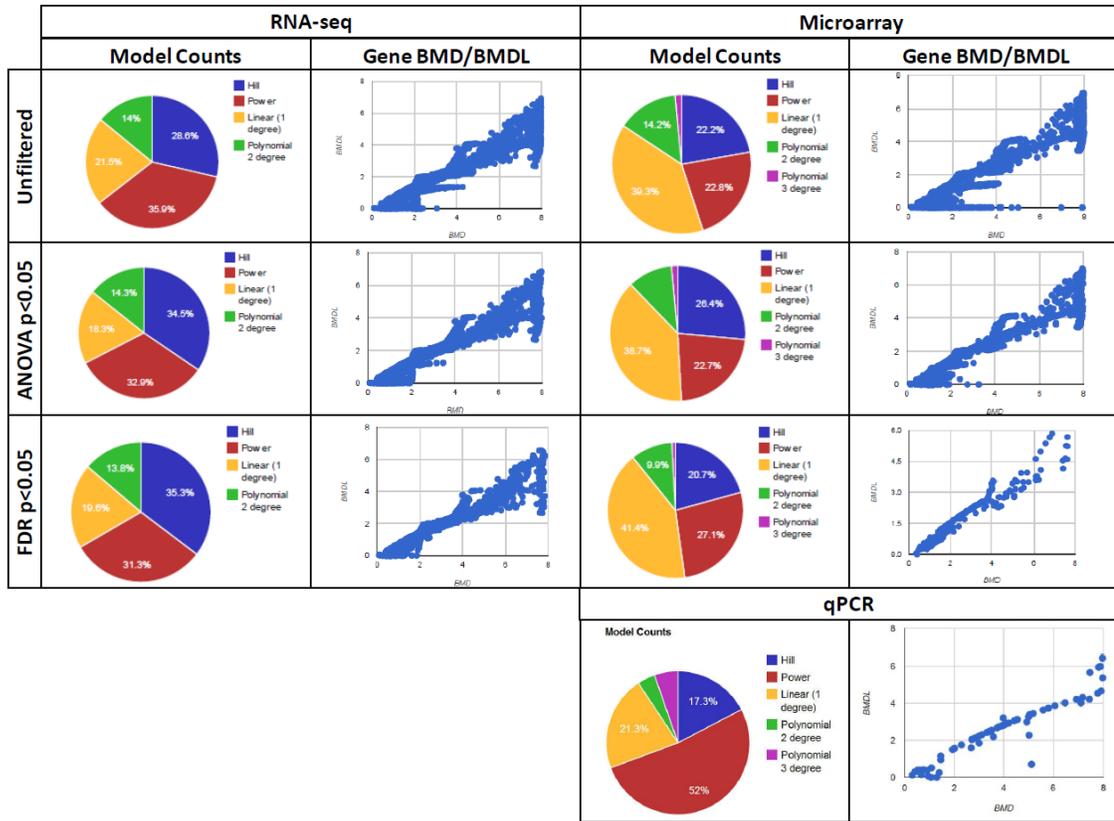


Figure C1 Model counts (pie charts) and gene BMD/BMDL plots (scatterplots) for each platform. Linear regressions comparing BMD/BMDL values were $R^2 > 0.9$ (linear regression $P < 0.0001$), with a slopes of 0.66-0.75 (corresponding to a BMD/BMDL ratio of 1.5-1.3).

