

Identifying a transcriptional signature for cell sensitivity to the cancer chemotherapy agent, BCNU

by

Chandni Rajan Valiathan

Bachelor of Science (B.S), Computer Science
Bachelor of Science (B.S), Mathematics
Brandeis University, Waltham, MA, 2005

Submitted to the Computational and Systems Biology Program
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computational and Systems Biology at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

©2011 Chandni Rajan Valiathan; All rights reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Author _____
Chandni Rajan Valiathan
Computational and Systems Biology Program
May 20, 2011

Certified by _____
Leona Samson
Director, Center for Environmental Health Sciences
Uncas and Helen Whitaker Professor of Biological Engineering and Biology
Thesis Supervisor

Accepted by _____
Christopher Burge
Professor of Biology and Biological Engineering
Director, CSB Graduate Program

Identifying a transcriptional signature for cell sensitivity to the cancer chemotherapy agent, BCNU

by
Chandni Rajan Valiathan

Abstract

Many organisms have evolved DNA damage response mechanisms to deal with the constant damage to DNA caused by endogenous and exogenous agents. These mechanisms activate cell cycle checkpoints to allow time for DNA repair or, in the case of severely damaged DNA, initiate cell death mechanisms to maintain genomic integrity. The cell's response to DNA damaging agents includes wide spread changes in the transcriptional state of the cell that have been implicated in cell death or survival decisions. However, we do not fully understand how the multiple and sometimes opposing transcriptional signals are interpreted to make these critical decisions. A computational and systems biology approach was taken to study the wide-spread transcriptional changes induced in human cell lines after exposure to a DNA damaging and chemotherapeutic agent, 1,3-bis-(2-chloroethyl)-1-nitrosourea (BCNU or carmustine).

Cell lines with extreme sensitivity or resistance to BCNU were identified from a set of twenty four genetically diverse human lymphoblastoid cell lines using a high-throughput method that was developed as part of this thesis. This assay has broad applications and can be used to simultaneously screen multiple cell lines and drugs for accurate measurements of cell proliferation and survival after drug treatment. The assay has the advantage of having a large dynamic range that allows sensitivity measurements on a multi-log scale allowing better resolution of comparative sensitivities.

Temporal transcription profiles were measured in cell lines with extreme BCNU sensitivity or resistance to generate a large transcription data set amenable to bioinformatics analysis. A transcriptional signature of 706 genes, differentially expressed between BCNU sensitive and resistant cell lines, was identified. Network and gene ontology enrichment identified these differentially expressed genes as being involved in key DNA damage response processes like apoptosis and mitosis. Experimental evidence showed that the transcription signature correlated with observed cellular phenotypes. Furthermore, the NF-Y transcription factor binding motif was enriched in the promoter region of 62 mitosis-related genes down-regulated in BCNU sensitive but not resistant cell lines. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) confirmed NF-Y occupancy in 54 of the 62 genes, thus implicating NF-Y as a possible regulator of the observed stalling of entry into mitosis.

Using experimental and computational techniques we deciphered the functional importance of differential transcription between BCNU sensitive and resistant cell lines and identified NF-Y as an important factor in the transcriptional and phenotypic cell response to BCNU such as the control of entry into mitosis.

Thesis Advisor: Leona Samson

Title: Director, Center for Environmental Health Sciences

Uncas and Helen Whitaker Professor of Biological Engineering and Biology

Acknowledgements

Many people have contributed to and enabled me to complete this thesis. They have advised, assisted and supported me throughout the duration of my graduate years and I would like to take this opportunity to acknowledge and thank them.

First and foremost, I would like to thank my thesis advisor, Leona Samson, for giving me the opportunity to work on an exciting project. Through her advice and guidance, she has taught me to be a better scientist and also taught me how to think about, write and present scientific ideas. I also want to thank my thesis committee members, Douglas Lauffenburger, Bevin Engelward and Ernest Fraenkel for their advice on my thesis project, and William Kaufmann, my external thesis defense committee member who has kindly agreed to travel to MIT from UNC for my defense. I am thankful to Bruce Tidor who has given me excellent academic and career advice throughout my time at MIT.

Past and present Samson lab members have taught me a lot about experimental techniques and I would like to thank them for this, and for making the Samson lab a wonderful place to work at. Prajit Limsirichai helped me tremendously during the last stretch of experiments and I am extremely grateful for his dedication and diligence. I also want to thank Shmulik Motola and Adam Labadorf from the Fraenkel lab and Mohini Jangi from the Sharp lab for all the help they provided for my CHIP-seq experiments.

Facilities provided by the BiomicroCenter have been fundamental in obtaining the data used in this thesis. I would especially like to thank Stuart Levine and Manlin Luo for working with me and providing timely and high-quality microarray and sequencing data. Additionally, I would like to acknowledge the Koch Institute flow cytometry facility where all the flow cytometry data was obtained.

Besides providing excellent opportunities for research, MIT also provided me the opportunity to explore and develop my artistic side. Dancing, choreographing and performing dances with MIT Natya was an important source for my energy and I would like to thank Gayathri Srinivasan, Charuleka Varadharajan and Mohini Jangi for keeping me motivated about dance. I also want to acknowledge my friends who have been with me through good times and bad.

Finally, and most importantly, I would like to thank my family for their unconditional support. I would not be here without the love and dedication my parents, Rajan Valiathan and Mridula Valiathan, gave me. My husband, Kartik Varadarajan, has been extremely supportive and has given me excellent guidance and advice in many difficult situations. My sister Rajeshwari Valiathan and her family have always put a smile on my face even during tough times. I want to thank my family for believing in me and providing me with the emotional support I needed during the past few years.

Contents

ACKNOWLEDGEMENTS	4
FIGURES	8
TABLES	10
CHAPTER 1: INTRODUCTION	11
1.1 DNA DAMAGE AND MAINTENANCE OF GENOMIC INTEGRITY	12
1.1.1 <i>The importance of DNA repair</i>	12
1.1.2 <i>DNA damage response mechanisms</i>	13
1.1.3 <i>Cell death or survival decisions</i>	16
1.2 THE ROLE OF TRANSCRIPTION IN CELL DEATH OR SURVIVAL DECISIONS.....	18
1.2.1 <i>Transcriptional response after exposure to DNA damaging agents</i>	18
1.2.2 <i>The possible role of transcriptional kinetics in cell death or survival mechanisms</i>	20
1.3 STUDYING THE ROLE OF TRANSCRIPTIONAL KINETICS AND CONTROL IN CELL DEATH OR SURVIVAL DECISIONS AFTER DNA DAMAGE	21
1.3.1 <i>BCNU as a DNA damaging agent and a chemotherapeutic agent</i>	22
1.3.2 <i>The experimental system - a panel of genetically varied cell lines</i>	25
1.3.3 <i>Computational approaches to better understand cell death/survival decisions after DNA damage</i>	28
1.4 GOALS OF THE THESIS AND PROJECT DESIGN	32
1.5 REFERENCES	33
CHAPTER 2: A HIGH-THROUGHPUT SURVIVAL ASSAY TO MEASURE DRUG-INDUCED CYTOTOXICITY AND CELL CYCLE EFFECTS.....	42
2.1 ABSTRACT.....	43
2.2 INTRODUCTION.....	44
2.3 MATERIALS AND METHODS	46
2.3.1 <i>Cell culture</i>	46
2.3.2 <i>Determining the optimal BrdU concentration</i>	48
2.3.3 <i>Cell cycle profile analysis by flow cytometry</i>	48
2.3.4 <i>Drug treatment</i>	48

2.3.5 Bromodeoxyuridine Addition	50
2.3.6 Nuclei isolation and staining for flow cytometry	50
2.3.7 Data collection	51
2.4 RESULTS	53
2.4.1 Detecting cells that have undergone zero, one or two cell divisions after drug treatment	53
2.4.2 Calculating the fraction of proliferated cells after drug treatment	53
2.4.3 The multi-well assay has a large dynamic range, yielding log scale killing for suspension and adherent cell lines treated with a cytotoxic agent	55
2.4.4 The multi-well assay detects cell-cycle effects of drug treatment	57
2.5 DISCUSSION	60
2.6 REFERENCES	63

CHAPTER 3: BCNU INDUCED TRANSCRIPTIONAL CHANGES IDENTIFY NF-Y AS A PUTATIVE REGULATOR OF MITOTIC ENTRY 65

3.1 ABSTRACT	66
3.2 INTRODUCTION	67
3.3 MATERIALS AND METHODS	68
3.3.1 Cell culture	68
3.3.2 Proliferation assay	68
3.3.3 Cell cycle profile analysis	68
3.3.4 Drug treatment	69
3.3.5 RNA isolation and hybridization	69
3.3.6 Microarray data analysis	69
3.3.7 Annexin/7AAD assay for measuring cell death	71
3.3.8 Phospho-histone H3 for measuring the mitotic fraction	71
3.3.9 Immunoblot analysis	72
3.3.10 Chromatin immunoprecipitation followed by sequencing and peak calling	73
3.4 RESULTS	74
3.4.1 Determining the optimal BrdU dose for lymphoblastoid cells	74
3.4.2 Cell lines show no G2/M arrest after BrdU addition	76