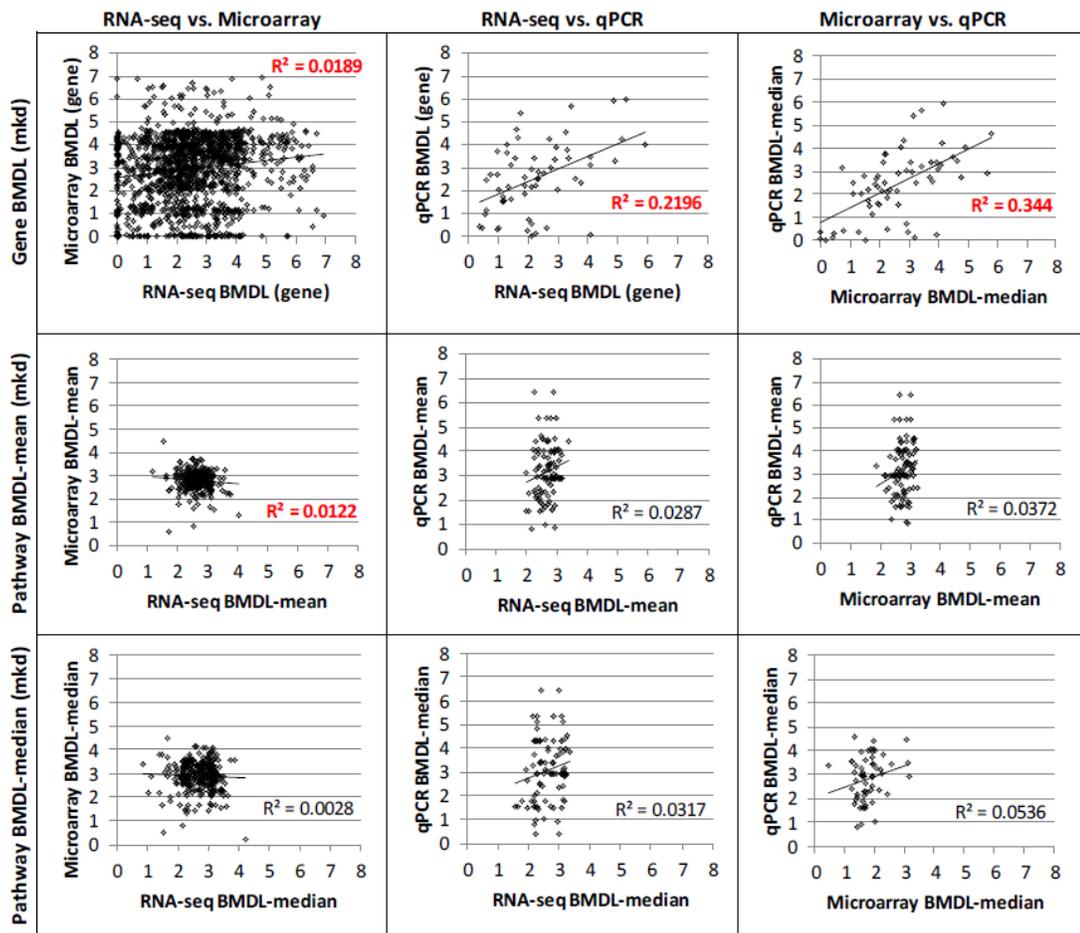


Figure C2 Inter-platform comparisons of BMDLs for genes (top row), pathway-means (middle row), and pathway medians (bottom row) for ANOVA filtered data. Statistically significant correlations are indicated in red (regression $P < 0.05$).

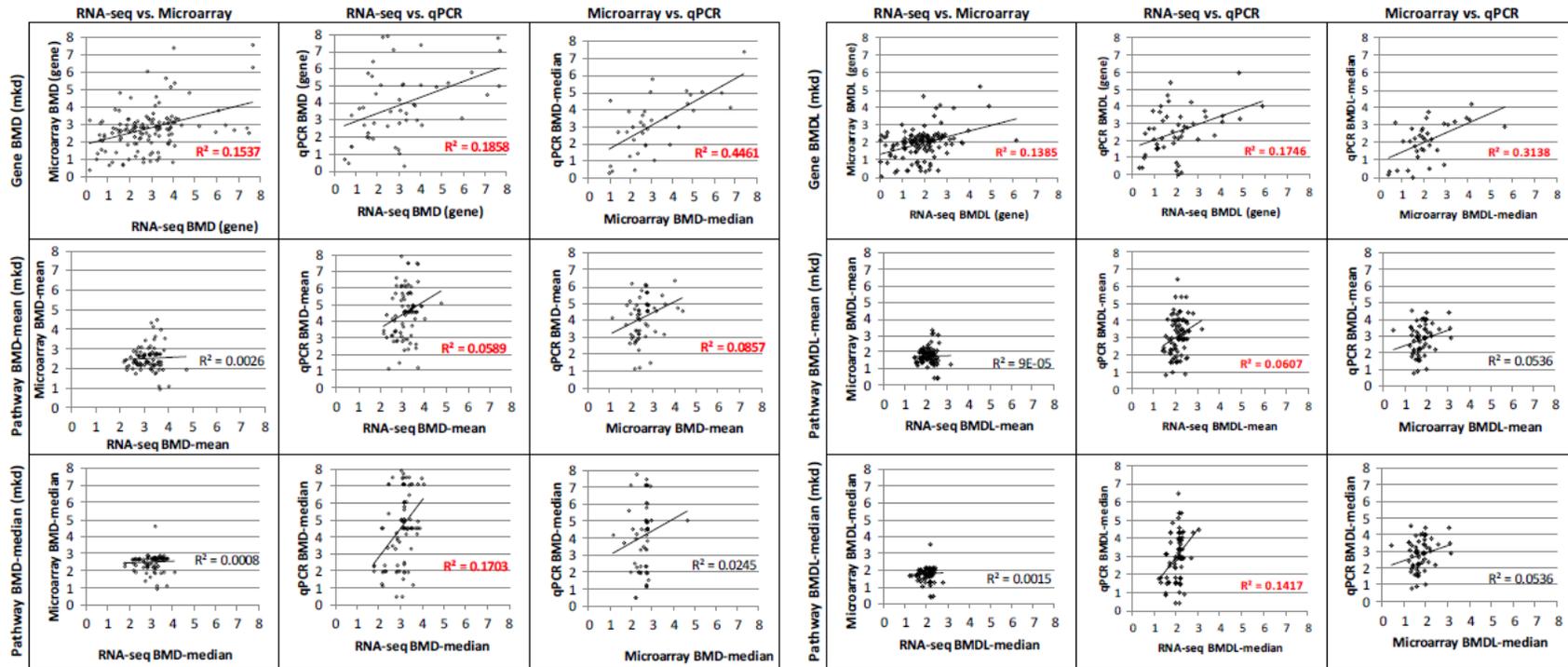


Figure C3 Inter-platform comparisons of BMDs (left) and BMDLs (right) for genes (top row), pathway-means (middle row), and pathway medians (bottom row) for FDR filtered data. Statistically significant correlations are indicated in red (regression $P < 0.05$).

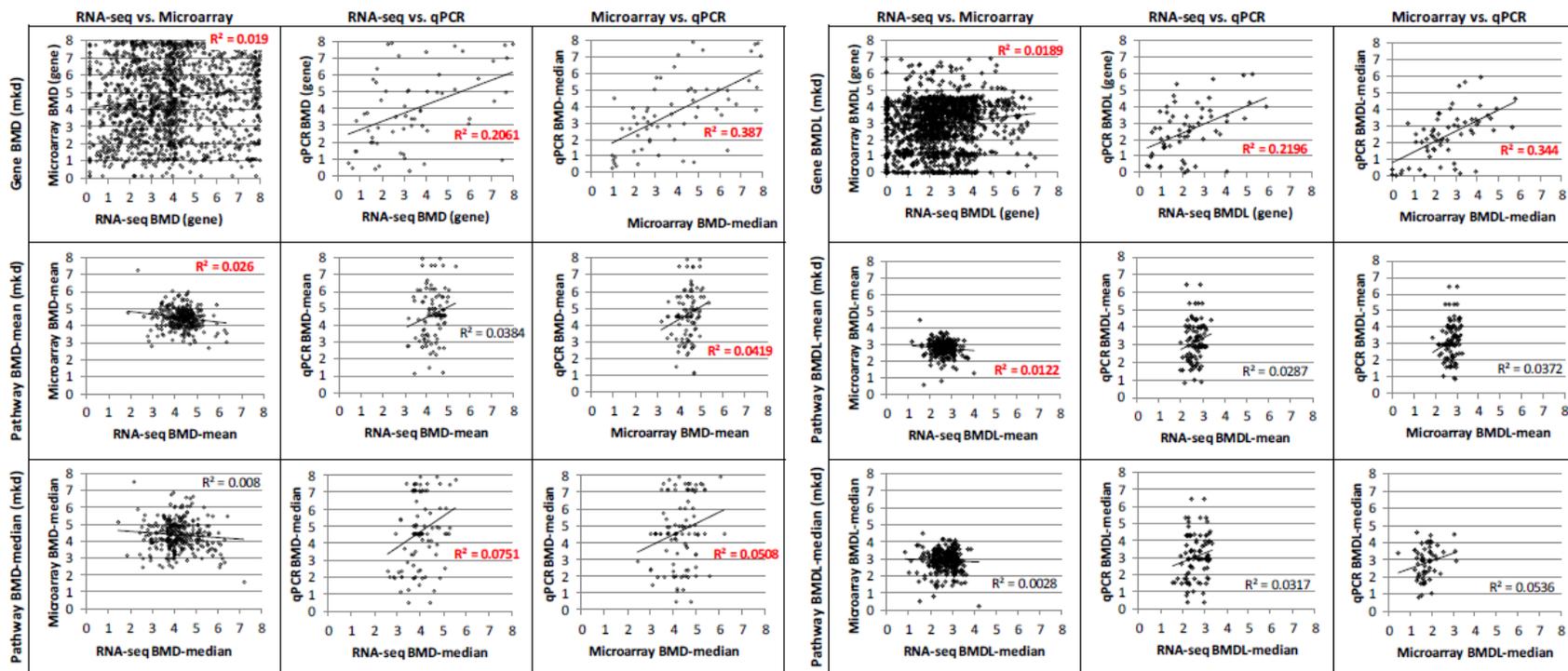


Figure C4 Inter-platform comparisons of BMDs (left) and BMDLs (right) for genes (top row), pathway-means (middle row), and pathway medians (bottom row) for unfiltered data. Statistically significant correlations are indicated in red (regression $P < 0.05$).

C.4 Chapter 5

C.4.1 Figures and Tables

Table C5 qPCR primers.

Gene	Direction	Product length	Primer Sequence	T _m
b-Actin	F	63bp	CACCGATCCACACAGAGTACTTG	64.6
	R		CTCTGGCTCCTAGCACCATGA	64.5
Cela2a	F	63bp	TTATGAGGTGGAGGATGATG	58.4
	R		CAGGTGTTGGGTGTGG	59.3
Ces6	F	70bp	CCAGGAGCTCAAGGGTTCTCA	55.48
	R		CTCACACACCTTCCCTGACAC	54.72
Ctrl	F	61bp	CAAGTCATGGCCTACAATA	58.4
	R		TGTTTATTAGATTGAGATGGGGA	57.4
Cyp2b10	F	67bp	TTCTATCAGACCTTTTCACTCATAAGC	61.6
	R		GGAAGCCAGAGAAGAGCTCAA	62.6
Cyp2b9	F	66bp	GGAGTCCTGCTCATGCTCAAGT	64.5
	R		CACCTGATCAATCTCCTTTTGGGA	61
Cyp2c55	F	77bp	GCCGTGTACCACCTTCATAC	52.65
	R		ATACCAGCATAGCAACAGCCAG	54.85
Cyp3a11	F	61bp	GTGTATATCCCCAAAGGGTCAACA	54.29
	R		GTGGGTCATGGTGAAGAGCATA	54.17
Esco2	F	80bp	CTGGTACTATGTGGGAAAGT	58.4
	R		GTATGTTTATTCTGCTACGA	56.3
Gstm2	F	73bp	CCTTGTCTCTGCCCTGCATC	54.95
	R		CGGGGATTCTCGGGTCTAGTA	54.29
Gstm3	F	60bp	GACGGCCCTCCCTAGAGTTA	54.19
	R		AAACAGGAACAGGCTGGCAC	55.26
Gucy2e	F	74bp	GATTAGAGACTGGGTAGGGG	62.4
	R		CCCTAAATCAGAGAGGAATGG	60.6

Hunk	F	77bp	TTATACCTGCCACACAAAGT	56.3
	R		TCCCTAAGTAAGTGTGTGGA	58.4
Mtap1a	F	65bp	CTCCATTATTCAAACCCATCCC	60.8
	R		CATCTCATCCTCCTTTCACAC	60.6
Por	F	60bp	CTTTGGGTGATGGGTGCCTT	54.61
	R		CCCAGCTCTGTGGAGAGAAGA	54.95
Ppp3cb	F	67bp	ACAGTAGCACACAATAAATTCC	57.1
	R		TCCAAATAAAACCCAACACAG	56.7
Setd1a	F	60bp	TGTATAGATGCTATTCTTGGGGG	61
	R		GGAAGTCCCCATTAACATGAAGC	62.8
Sgta	F	70bp	TCCTAACTTTGTGTCCAGTC	58.4
	R		CACAGATGGGTCAATGTCTA	58.4
Sspl3	F	65bp	CACCTCACACTTGGTTAGC	60.2
	R		GGGAAGGAACAGACTCTAGG	62.4
Ttc17	F	64bp	GAAGCTACAGCCAGAGTTTGTG	62.7
	R		CATGAGGTGGCACTGAATTGT	60.6

Table C6 QC metrics of the one-colour arrays (bold text indicates the guidelines followed).