3.4.3 The genetically varied panel of cell lines show a wide range of sensitivities to BCNU	76
3.4.4 Cell cycle profiles for cell lines with extreme BCNU sensitivity/resistance define the time frame for microarray measurements.....	92
3.4.5 Viable cell number measurements for BCNU sensitive/resistant cell lines show growth inhibition for sensitive cell lines but not resistant cell lines.....	98
3.4.6 Designing and generating a multi-dimensional transcriptional data set.....	100
3.4.7 Exploring the transcriptional data space	101
3.4.8 Identifying a general transcriptional signature for cell death/survival after DNA damage.....	104
3.4.9 Visualizing the DNA damage induced transcriptional signature	105
3.4.10 Network and canonical pathway analyses reveal the biological relevance of the transcriptional signature	105
3.4.11 The transcriptional signature can be partitioned into two functionally meaningful subsets.....	112
3.4.12 BCNU sensitive cell lines induce cell death by apoptosis	121
3.4.13 BCNU sensitive cell lines stall mitotic entry.....	125
3.4.14 p53 is activated to a greater extent in sensitive cell lines as compared to resistant cell lines	126
3.4.15 Computational promoter sequence analysis shows that a subset of down-regulated genes are enriched for NF-Y binding motifs and predicted to be novel NF-Y targets	134
3.4.16 Chromatin immunoprecipitation followed by sequencing (ChIP-seq) confirm NF-Y occupancy for novel predicted NF-Y targets	143
3.4.17 Gene Set Enrichment Analysis (GSEA) shows enrichment of acetyl-p53 repressed genes within the transcription signature.....	149
3.4.18 Immunoblot analysis shows no significant p53 acetylation in BCNU treated cells	149
3.5 DISCUSSION	154
3.6 REFERENCES	163
CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS	169
APPENDIX A:	178
APPENDIX B:	185

Figures

Figure 1.1: The wide range of cellular processes initiated by the DNA damage response.	15
Figure 1.2 Mechanism of BCNU interstrand cross-link formation	23
Figure 1.3 : The panel of cell lines shows a wide range of sensitivities to DNA alkylating agents	27
Figure 2.1: Experimental set up with anticipated timing.....	47
Figure 2.2: Data filtering and collection plots	52
Figure 2.3: Gates delineating cell cycle regions for the data analysis	54
Figure 2.4: Example of data obtained using the assay	58
Figure 2.5: Cell cycle effects of BCNU on U87MG cells	59
Figure 3.1: Determining the optimal BrdU concentration for lymphoblastoid cell lines	75
Figure 3.2: Checking for BrdU dependant G2/M arrest in Cell line 6	77
Figure 3.3: Checking for BrdU dependant G2/M arrest in Cell line 14	78
Figure 3.4: Checking for BrdU dependant G2/M arrest in Cell line 16	79
Figure 3.5: Checking for BrdU dependant G2/M arrest in Cell line 21	80
Figure 3.6: Checking for BrdU dependant G2/M arrest in Cell line 22	81
Figure 3.7: Checking for BrdU dependant G2/M arrest in Cell line 24	82
Figure 3.8: Survival curves for the 24 cell lines after BCNU treatment.....	85
Figure 3.9: Correlation between sensitivities of the 24 cell lines to MMS, MNNG and BCNU .	86
Figure 3.10: Genes whose basal expression correlates with BCNU sensitivity	88
Figure 3.11: Cell cycle progression of the BCNU sensitive cell line 4	94
Figure 3.12: Cell cycle progression of the BCNU sensitive cell line 5	95
Figure 3.13: Cell cycle progression of the BCNU resistant cell line 16.....	96
Figure 3.14: Cell cycle progression of the BCNU resistant cell line 13.....	97
Figure 3.15: Growth inhibition of cell lines showing extreme BCNU sensitivity/resistance.....	99
Figure 3.16: Transcripts that vary with time, cell line and treatment	102
Figure 3.17: Hierarchical clustering of the BCNU transcriptional signature	106
Figure 3.18: Network connectivity of the transcriptional signature	109
Figure 3.19: A p53-centric network with DNA replication, repair and cell cycle genes	110
Figure 3.20: Canonical pathways significantly enriched in the transcriptional signature	111
Figure 3.21: Gene Ontology enrichment of the up-regulated and down-regulated gene sets.....	113

Figure 3.22: Canonical pathway enrichment of the up-regulated and down-regulated gene sets	114
Figure 3.23: Top networks for the set of up-regulated genes are involved in cell death.....	117
Figure 3.24: Top two networks for the down-regulated genes are involved in cell cycle.....	119
Figure 3.25: Annexin V/ 7AAD plots for the four cell lines at 48h post BCNU treatment	123
Figure 3.26: Quantification from AnnexinV/7AAD plots at multiple times post BCNU treatment	124
Figure 3.27: Phospho-histone H3 at 48h post BCNU treatment.....	127
Figure 3.28: Quantified mitotic fraction at multiple time-points post BCNU treatment.....	128
Figure 3.29: Immunoblots probed for phosphor-p53 (ser20)	129
Figure 3.30: Total and phosphorylated p53 (Serine 20) as measured by immunoblot for BCNU treated samples.....	130
Figure 3.31: Immunoblots probed for phospho-p53 (ser15).....	131
Figure 3.32: Total and phosphorylated p53 (Serine 15) as measured by immunoblot for BCNU treated samples.....	132
Figure 3.33: Expression of canonical p53 targets from transcriptional profiles.....	133
Figure 3.34: Transcription Factor Binding Site enrichment in the down-regulated cluster using EXPANDER	136
Figure 3.35: Highly significant peaks from MACS analysis of NF-Y ChIP-seq data.....	146
Figure 3.36: Weakly significant peaks from MACS analysis of NF-Y ChIP-seq data	147
Figure 3.37: Genes with no significant peaks from MACS analysis NF-Y ChIP-seq data	148
Figure 3.38: GSEA of the transcriptional signature shows enrichment of genes repressed upon p53 acetylation	150
Figure 3.39: Etoposide treated samples probed with the antibody against acetyl-p53 (lys373)..	151
Figure 3.40: Immunoblot probed for acetyl-p53 (lysine 373)	152
Figure 3.41: p53 acetylation as measured by immunoblot after BCNU treatment.....	153
Figure 3.42: Model for repression of NF-Y mitotic targets mediated by p53 acetylation.....	162
Figure 4.1: Experimental and computational techniques driven by biologically motivated questions	172
Figure 4.2: Survival curves for cell lines 4, 5, 13 and 16 for four clinically used cancer treatments	177

Tables

Table 2.1 Example showing the process of calculating % control growth of treated samples.....	56
Table 3.1: The partition of the panel of 24 cell lines based on their doubling times.....	83
Table 3.2: 123 probesets with high positive or negative correlation to BCNU sensitivity	89
Table 3.3: Molecules in network A and B from Figure 3.23 for the up-regulated gene set	118
Table 3.4: Molecules in network A and B from Figure 3.24 for the down-regulated gene set ..	120
Table 3.5: Genes from the down-regulated gene set that are predicted to be NF-Y motif targets	137
Table 3.6: Genes from the down-regulated gene set that are enriched for the AhR motif	138
Table 3.7: Genes from the down-regulated gene set that are enriched for the GABP motif	139
Table 3.8: Genes from the down-regulated gene set that are enriched for the Nrf-1 motif.....	140
Table 3.9: Transcription factor binding motif enrichment for different promoter ranges for the up- and down-regulated genes	141
Table 3.10: Up-regulated genes containing p53 motifs in their promoter region.....	142
Table 3.11: Peaks called using MACS near the genes predicted to be NF-Y targets.....	145

Chapter 1: Introduction

1.1 DNA damage and maintenance of genomic integrity

1.1.1 The importance of DNA repair

DNA damaging agents are ubiquitous in our environment and within our bodies. Every day, these endogenous and exogenous DNA damaging agents generate up to 10^5 DNA lesions per cell in our bodies (1). Unrepaired DNA lesions pose a serious threat to our health since they lead to mutations or premature cell death which could exacerbate life-threatening diseases like cancer (2), neurodegenerative disease (3,4) and premature aging (5-7). Because of the harmful consequences of unrepaired DNA lesions, we have evolved DNA repair and damage response mechanisms that get rid of damaged DNA, thus maintaining genomic integrity and preventing disease (8,9).

The DNA damaging agents we are exposed to are of many different kinds. They include radiation from medical x-rays, UV from sunlight, chemical pollutants in the air we breathe, the food we eat and fluids we drink (10). In addition to these external agents, there are endogenous sources of DNA damage including nucleotide misincorporation during DNA replication, chemical instability of DNA (11,12) and most importantly internal processes like metabolism and the inflammatory response which generates reactive oxygen species that are extremely detrimental to DNA (13-15). These different sources of DNA damage generate various kinds of DNA lesions such as mismatches, small or bulky base adducts, single or double stranded breaks and intra- or inter-strand crosslinks.

To deal with the multitude of lesions seen by a cell at any given point in time, we have numerous types of repair mechanisms. Mismatches are detected by the mismatch repair complexes

followed by a single-strand incision that is then repaired and sealed by nucleases, polymerases and ligases (16). Small adducts can be removed by direct reversal proteins like *O*⁶-methylguanine methyl transferase (MGMT) and the alkB homolog (ALKBH) proteins. Other small lesions such as small alkylated or oxidized base lesions are generally removed by base excision repair where the base lesion is detected by a DNA glycosylase and removed. DNA is nicked at the resulting abasic site, followed by processing at the 5' and 3' ends which allows the filling and ligation of DNA (17,18). DNA distorting lesions and intrastrand lesions, such as pyrimidine dimers generated by UV, are removed by nucleotide excision repair (19), while inter-strand crosslinks such as those generated by many chemotherapeutics are repaired by the Fanconi anemia proteins, nucleotide excision repair proteins and homologous recombination proteins (19,20). Double strand breaks generated by radiation or as an intermediate during the processing of other lesions can be repaired by non-homologous end joining or homologous recombination (21-23). There are also translesion polymerases that can replicate DNA past certain poorly repaired lesions thus allowing cell survival but increasing mutation rates (24-26).

Though DNA repair is vital in maintaining genomic integrity, it does not work in isolation. In fact, there is a whole array of processes initiated after DNA damage to work in concert with repair mechanisms to ensure proper restoration and recovery of the cell after DNA damage. These processes are collectively termed as DNA damage response (DDR) mechanisms.

1.1.2 DNA damage response mechanisms

Following the detection of DNA damage, there are key DDR mechanisms initiated in our cells that amplify the DNA damage signal, recruit DNA repair proteins and trigger multiple

downstream processes that facilitate repair of damaged DNA (27-29). There are two groups of proteins that are currently thought to be key mediators of the DNA damage response; the phosphoinositol-3-kinase-like kinases DNA-PK, ATM and ATR, and members of the poly-ADP-ribose-polymerase (PARP) family, PARP1 and PARP2 (27,30,31). DNA-PK and ATM are activated in response to double strand breaks depending on the sensors that detect the damage which varies between cell types and cell cycle phase. If the heterodimer Ku70/Ku80 detects the double strand breaks, then DNAPK is recruited and preferentially initiates non-homologous end joining (32). On the other hand, if the double strand break is detected by the MRE11-RAD50-NBS1 complex (MRN), then ATM is activated, leading to preferential activation of homologous recombination (33). Single strand breaks are detected by PARP1/2 (31) whereas RPA coated single-stranded DNA generated during replication stress activates ATR (34,35). Once these mediators are activated, they in turn activate downstream proteins either by post-translational modifications of the targets or by recruiting other proteins to make these activating modifications. For example, PARP adds poly(ADP-ribose) chains to histones to initiate chromatin remodeling and recruitment of repair factors as well as chromatin modifiers such as histone deacetylases (31,36). Similarly ATM and ATR were recently shown to phosphorylate and modify proteins involved not only in repair, but in more general cellular processes like RNA splicing, metabolic signaling, cell cycle checkpoints, transcription and chromatin remodeling (37). All of these effects have been shown to be important for cell recovery upon exposure to DNA damaging agents, and the loss of any one of them results in compromised DNA repair and recovery. This suggests that the DNA damage response mechanisms are broad in nature and work hand in hand in a coordinated manner to help the cell recover from being exposed to DNA damaging agents (Figure 1.1).

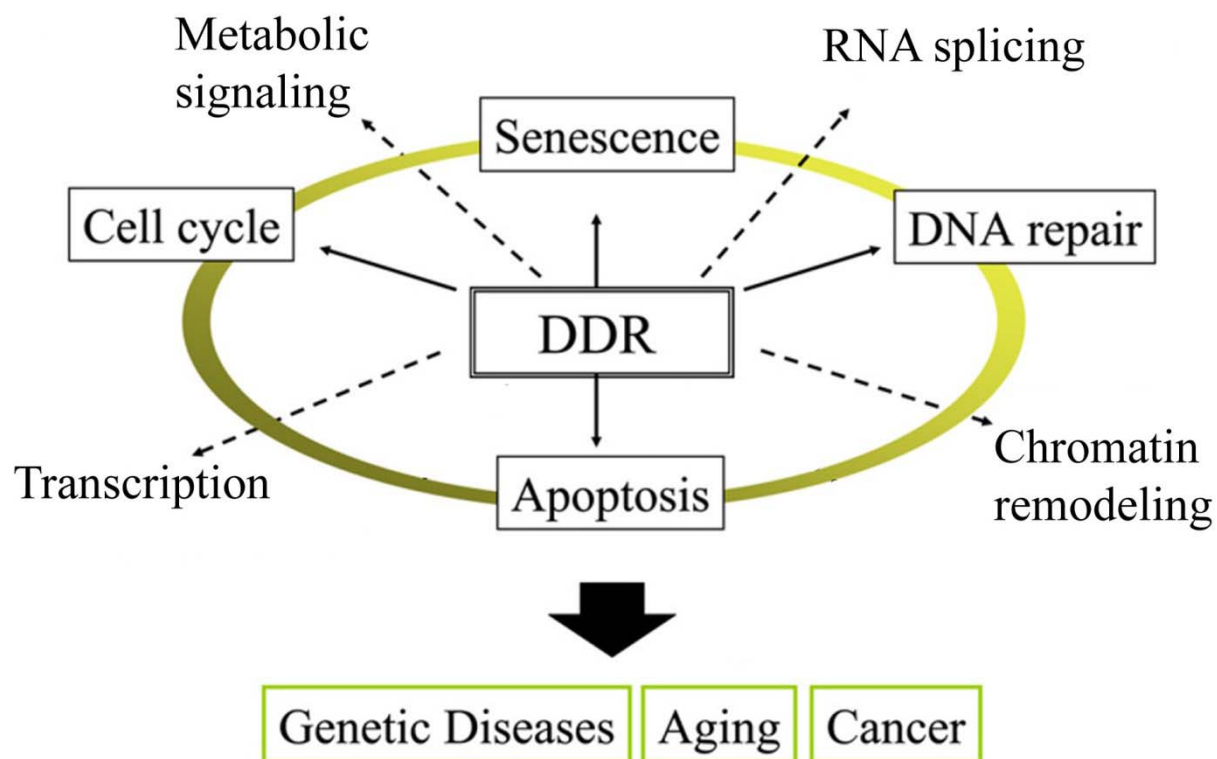


Figure 1.1: The wide range of cellular processes initiated by the DNA damage response. The figure is modified from (27) to highlight the wide range of cellular processes initiated after DNA damage (dotted arrows), the phenotypic outcome of the cell (black boxes) and the possible organism level consequences of aberrant DNA damage response mechanisms (green boxes)

The effects and importance of all the numerous and varied signals initiated after DNA damage are still not completely understood and are areas of active research. Of the numerous targets of ATM and ATR, CHK2 and CHK1 are the best studied so far. Once ATM and ATR are phosphorylated and activated, they phosphorylate CHK2 and CHK1, respectively (27,38,39). One of the functions of these checkpoint kinases is to phosphorylate and inactivate the Cdc25 phosphatases (40,41). During normal cell cycle, the Cdc25 phosphatases remove the inhibitory phosphate on their target cyclin dependent kinases thus enabling progression through cell cycle (42). Therefore, when these phosphatases are inhibited by DNA damage induced CHK2 or CHK1 phosphorylation, cells are stalled in G1, S or G2 (43), allowing time for the repair of DNA lesions, and preventing replication of damaged DNA, entry into mitosis and the segregation of damaged chromosomes.

In addition to activating cell cycle arrest, ATM and ATR also induce transcriptional changes through the direct or indirect activation of transcription factors such as P53, NF- κ B, E2F1 and Sp-1 (44-47). These transcription factors induce transcriptional changes such as regulation of cell cycle checkpoint genes, apoptotic genes or survival genes; these transcriptional responses are just as critical as DNA repair mechanisms for helping cells recover from exposure to a DNA damaging agent.

1.1.3 Cell death or survival decisions

The process by which a cell decides to die in the presence of irreparable damage is extremely important for the maintenance of genomic integrity in multi-cellular organisms. The decision against death in a cell unable to repair its damaged DNA can lead to mutations or gross

chromosomal changes that could continue to carcinogenesis. This is often observed as the increased incidence of cancer in people harboring mutations in DNA repair proteins such as ATM, BRCA1 and the Xeroderma Pigmentosum (XP) family of proteins. On the other hand, premature death of cells that have slow or low levels of DNA repair leads to unwanted cell death and premature aging for the organism (1,27,48). Therefore proper death or survival decisions at the cellular level are crucial for the whole organism's health and survival.

Although we are aware of the significance of accurate cell death or survival decisions and the many factors that affect these decisions, we do not fully understand how they integrate in the context of the cell to make these critical decisions. Experimental evidence shows that cell fate decisions vary greatly depending on the severity of damage and the cell type (49). This could be due to the intricate interplay and varying efficiencies of DNA repair, cell cycle control and cell death initiation between different cell types. All of these processes are, at least partially, regulated transcriptionally by DDR transcription factors. Interestingly, many of these DDR transcription factors are mutated in cancer. In particular, p53 is mutated in ~50% of all tumors (50). Other key DDR transcription factors such as E2F1, AP-1 and NF- κ B, are also linked to expression changes characteristic of human tumors (51-54). These observations suggest that the transcriptional response to DNA damage may be crucial in making correct cell death or survival decisions.

1.2 The role of transcription in cell death or survival decisions

1.2.1 Transcriptional response after exposure to DNA damaging agents

The prominent role of the transcriptional response to DNA damage is seen by the significant genome-wide transcriptional changes induced upon exposure to DNA damaging agents in yeast and mammalian cells, not only for genes involved in DNA damage related functions like cell cycle arrest and apoptosis, but also in other cellular processes such as protein degradation and metabolism (55-58). Moreover, in addition to the increased incidence of mutated transcription factors in cancer (59), the increased sensitivity to DNA damaging agents of yeast strains that were silenced in genes involved in transcription regulation (60,61) shows the importance of transcriptional control in protecting cells against DNA damaging agents. It is important to note that many of the DNA damaging agents used in these studies also damage other molecules in the cell. Therefore the transcriptional responses observed might also be important for the cell's response to cell-wide damage.