	Agilent Guidelines		1C, 1X		1C, 1.5X	
	<i>Standard GE</i>	1 colour FFPE	Mean (n=8)	SD	Mean (n=8)	SD
AnyColourPrcntFeatNonUnifOL	<1	<1	0.01	0.01	0.00	0.01
gNegCtlAveNetSig	<40	<100	65.73	9.93	232.76	22.82
gNegCtlAveBGSubSig	<i>(-10) to 5</i>	(-10.0) to 5	-4.27	1.41	<i>-13.53</i>	4.16
gNegCtlSDevBGSubSig	<10	<10	4.28	0.56	<i>11.50</i>	2.28
gSpatialDetrendRMSfit	<15	<15	5.81	0.71	14.60	1.39
gNonCntrlMedCVProcSig	<i>0 to 8</i>	0 to 8	7.74	0.66	5.16	0.52

Italics: value falls outside of guideline

Table C7 QC metrics of the two-colour arrays (bold text indicates the guidelines followed).

	Agilent Guidelines		2C, 1X		2C, 1.5X	
	<i>1C FFPE</i>	Standard GE	Mean (n=8)	SD	Mean (n=8)	SD
AnyColourPrcntFeatNonUnifOL	<1	<1	0.01	0.01	0.03	0.02
gNegCtlAveBGSubSig	<i>(-10.0) to 5</i>	(-20 to 10)	-4.14	1.16	-5.16	0.83
gNegCtlSDevBGSubSig	<10	<15	5.19	1.06	4.94	0.46
gNonCntrlMedCVProcSig	<i>0 to 8</i>	0 to 18	8.24	1.58	10.71	2.64
rNegCtlAveBGSubSig	<i>na</i>	(-20 to 4)	-2.11	0.49	-0.69	0.24
rNegCtlSDevBGSubSig	<i>na</i>	<6	4.53	0.51	3.27	0.12
rNonCntrlMedCVProcSig	<i>na</i>	0 to 18	9.75	2.75	<i>28.44</i>	10.06

Italics: value falls outside of guideline

C.4.2 Supplemental Protocol

PROTOCOL FOR: Preparation of archival FFPE mouse liver samples for use with the Agilent gene expression microarray platform

Anna Francina Jackson^{1,2}, Andrew Williams¹, Ivy Moffat¹, Suzanne Phillips³, Les Recio³, Mike Waters³, Iain Lambert², Carole Yauk¹

¹*Mechanistic Studies Division, Environmental Health Science and Research Bureau, Health Canada, Ottawa, Canada, K1A 0K9*

²*Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada*

³*Integrated Laboratory Systems, Inc., P.O. Box 13501, Research Triangle Park, NC 27709, USA*

REAGENTS

- RNeasy FFPE RNA Extraction kit (Qiagen, Hilden, Germany, 73504)
- Xylene (Sigma-Aldrich, Oakville, Ontario, Canada)
- 100% Ethanol, molecular grade (Sigma-Aldrich, Oakville, Ontario, Canada)
- WTA2 Kit (Sigma-Aldrich, Oakville, Ontario, Canada, WTA2-50RXN or WTA2-10RXN)
- QiaQuick PCR Clean-up kit (Qiagen, Hilden, Germany, 28104)
- Genomic DNA ULS labeling kit (Agilent Technologies, Santa Clara, California, USA, 5190-0419)
- Hi-RPM Gene Expression Hybridization kit (Agilent Technologies, Santa Clara, California, USA, 5190-0404)
- Gene Expression Wash Buffer Kit (Agilent Technologies, Santa Clara, California, USA, 5188-5327)
- SurePrint G3 Mouse GE 8x60K Microarray kit (Agilent Technologies, Santa Clara, California, USA, G4852A)
- RNaseZAP (Ambion, Carlsbad, California, USA. AM9780)

EQUIPMENT

- Microcentrifuge (ThermoScientific, Wilmington, Delaware, USA)
- Heat Block (Fisher Scientific Ltd, Ottawa, Ontario, Canada)
- Waterbath (ThermoScientific, Wilmington, Delaware, USA)
- Pipettes (Fisher Scientific Ltd, Ottawa, Ontario, Canada)
- NanoDrop (Thermo Scientific, Wilmington, Delaware, USA)
- BioAnalyzer (Agilent Technologies, Santa Clara, California, USA)
- Thermocycler (Vendor, City, State, Country)
- Agilent G2505B scanner (Agilent Technologies, Santa Clara, California, USA)
- Coplin jars/staining dishes (Fisher Scientific Ltd, Ottawa, Ontario, Canada)
- Forceps (Fisher Scientific Ltd, Ottawa, Ontario, Canada)
- Sterile single-use razor blades (Fisher Scientific Ltd, Ottawa, Ontario, Canada)
- Hybridization Oven (ThermoScientific, Wilmington, Delaware, USA)

PROCEDURE

LEGEND

HINT: Tip to ensure success at this point in the protocol

ATTENTION: Include critical information

REST: Include instructions for storage of materials

RNA EXTRACTION (adapted from the RNeasy FFPE RNA extraction kit protocol)

HINT: Step-by-step instructions for RNA extraction are also available in Table P1.

1. Follow the protocol as set-out in the RNeasy FFPE RNA extraction kit with the following exceptions:
 - a. If you are extracting RNA from freshly sectioned FFPE tissue, deparaffinize with xylene as per appendix A (Qiagen manual, pg 27)
 - b. If you are extracting RNA from slide-mounted sections, proceed to step 2 of this protocol (Adapted from Absolutely RNA FFPE kit manual, Appendix I, Agilent Technologies)
2. Place slides in a slide rack or Coplin jar
3. Incubate slides in a staining dish or Coplin jar containing a sufficient volume of xylene such that the slides are covered for 15 minutes at room temperature.
4. Transfer the slides to a second dish containing 100% EtOH and incubate for 10 minutes
5. Transfer the slides to a third dish containing 90% EtOH and incubate for 10 minutes
6. Transfer the slides to a fourth dish containing 70% EtOH and incubate for 10 minutes

7. Prepare one 1.5mL tube per sample/slide with 0.5mL of 70% EtOH
8. Remove one slide from the dish. Using a single-use, sterile razor blade, scrape the section into its tube. Repeat until all of the samples are in their tubes.

HINT: Avoid scraping dry tissue sections into the tubes. The plastic tube creates static electricity that causes the section to jump around making it difficult to ensure the entire sample has entered the tube.

9. Spin the tubes down at maximum speed for 5 minutes. Carefully discard the supernatant using a pipette. *If necessary, spin again for 1 minute at maximum speed and discard the remaining supernatant using a pipette.*
10. Allow the pellet to air dry (5-10 minutes).
11. Add Buffer PKD, vortex to mix
12. Add 10 µl Proteinase K, pipette to mix.
13. Incubate at 56°C for 15 minutes and then 80°C for 15 minutes.
14. Continue as per the published protocol, Step 11: Incubate on ice for 3 minutes and spin down for 15 minutes at 13 000 rpm.
15. Following the completion of the protocol, quantify the extracted RNA using a NanoDrop 2000 spectrophotometer and measure its quality using a Bioanalyzer.

REST: Samples can be stored at -80°C.

HINT: For additional quality control measures at this point, we recommend performing cDNA synthesis (using the High Capacity cDNA Reverse Transcription kit, Applied Biosystems) and real-time PCR analysis on a few positive and negative control genes.

WHOLE TRANSCRIPTOME AMPLIFICATION (WTA)

16. Follow the WTA2 protocol according to the manufacturer's instructions.

- a. For FFPE samples begin with at least 250 ng RNA per sample (for 250 ng RNA from mouse liver, you can expect 7-10 µg of product).

HINT: If you choose to use the WTA1 kit (Sigma), be aware that you must order the enzymes separately.

17. For sample clean-up use the QiaQuick PCR clean-up kit (Qiagen) as per the manufacturer's protocol.

HINT: At the end of the WTA2 Amplification reaction you will have 5x75 µl reactions. Instead of using one QiaQuick column for each of these aliquots combine aliquots before purification. The maximum binding capacity of the column is 10ug so combine aliquots such that 10ug/column is not exceeded (ex.: 2x187.5).

HINT: To avoid having to concentrate samples upon elution, elute in 30 µL (not the recommended 50 µL)

18. Combine eluted aliquots (2x30 µL=60 µL per sample) and quantify using a NanoDrop spectrophotometer

REST: Samples can be stored at -20°C for short-term and -80°C for long-term storage.

cDNA LABELLING (Adapted from the Gene Expression FFPE Workflow, Agilent)

ATTENTION: Cy3 and Cy5 dyes are light sensitive. From here on, the protocol should be carried-out in a room with as little direct light as possible.

HINT: See also Tables P2 and P3 for step-by-step instructions.

19. Label tubes for each labeling reaction. *For a two-colour experiment we suggest using a dye-swap array design.*

20. Prepare the relevant labeling reactions as per Table P2.

HINT: We have found that the amount of dye suggested by the manufacturer to be insufficient. Dye amounts in Table P2 are increased to 1.5x Cy3 and 1.75x Cy5.

21. Calculate volumes of cDNA to be labeled. You will need 1650 ng per sample for 20 µl labeling reaction or 825 ng per sample for 10 µl labeling reaction (Table P2, row 1). Do not exceed the maximum volumes indicated in Table P2 (row 2).

22. Aliquot water, labeling solution, dye and cDNA (as per Table P2, row 3-5).

23. Mix by pipetting or gentle vortex.

HINT: If working with a lot of samples, make a master mix of your dye and labeling solution.

24. Incubate tubes at 85°C for 30 minutes in the dark.

25. Incubate on ice for at least 3 minutes.

26. Spin down for 1 minute at 6000 x g in a 4°C centrifuge.

27. Keep on ice.

REMOVAL OF NON-REACTED Cy-ULS (Adapted from the Gene Expression FFPE Workflow, Agilent)

HINT: See also Table P3, steps 6-18, for step-by-step instructions.

28. Follow the manufacturer's protocol for sample purification using KREApure columns
29. Quantify each purified, labeled cDNA samples twice (2 x 1 µl) using a NanoDrop spectrophotometer on the 'Microarray' setting
30. Record the average concentration and pmol per µl dye of each sample.
Calculate the degree of labeling of each sample (should be between 1.5-3%)
Degree of labeling = $[(340 \times \text{pmol per } \mu\text{l dye}) / (\text{ng per } \mu\text{l cDNA} \times 1000)] \times 100\%$

HINT: Step 29 is achieved most easily by using an excel spreadsheet that you have created for your experiment.

HYBRIDIZATION AND SCANNING (Adapted from the Gene Expression FFPE Workflow, Agilent)

ATTENTION: FFPE tissue-derived cDNA is already degraded and therefore does not need to undergo a fragmentation reaction (as would high-quality samples)

HINT: See also Table P3, steps 19-34, for step-by-step instructions.