Some mammalian transcription factors thought to regulate these gene expression changes after damage include p53, E2F1, NF- κ B, AP-1, FKHR, ATF and c-Myc. These are by no means the only transcription factors involved in the transcriptional control of the DDR but are the few that have been studied to some extent in the context of DNA damage response. Yet, the temporal activation and the co-ordination of their targets to elicit the observed phenotype are largely unknown. Moreover, what is known is confounding; many of these transcription factors can induce both anti- and pro-apoptotic genes and therefore the outcome of their transcriptional activation after DNA damage is not easily deducible. These transcription factors also induce or repress transcriptional activity of each other, further complicating the picture.

p53 is activated by many post-translational mechanisms including phosphorylation by ATM, ATR, Chk1 or Chk2. p53 phosphorylation prevents its nuclear export and degradation, and results in its nuclear accumulation and activation (47,62-65). Activated p53 causes G1/S arrest by induction of *CDKN1A* (*p21*), the product of which is an inhibitor of the cyclin dependent kinase (*CDK*)-2 (66). Activated p53 also maintains a G2/M arrest by repressing *CCNBI* (cyclin B1) and inducing the CDK1 inhibitors *GADD45A* and *14-3-3 σ* (67). Apoptosis related genes such as *BCL2* (anti-apoptotic) plus *PUMA*, *NOXA* and *BAX* (pro-apoptotic) are transactivated by p53. Since p53 induces both pro- and anti-apoptotic genes, the outcome of p53 activation is difficult to predict (68). Like p53, E2F1 has both pro- and anti-apoptotic arms that require strict control for normal cell function. E2F1 regulates genes required for progression into S-phase and is negatively regulated by pRB which is a p53 target. Targets of E2F1 include DNA repair genes (e.g *Rad51*), signal transducers (e.g *ATM*) and pro-apoptotic genes (e.g *PUMA* and *NOXA*) (69,70). E2F1 also induces an anti-apoptotic function through indirect activation of the PI3K/AKT survival pathway (71,72). Similar to p53 and E2F1, NF- κ B is activated in response to DNA damage by IKK or ATM mediated phosphorylation of I κ B (73). NF- κ B activates anti-apoptotic genes in most cases, though in some tissues, pro-apoptotic genes are also induced. NF- κ B has also been shown to repress p53 (74).

Because these DDR transcription factors induce genes involved cell death as well as cell survival, the process by which cells decide between the survival arm and the death arm is difficult to comprehend and deduce. This decision process in the presence of two orthogonal signals is not yet understood, and is an area of active research.

1.2.2 The possible role of transcriptional kinetics in cell death or survival mechanisms

The transcriptional response to DNA damage is inherently dynamic because of the kinetics implicit in upstream signaling, varying rates of mRNA transcription and translation as well as the cross-regulation and feed-back loops within transcriptional networks. Moreover, as DNA repair progresses, the initiating signal - the amount and extent of DNA damage - also changes dynamically thus altering downstream transcription control.

Because of the dynamic nature of the transcriptional response to DNA damage, the kinetics of transcription might play an important role in the phenotypic outcome of the cell. For example, the strength of induction of anti- and pro-apoptotic genes may differ, thus enabling one of the processes to win out over the other as time progresses. Similarly, there might be a difference in transcription rates or mRNA degradation rates for genes involved in the two opposing arms of the cell decision process. Additionally, the level of transcription factor activation may vary depending upon the severity of the damage present, which, in the presence of active DNA repair, will also change over time. This could confer different transcript induction patterns at different times post DNA damage, thus influencing the cell death/survival decision in the cell.

Temporal transcription patterns could also play a role in keeping track of the time elapsed since damage, an important aspect in the decision between continued cell cycle arrest or cell death. For example if DNA damage persisted for a long time, the continued induction of apoptotic genes or continued repression of anti-apoptotic genes, could drive the cell to induce cell death.

1.3 Studying the role of transcriptional kinetics and control in cell death or survival decisions after DNA damage

The cell death/survival decision process after DNA damage is a complex one, involving many signaling proteins as well as drastic changes in the transcriptional state of the cell (37,57). Other complex biological processes, such as signaling cascades affecting cell death or growth in response to external stimuli, have previously been successfully studied using systems measurements and computational modeling techniques (75-77). Stimulated by these studies, we decided to apply a systems-level approach to better understand the temporal transcriptional coordination of various DNA damage response pathways as well as the involvement of a wide range of cellular processes in cell death/survival decision processes.

With the knowledge that significant genome-wide transcriptional changes occur after DNA damage, we decided to measure global transcriptional changes in human cells after DNA damage induced by a chemotherapeutic agent, 1,3-bis-(2-chloroethyl)-1-nitrosourea (BCNU). Our decision to study transcription was motivated not only by the potential role of transcription in cell decision processes but also by the fact that, at the beginning of the project, whole genome microarray technology was the accepted and readily available method available for systems-wide transcription measurements. Computational methods could then be applied to the large data set generated in this manner, to extract pertinent information about the transcriptional signatures for cell death or survival after DNA damage. Additionally, the unbiased nature of genome-wide measurements would also facilitate the discovery of novel pathways and processes important in the decision between cell death and survival.

The approach we took to study the transcriptional control of cell death or survival after DNA damage was to compare transcriptional kinetics between DNA damage sensitive cells that are destined to die and resistant cells that are destined to survive in the presence of damage, anticipating that this might give us insight into temporal gene expression changes related to cell death or survival. This approach has been used successfully by us and others to identify genes that can accurately differentiate between distinct populations (78-81). The DNA damaging agent, cell lines and computational techniques we used to generate a genome-wide transcriptional data set and study transcriptional kinetics in damage sensitive or resistant cell lines are discussed below.

1.3.1 BCNU as a DNA damaging agent and a chemotherapeutic agent

1.3.1.1 Cellular reactions of BCNU

BCNU is a member of the chloroethylnitrosoureas (CENUs), known to generate nucleophilic chloroethylenium ions in aqueous solutions that react with DNA to produce a wide range of DNA lesions (82). CENUs also react with and inhibit protein function by carbamylating them at cysteine residues.

Although BCNU can react with both cellular proteins and DNA, the major biologically and clinically relevant reactions are thought to be the DNA interstrand crosslinks generated by BCNU (83). These crosslinks are generated in a multi-step process (Figure 1.2), the first of which is the formation of the O^6 -chloroethylguanine DNA lesion (84). This intermediate undergoes intramolecular rearrangement to form 1- O^6 -ethanoguanine which in turn reacts with cytosine in the complementary DNA strand to form a G-C interstrand crosslink (82,84). The final interstrand crosslinks are hard to repair and contribute to cellular cytotoxicity by BCNU.

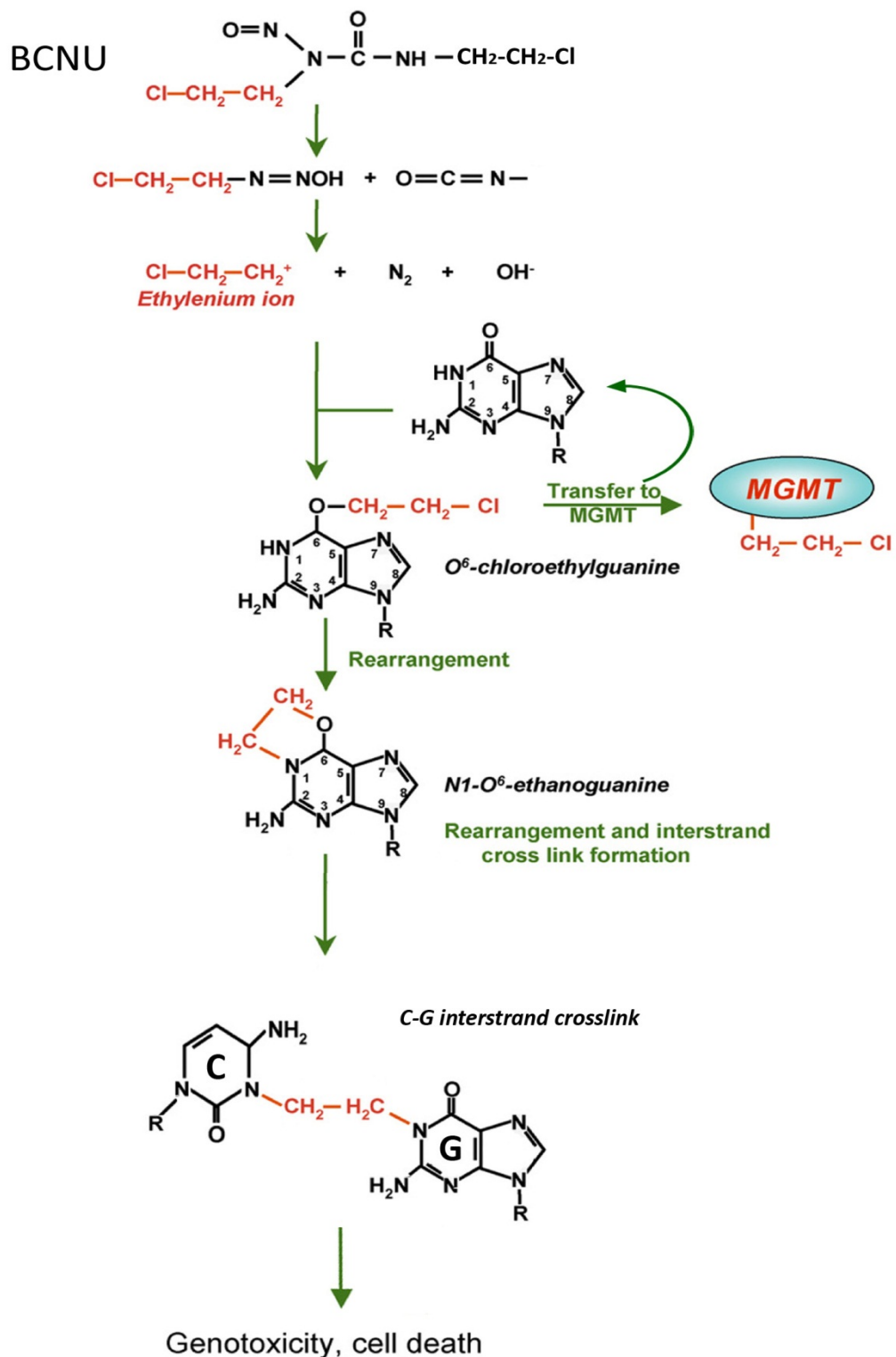


Figure 1.2 Mechanism of BCNU interstrand cross-link formation

The steps in the formation of interstrand cross-links by BCNU. The figure also shows the intermediate that can be successfully removed by MGMT (modified from 150).