31. Prepare 100X GE blocking agent by adding 125 µL nuclease-free water to one vial of lyophilized 10X GE Blocking Agent (in Hi-RPM Gene Expression Hybridization kit)
32. Bring Agilent-CGHBlock (in ULS Labeling Kit) to room temperature
33. Bring heat block to 95 °C
34. Label one tube for each array

35. Add 2x Hi-RPM GE Hyb Buffer to each tube:
 - a. 4x44K: 55 μ L
 - b. 8x60K: 27.5 μ L
36. Add 100X GE Blocking Agent to each tube:
 - a. 4x44K: 1.1 μ L
 - b. 8x60K: 0.55 μ L
37. Add labeled cDNA to each tube:

One-colour:

 - a. 4x44K: 1650 ng (up to 27 μ L)
 - b. 8x60K: 600 ng (up to 13.5 μ L)

Two-colour:

 - c. 4x44K: 825 ng per sample (up to 27 μ L together)
 - d. 8x60K: 300 ng per sample (up to 13.5 μ L together)
38. Add nuclease-free water:
 - a. 4x44K: up to 83 μ L total volume
 - b. 8x60K: up to 41.5 μ L total volume
39. Mix by pipetting. *Avoid the introduction of bubbles.*
40. Incubate at 95 °C for 3 minutes, then put on ice.
41. Spin samples down
42. Add room-temperature CGHBlock
 - a. 4x44K: 27 μ L (total volume is now 110 μ L)
 - b. 8x60K: 13.5 μ L (total volume is now 55 μ L)
43. Keep samples at room-temperature until arrays are loaded

44. Load arrays:

a. 4x44K: 100 μ L

b. 8x60K: 45 μ L

45. Hybridize at 65 °C for 17 hours at 20 RPM in the dark.

ARRAY WASHING AND SCANNING

Use standard array washing and scanning protocols provided with the GE arrays by Agilent-Technologies.

TROUBLESHOOTING

PROBLEM 1: High background signal.

This occurs if you follow the standard gene expression protocol which does not include the CGH blocking buffer. Solution: include CGH blocking buffer.

Another reason this may occur is if the labeled-cDNA input amount has been increased.

Solution: Decrease amount of cDNA hybridized.

PROBLEM 2: Degree of labeling is insufficient.

Solution: Increase the amount of ULS-Cy3/5 used in the labeling reaction.

PROTOCOL TABLES

Table P1: Step-by-step protocol for RNA Extraction from FFPE sections (S1-S6) and slide-mounted sections (SM1-SM9). All steps until step 7 should be carried out in a fumehood. Allowable sample thickness and quantity can be found in the Qiagen RNeasy Handbook (page 13, 09/2011).

Instructions for freshly sectioned tissues. Source: the RNeasy FFPE kit handbook, page 27 (Qiagen, 09/2011)		Instructions for slide-mounted tissues. Source: Adapted from the Absolutely RNA FFPE Kit Manual, Appendix I (Agilent Technologies)	
S1	Use a microtome to section FFPE blocks and place fresh sections directly into 1.5mL tubes.	SM 1	Place slides in a Coplin jar or slide rack and staining dish that is filled with enough xylene so that slides are covered. Incubate at room-temperature for 15-20 minutes.
S2	Add 1mL of xylene to each tube. Vortex vigorously for 10 seconds, centrifuge at maximum speed for 2 minutes.	SM 2	Removes slides, blot excess xylenes and place in a second dish filled with 100% ethanol. Incubate for 10 minutes
S3	Use a pipette to remove the supernatant.	SM 3	Removes slides, blot and place in a third dish filled with 90% ethanol. Incubate for 10 minutes
S4	Add 1mL of 100% ethanol to each tube. Vortex vigorously for 10 seconds, centrifuge at maximum speed for 2 minutes.	SM 4	Removes slides, blot and place in a third dish filled with 70% ethanol. Incubate for 15 minutes.
S5	Use a pipette to remove the supernatant.	SM 5	Label one 1.5mL tube per sample. Fill each tube with 0.5mL of 70% ethanol
S6	Allow the pellets to air dry (this can take up to 20 minutes).	SM 6	One slide at-a-time: remove from 70% ethanol, blot and scrape the deparaffinized tissue into its 1.5mL tube.
	Proceed to step 7	SM 7	Centrifuge the tubes for 5 minutes at maximum speed. Carefully remove the supernatant with a pipette.
		SM 8	Centrifuge the tubes for 1 minute at maximum speed. Carefully remove the supernatant with a pipette.

SM 9 Allow the pellets to air dry (this can take up to 20 minutes).

Proceed to **step 7**

7 Add 150ul of buffer PKD to each tube, vortex.

8 Add 10uL proteinase K to each tube, mix by pipetting.

9 Incubate at 56°C for 15 minutes. Bring a second heat block to 80°C now.

10 Incubate at 80°C for exactly 15 minutes

11 Incubate on ice for 3 minutes

12 Centrifuge for 15 minutes at 13 500 rpm (20 000 x g)

13 Transfer supernatant into a new 1.5mL tube

14 Add 16uL DNase Booster Buffer to each tube

15 Add 10uL DNase I to each tube and mix by inversion. Briefly spin tubes down.

16 Incubate at room-temperature for 15 minutes.

17 Add 320 uL buffer RBC. Mix thoroughly.

18 Add 720uL 100% ethanol. Mix by pipetting.

19 Transfer 700uL (from #18) into an RNeasy MinElute spin column. Close lid and spin for 15 seconds at > 10 000 rpm (> 8 000 x g). Discard flow through.

20 Transfer remaining (from #18) into correct RNeasy MinElute spin column. Close lid and spin for 15 seconds at > 10 000 rpm (> 8 000 x g). Discard flow through.

21 Add 500 uL Buffer RPE to the column, close the lid, spin for 15 seconds at > 10 000 rpm (> 8 000 x g). Discard flow-through.

22 Add 500 uL Buffer RPE to the column, close the lid, spin for 2 minutes at > 10 000 rpm (> 8 000 x g). Discard flow-through and collection tube.