1.3.1.2 DNA damage response mechanisms initiated by BCNU lesions

Multiple studies have shown that the repair of O^6 -chloroethylguanine lesions by the O^6 -methylguanine DNA methyltransferase protein (MGMT) prevents DNA interstrand crosslink formation. Cells expressing high levels of the MGMT protein are, in fact, resistant to BCNU treatment (85-87). In the absence of MGMT, the reaction proceeds to completion and interstrand crosslinks are formed. These ICLs are extremely toxic because they induce severe replication stress, inhibit transcription and are hard to repair and resolve. The repair mechanisms employed to remove these ICLs are not known yet but are thought to include nucleotide excision repair and ICL repair mediated by the Fanconi anemia proteins and homologous recombination (88-90). Base excision repair is also thought to play a role in protecting cells from BCNU cytotoxicity (91).

1.3.1.3 Chemotherapeutic application of BCNU

The DNA inter-strand cross-links are particularly cytotoxic for cells undergoing frequent replication due to stalled replication and the formation of double-strand breaks (92,93). Because of its preferential cytotoxicity for rapidly dividing cells, BCNU is commonly used to treat glioblastomas. BCNU is lipophilic and can therefore easily cross through the blood brain barrier, making it more suitable for treating glioblastomas (94,95). Currently, the treatment regimen includes combination therapy with another DNA alkylating agent, temozolomide, radiotherapy and surgical resection. Even with these combination therapies, the average survival time of the patients is about one year (96-98).

One of the drawbacks of BCNU is that tumor cells with functional MGMT are resistant to low doses of the drug (99,100). These types of tumors have been targeted by administering the

MGMT inhibitor, *O*⁶-benzylguanine, along with BCNU (101). Besides MGMT-mediated resistance of BCNU, another undesired outcome of BCNU treatment is the drastic side effects it causes such as myelosuppression (98).

Few studies have measured the effects of BCNU treatment in healthy human cell lines. The DNA response mechanisms initiated in healthy cells, the mechanisms of cell cycle arrest or cell death and the factors responsible for the difference in side effects observed between patients in the clinic are not known. Further study could reveal other factors that may improve BCNU treatment strategies by identifying tumors that will respond to BCNU treatment and patients that will have less severe side effects. Moreover, knowledge of factors that render blood and epithelial cells sensitive to BCNU could potentially identify targets to be inhibited or manipulated to ameliorate the severe and adverse side effects that currently limit cancer therapy.

With this in mind, in this study, we used a panel of 24 cell lines derived from healthy, ethnically diverse humans to study factors that affect the sensitivity of healthy cells to BCNU, and to better understand the difference between cells that are sensitive and resistant in terms of their transcriptional kinetics, cell death initiation mechanism and cell cycle arrest after BCNU treatment.

1.3.2 The experimental system - a panel of genetically varied cell lines

A panel of twenty-four genetically diverse cell lines was used for this study in the hope that they could be used to find novel genes responsible for the cell death/survival decision mechanism. The 24 cell lines are a subset of a larger set of 450 lymphoblastoid cell lines derived from the

blood of healthy humans with ancestry from around the globe, and that maintains the genetic diversity of the panel. These cell lines were developed from B lymphocytes by EBV transformation. It has been observed that EBV transformation maintains p53 function and does not induce p53 mutations (102). Therefore, we believe that these cells induce proper DNA damage response mechanisms. Additionally, these cell lines have varying basal expression levels of hundreds of genes that could provide new information on novel mechanisms for cell death or survival after DNA damage.

1.3.2.1 Cell survival of the panel of cell lines to various DNA damaging agents

Previous studies showed that the 24 cell lines showed a wide range of sensitivity spanning from ~10% to ~90% control growth when exposed to the alkylating agent N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), an alkylating agent (79). By studying the expression of the four most sensitive and four most resistant cell lines, we found that there was a strong basal gene expression signature of 48 genes that could accurately predict the sensitivity of the remaining cell lines to MNNG. The panel of cell lines also showed a broad range of sensitivity to methylmethanesulfonate (MMS), another alkylating agent (Figure 1.3). Interestingly, the cell lines do not maintain order of sensitivity so that cell lines that were considered resistant for MNNG are not necessarily resistant to MMS treatment and vice versa. These observations show that the genetic variation present in the panel of cell lines provides an interesting and informative experimental system in which to study resistance or sensitivity to DNA damage.

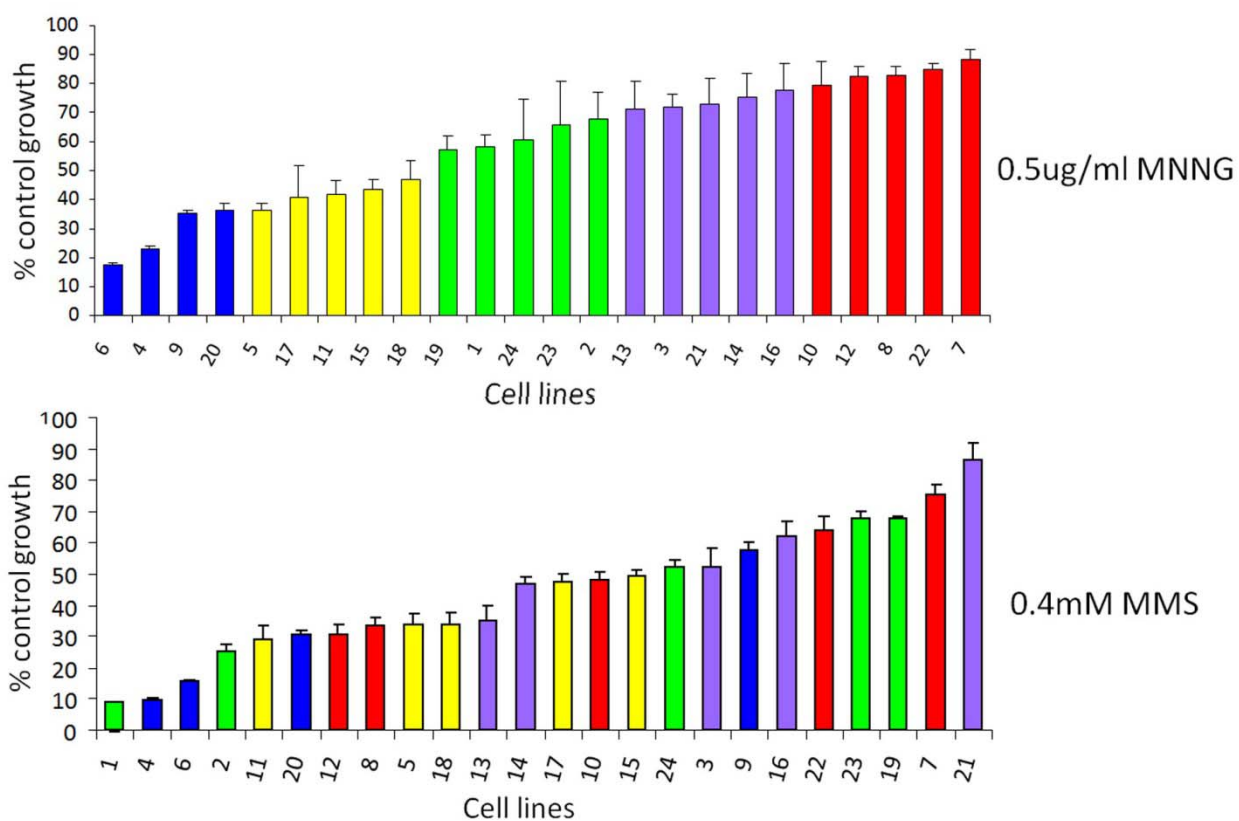


Figure 1.3 : The panel of cell lines shows a wide range of sensitivities to DNA alkylating agents MNNG sensitivity (79) (top) and MMS sensitivity (Unpublished data, Samson Lab) of the panel of cell lines. Cell lines are colored according to MNNG sensitivity to show the different order of sensitivity for MMS.(figure from Leona Samson)

1.3.2.2 Studying DNA damage recovery in the panel of cell lines

Based on the wide range of sensitivity observed for MNNG and MMS, we hypothesized that the panel of cell lines would show a similarly broad range of sensitivity to other DNA damaging agents. In particular we were interested in studying the range of sensitivity of these cell lines to the clinically prescribed DNA damaging agent, BCNU, used in cancer therapy. If indeed this was the case, we would have a set of cell lines with extreme BCNU sensitivity or resistance. The extremely sensitive cell lines would provide a platform to study transcriptional kinetics in cells that decide to die. Similarly, transcriptional kinetics measured in extremely resistant cells would provide insight into the decision to survive. With this as our experimental goal, we decided to screen the panel of cell lines for sensitivity to BCNU, and measure BCNU induced transcriptional changes in the two most sensitive and two most resistant cell lines to gain insight into the transcriptional control of cell death or survival after DNA damage.

1.3.3 Computational approaches to better understand cell death/survival decisions after DNA damage

Our goal was to use computational techniques to mine a large transcriptional data set to better understand transcriptional kinetics in the presence of DNA damage. A major challenge in microarray data analysis is the identification of meaningful signals from the confounding noise. This problem of feature selection has been studied extensively and is an area of active research. Some proposed methods of feature selection from microarray data include the use of genetic algorithms (103,104), principle component analysis (105-109), Bayesian methods (110,111) and numerous other techniques (112-114). These techniques try to identify the minimal subset of genes that incorporate the maximum information content from the data set.

Although such minimal subsets have applicability as biomarkers, they are not often sufficient or necessary to understand all of the important aspects about the biological system. Instead, simpler techniques such as statistically significant fold-changes are commonly used to identify genes that are differentially expressed between two biologically distinct conditions. We, and others, have successfully used this method on data sets consisting of multiple comparable measurements for each of two conditions (78-81). Some studies have also used statistically significant fold-change on time-series data to identify genes that are induced or repressed at a particular time-point between treated and control samples (115,116). Another commonly used technique for gene selection from time-series data is analysis of variance (ANOVA) (117-120). With ANOVA, genes that vary across multiple dimensions can be identified and studied in detail.

Once an interesting and meaningful set of genes have been identified, network and enrichment algorithms are commonly used to identify the underlying pathways and cellular process that are described by a selected set of genes. For example numerous algorithms have been described that build networks from gene expression data based on co-expression, mutual information or biological functions (121-126). Additionally, platforms like Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) can be used to build networks from data bases containing curated gene and protein interaction or regulation information. These methods allow the analysis of selected genes in terms of their underlying biology, thus facilitating a better understanding of their involvement in the biological system being studied.

Besides gene-set selection and network identification, time-series microarray data is amenable to analysis by various other computational techniques, the choice of which depends greatly on the

biological question of interest. For example differential equation models of transcript kinetics have been used to predict transcription factor activity from gene expression measurements (127,128). Hidden Markov Models have also been used to build time-series expression trees separating or clustering together genes with similar expression (129). This and other clustering techniques give us the power to identify transcripts that follow similar expression patterns across multiple conditions, thus revealing genes that are co-regulated or that form interconnected modules involved in DDR. In fact, previous studies have used clustering analyses with expression data in yeast to identify core stress-response pathways or proteins induced under various stresses (130-132).

For our data set, we implemented many of the afore-mentioned computational techniques to understand transcriptional changes upon exposure to a DNA damaging agent. First, we used statistical techniques to extract meaningful signals from the confounding noise. The statistical analyses were complemented with network and enrichment techniques to better understand cellular modules involved in the transcriptional response to DNA damage. These analyses directed us towards specific pathways that are transcriptionally activated or repressed to a greater extent in the sensitive cell lines as compared to the resistant cell lines after BCNU treatment.

The transcriptional signature identified from our data was further used to identify transcription factors that might control expression of genes in the signature set. Transcription factors act as integrators of signals and are key regulators that can change the state of the cell based on the signals. Knowledge of possible transcriptional regulators of the transcription signature in our

data set would give us insight into mechanisms that activate the cell death signature in cells that decide to die after DNA damage.

Computational techniques have previously been used to predict putative transcriptional regulation of co-expressed gene sets in yeast as well as human cells (129,133,134). These techniques have been successful in yeast studies because of simpler transcriptional control in these single-celled organisms. Transcriptional regulation is more complex in human cells because of the increased number of transcription factors and the presence of more complicated and dynamic combinatorial regulatory complexes (135). However, our knowledge about transcriptional regulation in humans is continuously being improved. Chromatin immunoprecipitation followed either by microarray hybridization (ChIP-chip) or sequencing (ChIP-seq) experiments have provided data-sets to identify high confidence transcription factor binding motifs for many transcription factors in various cell types. These data have been curated in databases like TRANSFAC (136-139) and JASPAR (140-143), from which high confidence position weight matrices have been calculated for transcription factors with available data. Many computational techniques take advantage of these experimentally determined position weight matrices to predict putative transcription factors that regulate a set of co-expressed genes (144,145). For our data set, we used one such algorithm, PRIMA, which is built into the EXPANDER (146-149) package, to identify transcription factors whose binding sites were enriched within the promoter regions of genes in our signature set.

These computational techniques allowed the identification and interpretation of a transcriptional signature for BCNU induced cell death, and also provided a putative model for the transcriptional control of a subset of the transcriptional signature.

1.4 Goals of the thesis and project design

Using the set of 24 cell lines, DNA damaging agent and computational techniques described, the goal of this thesis was to identify transcriptional signatures that are characteristic of cell death or cell survival after DNA damage. This was systematically pursued following the steps below: i) a high-throughput assay was developed to measure the sensitivity of lymphoblastoid cell lines to genotoxic and cytotoxic agents; ii) The panel of 24 genetically varied lymphoblastoid cell lines were screened for sensitivity to BCNU; iii) a multi-dimensional data set was generated by making temporal transcriptional profile measurements in the two most sensitive and the two most resistant cell lines after BCNU treatment; iv) computational techniques were used to identify a biologically meaningful transcriptional signature for cell death after BCNU induced DNA damage; v) various functions represented in the transcriptional signature were experimentally shown to explain the phenotypic cell behavior after DNA damage; vi) using computational and experimental techniques, Nuclear Factor Y was identified as a putative regulator of a subset of genes in the cell death signature, and a possible model for transcriptional regulation was proposed.

These steps are described in detail in Chapter 2 (step (i)) and Chapter 3 (steps (ii) to (vi)).

1.5 References

1. Hoeijmakers, J.H. (2009) DNA damage, aging, and cancer. *N Engl J Med*, **361**, 1475-1485.
2. Bartek, J., Bartkova, J. and Lukas, J. (2007) DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene*, **26**, 7773-7779.
3. Wilson, D.M., 3rd and Bohr, V.A. (2007) The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair (Amst)*, **6**, 544-559.
4. Biton, S., Barzilai, A. and Shiloh, Y. (2008) The neurological phenotype of ataxia-telangiectasia: solving a persistent puzzle. *DNA Repair (Amst)*, **7**, 1028-1038.
5. Schumacher, B., Garinis, G.A. and Hoeijmakers, J.H. (2008) Age to survive: DNA damage and aging. *Trends Genet*, **24**, 77-85.
6. Campisi, J. (2005) Aging, tumor suppression and cancer: high wire-act! *Mech Ageing Dev*, **126**, 51-58.
7. Akbari, M. and Krokan, H.E. (2008) Cytotoxicity and mutagenicity of endogenous DNA base lesions as potential cause of human aging. *Mech Ageing Dev*, **129**, 353-365.
8. Ciccia, A. and Elledge, S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol Cell*, **40**, 179-204.
9. Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A. and Ellenberger, T. (2006) *DNA damage, repair and mutagenesis*. 2 ed. American Society for Microbiology.
10. Jackson, S.P. and Bartek, J. (2009) The DNA-damage response in human biology and disease. *Nature*, **461**, 1071-1078.
11. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709-715.
12. Lindahl, T. and Barnes, D.E. (2000) Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol*, **65**, 127-133.
13. De Bont, R. and van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*, **19**, 169-185.
14. Kawanishi, S., Hiraku, Y., Pinlaor, S. and Ma, N. (2006) Oxidative and nitrative DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis. *Biol Chem*, **387**, 365-372.
15. Valko, M., Rhodes, C.J., Moncol, J., Izakovic, M. and Mazur, M. (2006) Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact*, **160**, 1-40.
16. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*, **7**, 335-346.
17. David, S.S., O'Shea, V.L. and Kundu, S. (2007) Base-excision repair of oxidative DNA damage. *Nature*, **447**, 941-950.
18. Caldecott, K.W. (2004) DNA single-strand breaks and neurodegeneration. *DNA Repair (Amst)*, **3**, 875-882.
19. Gillet, L.C. and Scharer, O.D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev*, **106**, 253-276.
20. Moldovan, G.L. and D'Andrea, A.D. (2009) How the fanconi anemia pathway guards the genome. *Annu Rev Genet*, **43**, 223-249.
21. Shrivastav, M., De Haro, L.P. and Nickoloff, J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res*, **18**, 134-147.