23 Place each column into a new 2 mL collection tube. Leave the lids open and spin at maximum speed for 5 minutes. Discard flow-through and collection tube.

24 Place the column in a 1.5 mL tube. Add 30uL nuclease-free water to the center of the column membrane. Close the lid and spin for 1 minute at maximum.

25 Place eluted RNA on ice. Analyze RNA quantity with a NanoDrop 2000 spectrophotometer (Thermo Scientific) and quality with a Bioanalyzer (Agilent Technologies)

Table P2: cDNA labeling reaction adapted from Agilent's 4x44K GE, one-colour, FFPE protocol for use one one- and two-colour Agilent 8x60K GE arrays.

	Reagent	One-colour (Agilent)		Two-colour (Modified protocol}
		4x44K	8x60K	8x60K
1	Amount of cDNA required for labeling (ng)	1650	1650	825
2	Maximum volume of cDNA (μL) (Cy3 / Cy5 reaction)*	16.35	15.52	7.76 / 7.35
3	Labelling solution (μL)	2	2	1
4	ULS-Cy3 (μL) / ULS-Cy5 (μL)**	1.65	2.48	1.24 / 1.44
5	Fill with nuclease-free water to a total reaction volume (μL) of	20	20	10
		1x cDNA input for hybridization		
6	Cy3-cDNA to hyb (ng)	1650***	600	300
7	Cy5-cDNA to hyb (ng)			300
		1.5x cDNA input for hybridization		
8	Cy3-cDNA to hyb (ng)	na	900	450
9	Cy5-cDNA to hyb (ng)			450
* Less Cy3 dye is used than Cy5 therefore a greater volume of cDNA can be accommodated in the Cy3 reactions				
**Only use ONE DYE per sample. ULS-Cy3 is 1.5X original recommendation; ULS-Cy5 is 1.75X original recommendation (original recommendation is 1.65μL dye in a 20μL reaction)				
***In our experience, it is unrealistic to expect 100% recovery of labeled product				

Table P3: Step-by-step protocol for microarray preparation

Instructions for preparing samples for microarray analysis. Source: Adapted from Gene Expression Workflow, version 1.0 (Agilent Technologies, February 2012)	
Label with ULS:	
1	Aliquot and dilute cDNA according to your array platform (Table P2, rows 1-2)
2	Make a labeling master mix of your labeling solution and dye (Cy3 <u>or</u> Cy 5) and then aliquot to the appropriate tubes (Table P2, rows 3-5). Note: dyes are light sensitive, try to minimize light exposure of samples from this point on.
3	Incubate at 85°C for 30 minutes
4	Incubate on ice for at least 3 minutes
5	Centrifuge for 1 minute at 6 000 x g and store on ice.
Remove excess dye according to the manufacturer's protocol using KREApure columns (Agilent Technologies):	
6	Resuspend columns by vortexing (columns are supplied with the Agilent ULS labeling kit)
7	Loosen cap, snap off bottom and place column in a collection tube
8	Centrifuge for 1 minute at maximum speed.
9	Discard flow-through
10	Add 300 µL water to the column
11	Centrifuge for 1 minute at maximum speed.
12	Discard flow-through and collection tube.
13	Place column in a new, labeled 1.5 mL tube.
14	Apply sample from step 5 to the columns.
15	Centrifuge for 1 minute at maximum speed.
16	Store samples on ice.
17	Analyze labeled cDNA using a NanoDrop spectrophotometer.
	Calculate the degree of labeling (DL) and ensure that it is between 1.5 and 3%.
18	$DL = [(340 \times \text{pmol per } \mu\text{L dye}) / (\text{ng per } \mu\text{L cDNA} \times 1000)] \times 100\%$ Calculate the volume for the amount of cDNA that you would like to hybridize (Table P2, rows 6-7 or 8-9)
Hybridization: Before starting	
19	Prepare 100X blocking agent by adding 125 µL nuclease-free water to a vial of lyophilized 10X GE Blocking Agent (supplied in the Hi-RPM Gene Express Hyb kit). Vortex and leave at room-temperature for 60 minutes before use.
20	Bring Agilent CGHblock to room temperature (supplied with Agilent ULS labeling kit)

21 Bring heat block to 95°C

Hybridization:

23 Label one 1.5 mL tube per array

Make a hybridization master mix and aliquot to each tube

24 4x44K, 1 reaction: 1.1 µL 100X GE Blocking Agent, 55 µL Hi-RPM GE Hyb Buffer
(total volume 56.1 µL)

8x60K, 1 reaction: 0.55 µL 100X GE Blocking Agent, 27.5 µL Hi-RPM GE Hyb Buffer
(28.05 µL)

25 Aliquot cDNA (volumes calculated in step 18)

26 Top-up each tube with nuclease-free water for a total volume of 83 µL (4x44K) or 41.5
µL (8x60K)

27 Incubate at 95°C for 3 minutes

28 Incubate on ice for at least 3 minutes

29 Briefly spin tubes down

30 Add Agilent CGHblock to each tube 27 µL (4x44K) or 13.5 µL (8x60K). The total
volume should be 110 µL or 55 µL, respectively.

31 Keep samples at room-temperature until the array is loaded.

32 Dispense 100 µL (4x44K) or 45 µL (8x60K) on to each array

33 Hybridize at 65°C in the dark for 17 hours at 20 rpm.

34 Use standard Agilent Gene Expression wash and scan conditions to visualize arrays.

C.5 Chapter 6

Table C8 RNA quality and yield.

Time in formalin	Animal ID	[RNA] (ng/ul)	260/280	260/230	RIN
Fresh-frozen	2	150.75	2.07	2.24	9.5
Fresh-frozen	3	153.06	2.08	2.25	9.3
Fresh-frozen	4	163.21	2.07	2.23	9
Fresh-frozen	6	153.88	2.05	2.19	9.3
Fresh-frozen	41	157.81	2.02	2.27	9.5
Fresh-frozen	42	145.53	2.06	2.16	9.2
Fresh-frozen	43	155.736	2.05	2.25	9.4
Fresh-frozen	44	148.05	2.03	2.06	8.9
18 hours	2	253.55	2.06	2	2.1
18 hours	3	315.79	2.06	1.92	2
18 hours	4	173	2.12	2.05	2.2
18 hours	6	318.29	2.05	1.99	2
18 hours	41	275.76	2.05	1.93	2
18 hours	42	429.72	2.06	1.97	2
18 hours	43	276.67	2.08	2.06	1.9
18 hours	44	300.55	2.06	2	2.3
3 weeks	2	36.82	1.79	1.33	2.4
3 weeks	3	43.14	1.91	1.84	2.4
3 weeks	4	36.89	1.81	2.11	2.4
3 weeks	6	31.87	1.78	1.94	2.4
3 weeks	41	27.11	1.85	1.88	1.3
3 weeks	42	17.28	1.89	1.88	2.2
3 weeks	43	8.8	1.74	2.32	2.15
3 weeks	44	27.48	1.87	2.03	NA

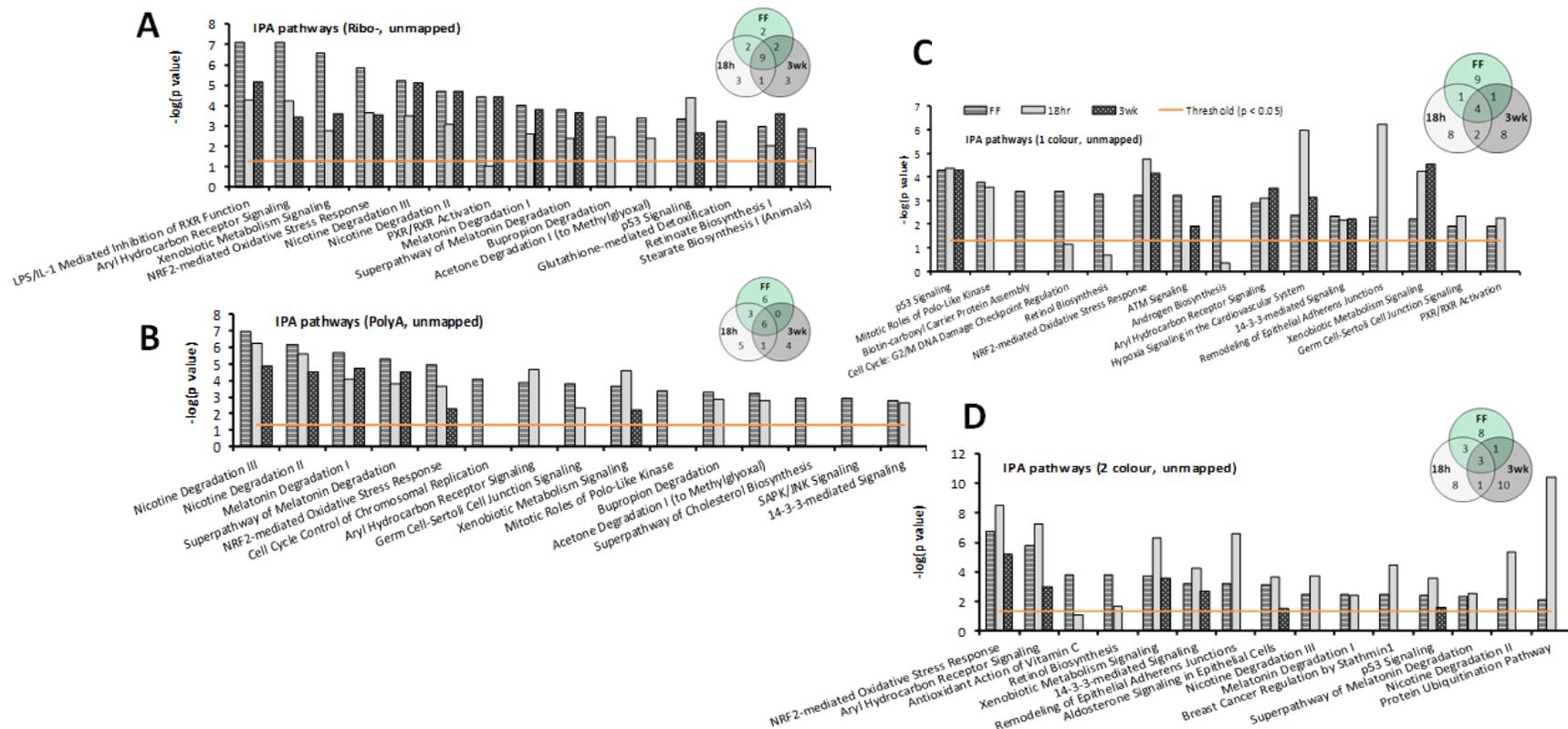


Figure C5 Pathway analysis for (A) ribo-depletion RNA-seq, (B) polyA-enrichment RNA-seq, (C) 1-colour microarrays, and (D) 2-colour microarrays, showing the top 15 enriched pathways for the fresh-frozen samples and enrichment levels of these pathways in the corresponding FFPE samples (pathway significance was calculated in IPA using a Fisher Exact test, $-\log(p \text{ value}) = 1.3$ corresponds to a $p = 0.05$). Venn diagrams show the overlap of the top 15 pathways in each FRO, 18hr-FFPE, and 3wk-FFPE. For all analyses, all differentially expressed genes (i.e., unmapped gene list) with $FDR P \leq 0.05$, fold change $> \pm 1.5$ were used.