22. Kanaar, R., Wyman, C. and Rothstein, R. (2008) Quality control of DNA break metabolism: in the 'end', it's a good thing. *EMBO J*, **27**, 581-588.
23. O'Driscoll, M. and Jeggo, P.A. (2006) The role of double-strand break repair - insights from human genetics. *Nat Rev Genet*, **7**, 45-54.
24. Andersen, P.L., Xu, F. and Xiao, W. (2008) Eukaryotic DNA damage tolerance and translesion synthesis through covalent modifications of PCNA. *Cell Res*, **18**, 162-173.
25. Loeb, L.A. and Monnat, R.J., Jr. (2008) DNA polymerases and human disease. *Nat Rev Genet*, **9**, 594-604.
26. Lange, S.S., Takata, K. and Wood, R.D. (2011) DNA polymerases and cancer. *Nat Rev Cancer*, **11**, 96-110.
27. Harper, J.W. and Elledge, S.J. (2007) The DNA damage response: ten years after. *Mol Cell*, **28**, 739-745.
28. Rouse, J. and Jackson, S.P. (2002) Interfaces between the detection, signaling, and repair of DNA damage. *Science*, **297**, 547-551.
29. Harrison, J.C. and Haber, J.E. (2006) Surviving the breakup: the DNA damage checkpoint. *Annu Rev Genet*, **40**, 209-235.
30. Meek, K., Dang, V. and Lees-Miller, S.P. (2008) DNA-PK: the means to justify the ends? *Adv Immunol*, **99**, 33-58.
31. Schreiber, V., Dantzer, F., Ame, J.C. and de Murcia, G. (2006) Poly(ADP-ribose): novel functions for an old molecule. *Nat Rev Mol Cell Biol*, **7**, 517-528.
32. Mahaney, B.L., Meek, K. and Lees-Miller, S.P. (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J*, **417**, 639-650.
33. Williams, R.S., Williams, J.S. and Tainer, J.A. (2007) Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. *Biochem Cell Biol*, **85**, 509-520.
34. Cimprich, K.A. and Cortez, D. (2008) ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol*, **9**, 616-627.
35. Byun, T.S., Pacek, M., Yee, M.C., Walter, J.C. and Cimprich, K.A. (2005) Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev*, **19**, 1040-1052.
36. Polo, S.E., Kaidi, A., Baskcomb, L., Galanty, Y. and Jackson, S.P. (2010) Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J*, **29**, 3130-3139.
37. Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., 3rd, Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lereenthal, Y. *et al.* (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*, **316**, 1160-1166.
38. Abraham, R.T. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes Dev*, **15**, 2177-2196.
39. Bartek, J. and Lukas, J. (2003) Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell*, **3**, 421-429.
40. Furnari, B., Blasina, A., Boddy, M.N., McGowan, C.H. and Russell, P. (1999) Cdc25 inhibited in vivo and in vitro by checkpoint kinases Cds1 and Chk1. *Mol Biol Cell*, **10**, 833-845.

41. Furnari, B., Rhind, N. and Russell, P. (1997) Cdc25 mitotic inducer targeted by chk1 DNA damage checkpoint kinase. *Science*, **277**, 1495-1497.
42. Draetta, G. and Eckstein, J. (1997) Cdc25 protein phosphatases in cell proliferation. *Biochim Biophys Acta*, **1332**, M53-63.
43. Stark, G.R. and Taylor, W.R. (2006) Control of the G2/M transition. *Mol Biotechnol*, **32**, 227-248.
44. Wu, Z.H., Shi, Y., Tibbetts, R.S. and Miyamoto, S. (2006) Molecular linkage between the kinase ATM and NF-kappaB signaling in response to genotoxic stimuli. *Science*, **311**, 1141-1146.
45. Carcagno, A.L., Ogara, M.F., Sonzogni, S.V., Marazita, M.C., Sirkin, P.F., Ceruti, J.M. and Canepa, E.T. (2009) E2F1 transcription is induced by genotoxic stress through ATM/ATR activation. *IUBMB Life*, **61**, 537-543.
46. Olofsson, B.A., Kelly, C.M., Kim, J., Hornsby, S.M. and Azizkhan-Clifford, J. (2007) Phosphorylation of Sp1 in response to DNA damage by ataxia telangiectasia-mutated kinase. *Mol Cancer Res*, **5**, 1319-1330.
47. Shieh, S.Y., Taya, Y. and Prives, C. (1999) DNA damage-inducible phosphorylation of p53 at N-terminal sites including a novel site, Ser20, requires tetramerization. *EMBO J*, **18**, 1815-1823.
48. Andressoo, J.O. and Hoeijmakers, J.H. (2005) Transcription-coupled repair and premature ageing. *Mutat Res*, **577**, 179-194.
49. Gudkov, A.V. and Komarova, E.A. (2003) The role of p53 in determining sensitivity to radiotherapy. *Nat Rev Cancer*, **3**, 117-129.
50. Vousden, K.H. and Lu, X. (2002) Live or let die: the cell's response to p53. *Nat Rev Cancer*, **2**, 594-604.
51. Xiong, H.Q., Abbruzzese, J.L., Lin, E., Wang, L., Zheng, L. and Xie, K. (2004) NF-kappaB activity blockade impairs the angiogenic potential of human pancreatic cancer cells. *Int J Cancer*, **108**, 181-188.
52. Campisi, J. (2001) Cellular senescence as a tumor-suppressor mechanism. *Trends Cell Biol*, **11**, S27-31.
53. Liu, Y., Ludes-Meyers, J., Zhang, Y., Munoz-Medellin, D., Kim, H.T., Lu, C., Ge, G., Schiff, R., Hilsenbeck, S.G., Osborne, C.K. *et al.* (2002) Inhibition of AP-1 transcription factor causes blockade of multiple signal transduction pathways and inhibits breast cancer growth. *Oncogene*, **21**, 7680-7689.
54. Libermann, T.A. and Zerbini, L.F. (2006) Targeting transcription factors for cancer gene therapy. *Curr Gene Ther*, **6**, 17-33.
55. Friedberg, E.C. (2006) *DNA repair and mutagenesis*. 2nd ed. ASM Press, Washington, D.C.
56. Jelinsky, S.A. and Samson, L.D. (1999) Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc Natl Acad Sci U S A*, **96**, 1486-1491.
57. Jen, K.Y. and Cheung, V.G. (2003) Transcriptional response of lymphoblastoid cells to ionizing radiation. *Genome Res*, **13**, 2092-2100.
58. Zhou, T., Chou, J., Mullen, T.E., Elkon, R., Zhou, Y., Simpson, D.A., Bushel, P.R., Paules, R.S., Lobenhofer, E.K., Hurban, P. *et al.* (2007) Identification of primary transcriptional regulation of cell cycle-regulated genes upon DNA damage. *Cell Cycle*, **6**, 972-981.

59. Mees, C., Nemunaitis, J. and Senzer, N. (2009) Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. *Cancer Gene Ther*, **16**, 103-112.
60. Begley, T.J., Rosenbach, A.S., Ideker, T. and Samson, L.D. (2002) Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res*, **1**, 103-112.
61. Begley, T.J., Rosenbach, A.S., Ideker, T. and Samson, L.D. (2004) Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. *Mol Cell*, **16**, 117-125.
62. Tibbetts, R.S., Brumbaugh, K.M., Williams, J.M., Sarkaria, J.N., Cliby, W.A., Shieh, S.Y., Taya, Y., Prives, C. and Abraham, R.T. (1999) A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev*, **13**, 152-157.
63. Unger, T., Juven-Gershon, T., Moallem, E., Berger, M., Vogt Sionov, R., Lozano, G., Oren, M. and Haupt, Y. (1999) Critical role for Ser20 of human p53 in the negative regulation of p53 by Mdm2. *EMBO J*, **18**, 1805-1814.
64. Shieh, S.Y., Ahn, J., Tamai, K., Taya, Y. and Prives, C. (2000) The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev*, **14**, 289-300.
65. Chehab, N.H., Malikzay, A., Appel, M. and Halazonetis, T.D. (2000) Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. *Genes Dev*, **14**, 278-288.
66. el-Deiry, W.S., Harper, J.W., O'Connor, P.M., Velculescu, V.E., Canman, C.E., Jackman, J., Pietenpol, J.A., Burrell, M., Hill, D.E., Wang, Y. *et al.* (1994) WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. *Cancer Res*, **54**, 1169-1174.
67. Taylor, W.R. and Stark, G.R. (2001) Regulation of the G2/M transition by p53. *Oncogene*, **20**, 1803-1815.
68. Fei, P. and El-Deiry, W.S. (2003) P53 and radiation responses. *Oncogene*, **22**, 5774-5783.
69. Hershko, T. and Ginsberg, D. (2004) Up-regulation of Bcl-2 homology 3 (BH3)-only proteins by E2F1 mediates apoptosis. *J Biol Chem*, **279**, 8627-8634.
70. Berkovich, E. and Ginsberg, D. (2003) ATM is a target for positive regulation by E2F-1. *Oncogene*, **22**, 161-167.
71. Ginsberg, D. (2002) E2F1 pathways to apoptosis. *FEBS Lett*, **529**, 122-125.
72. Chaussepied, M. and Ginsberg, D. (2004) Transcriptional regulation of AKT activation by E2F. *Mol Cell*, **16**, 831-837.
73. Habraken, Y. and Piette, J. (2006) NF-kappaB activation by double-strand breaks. *Biochem Pharmacol*, **72**, 1132-1141.
74. Wu, H. and Lozano, G. (1994) NF-kappa B activation of p53. A potential mechanism for suppressing cell growth in response to stress. *J Biol Chem*, **269**, 20067-20074.
75. Gaudet, S., Janes, K.A., Albeck, J.G., Pace, E.A., Lauffenburger, D.A. and Sorger, P.K. (2005) A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol Cell Proteomics*, **4**, 1569-1590.
76. Janes, K.A., Kelly, J.R., Gaudet, S., Albeck, J.G., Sorger, P.K. and Lauffenburger, D.A. (2004) Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J Comput Biol*, **11**, 544-561.

77. Miller-Jensen, K., Janes, K.A., Brugge, J.S. and Lauffenburger, D.A. (2007) Common effector processing mediates cell-specific responses to stimuli. *Nature*, **448**, 604-608.
78. Fry, R.C., Navasumrit, P., Valiathan, C., Svensson, J.P., Hogan, B.J., Luo, M., Bhattacharya, S., Kandjanapa, K., Soontararuks, S., Nookabkaew, S. *et al.* (2007) Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *PLoS Genet*, **3**, e207.
79. Fry, R.C., Svensson, J.P., Valiathan, C., Wang, E., Hogan, B.J., Bhattacharya, S., Bugni, J.M., Whittaker, C.A. and Samson, L.D. (2008) Genomic predictors of interindividual differences in response to DNA damaging agents. *Genes Dev*, **22**, 2621-2626.
80. Tseng, Y.H., Butte, A.J., Kokkotou, E., Yechoor, V.K., Taniguchi, C.M., Kriauciunas, K.M., Cypess, A.M., Niinobe, M., Yoshikawa, K., Patti, M.E. *et al.* (2005) Prediction of preadipocyte differentiation by gene expression reveals role of insulin receptor substrates and neclin. *Nat Cell Biol*, **7**, 601-611.
81. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-537.
82. Ludlum, D.B. (1990) DNA alkylation by the haloethylnitrosoureas: nature of modifications produced and their enzymatic repair or removal. *Mutat Res*, **233**, 117-126.
83. Gonzaga, P.E., Potter, P.M., Niu, T.Q., Yu, D., Ludlum, D.B., Rafferty, J.A., Margison, G.P. and Brent, T.P. (1992) Identification of the cross-link between human O6-methylguanine-DNA methyltransferase and chloroethylnitrosourea-treated DNA. *Cancer Res*, **52**, 6052-6058.
84. Tong, W.P., Kirk, M.C. and Ludlum, D.B. (1982) Formation of the cross-link 1-[N3-deoxycytidyl],2-[N1-deoxyguanosinyl]ethane in DNA treated with N,N'-bis(2-chloroethyl)-N-nitrosourea. *Cancer Res*, **42**, 3102-3105.
85. Erickson, L.C., Laurent, G., Sharkey, N.A. and Kohn, K.W. (1980) DNA cross-linking and monoadduct repair in nitrosourea-treated human tumour cells. *Nature*, **288**, 727-729.
86. Samson, L., Derfler, B. and Waldstein, E.A. (1986) Suppression of human DNA alkylation-repair defects by *Escherichia coli* DNA-repair genes. *Proc Natl Acad Sci U S A*, **83**, 5607-5610.
87. Robins, P., Harris, A.L., Goldsmith, I. and Lindahl, T. (1983) Cross-linking of DNA induced by chloroethylnitrosourea is presented by O6-methylguanine-DNA methyltransferase. *Nucleic Acids Res*, **11**, 7743-7758.
88. De Silva, I.U., McHugh, P.J., Clingen, P.H. and Hartley, J.A. (2000) Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Mol Cell Biol*, **20**, 7980-7990.
89. Vasquez, K.M. (2010) Targeting and processing of site-specific DNA interstrand crosslinks. *Environ Mol Mutagen*, **51**, 527-539.
90. McHugh, P.J., Spanswick, V.J. and Hartley, J.A. (2001) Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet Oncol*, **2**, 483-490.
91. Engelward, B.P., Dreslin, A., Christensen, J., Huszar, D., Kurahara, C. and Samson, L. (1996) Repair-deficient 3-methyladenine DNA glycosylase homozygous mutant mouse cells have increased sensitivity to alkylation-induced chromosome damage and cell killing. *EMBO J*, **15**, 945-952.

92. Dronkert, M.L. and Kanaar, R. (2001) Repair of DNA interstrand cross-links. *Mutat Res*, **486**, 217-247.
93. Muniandy, P.A., Liu, J., Majumdar, A., Liu, S.T. and Seidman, M.M. (2010) DNA interstrand crosslink repair in mammalian cells: step by step. *Crit Rev Biochem Mol Biol*, **45**, 23-49.
94. Hart, M.G., Grant, R., Garside, R., Rogers, G., Somerville, M. and Stein, K. (2008) Chemotherapeutic wafers for High Grade Glioma. *Cochrane Database Syst Rev*, CD007294.
95. Perry, J., Chambers, A., Spithoff, K. and Laperriere, N. (2007) Gliadel wafers in the treatment of malignant glioma: a systematic review. *Curr Oncol*, **14**, 189-194.
96. Buckner, J.C. (2003) Factors influencing survival in high-grade gliomas. *Semin Oncol*, **30**, 10-14.
97. Sathornsumetee, S. and Rich, J.N. (2006) New treatment strategies for malignant gliomas. *Expert Rev Anticancer Ther*, **6**, 1087-1104.
98. Reithmeier, T., Graf, E., Piroth, T., Trippel, M., Pinsker, M.O. and Ninkkrah, G. (2010) BCNU for recurrent glioblastoma multiforme: efficacy, toxicity and prognostic factors. *BMC Cancer*, **10**, 30.
99. Yang, D.I., Yin, J.H., Ju, T.C., Chen, L.S. and Hsu, C.Y. (2004) Nitric oxide and BCNU chemoresistance in C6 glioma cells: role of S-nitrosoglutathione. *Free Radic Biol Med*, **36**, 1317-1328.
100. Jin, G., Cook, S., Cui, B., Chen, W.C., Keir, S.T., Killela, P., Di, C., Payne, C.A., Gregory, S.G., McLendon, R. *et al.* (2010) HDMX regulates p53 activity and confers chemoresistance to 3-bis(2-chloroethyl)-1-nitrosourea. *Neuro Oncol*, **12**, 956-966.
101. Schold, S.C., Jr., Kokkinakis, D.M., Rudy, J.L., Moschel, R.C. and Pegg, A.E. (1996) Treatment of human brain tumor xenografts with O6-benzyl-2'-deoxyguanosine and BCNU. *Cancer Res*, **56**, 2076-2081.
102. Forte, E. and Luftig, M.A. (2009) MDM2-dependent inhibition of p53 is required for Epstein-Barr virus B-cell growth transformation and infected-cell survival. *J Virol*, **83**, 2491-2499.
103. Chuang, L.Y., Yang, C.S., Li, J.C. and Yang, C.H. (2009) Chaotic genetic algorithm for gene selection and classification problems. *OMICS*, **13**, 407-420.
104. Yu, J., Almal, A.A., Dhanasekaran, S.M., Ghosh, D., Worzel, W.P. and Chinnaiyan, A.M. (2007) Feature selection and molecular classification of cancer using genetic programming. *Neoplasia*, **9**, 292-303.
105. Li, G.Z., Bu, H.L., Yang, M.Q., Zeng, X.Q. and Yang, J.Y. (2008) Selecting subsets of newly extracted features from PCA and PLS in microarray data analysis. *BMC Genomics*, **9 Suppl 2**, S24.
106. Ringner, M. (2008) What is principal component analysis? *Nat Biotechnol*, **26**, 303-304.
107. Nguyen, D.V. and Roche, D.M. (2002) Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, **18**, 39-50.
108. Nguyen, D.V. (2005) Partial least squares dimension reduction for microarray gene expression data with a censored response. *Math Biosci*, **193**, 119-137.
109. Mishra, D., Dash, R., Rath, A.K. and Acharya, M. (2011) Feature selection in gene expression data using principal component analysis and rough set theory. *Adv Exp Med Biol*, **696**, 91-100.

110. Zhang, J.G. and Deng, H.W. (2007) Gene selection for classification of microarray data based on the Bayes error. *BMC Bioinformatics*, **8**, 370.
111. Roth, V. and Lange, T. (2004) Bayesian class discovery in microarray datasets. *IEEE Trans Biomed Eng*, **51**, 707-718.
112. Sivaraksa, M. and Lowe, D. (2008) Predictive gene lists for breast cancer prognosis: a topographic visualisation study. *BMC Med Genomics*, **1**, 8.
113. Mutsabayashi, H., Aso, S., Nagashima, T. and Okada, Y. (2008) Accurate and robust gene selection for disease classification using a simple statistic. *Bioinformatics*, **3**, 68-71.
114. Zhu, C.Q., Ding, K., Strumpf, D., Weir, B.A., Meyerson, M., Pennell, N., Thomas, R.K., Naoki, K., Ladd-Acosta, C., Liu, N. *et al.* (2010) Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol*, **28**, 4417-4424.
115. Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O. *et al.* (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science*, **326**, 257-263.
116. Otomo, T., Hishii, M., Arai, H., Sato, K. and Sasai, K. (2004) Microarray analysis of temporal gene responses to ionizing radiation in two glioblastoma cell lines: up-regulation of DNA repair genes. *J Radiat Res (Tokyo)*, **45**, 53-60.
117. Kerr, M.K., Martin, M. and Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *J Comput Biol*, **7**, 819-837.
118. de Haan, J.R., Wehrens, R., Bauerschmidt, S., Piek, E., van Schaik, R.C. and Buydens, L.M. (2007) Interpretation of ANOVA models for microarray data using PCA. *Bioinformatics*, **23**, 184-190.
119. Pavlidis, P. (2003) Using ANOVA for gene selection from microarray studies of the nervous system. *Methods*, **31**, 282-289.
120. Nueda, M.J., Conesa, A., Westerhuis, J.A., Hoefsloot, H.C., Smilde, A.K., Talon, M. and Ferrer, A. (2007) Discovering gene expression patterns in time course microarray experiments by ANOVA-SCA. *Bioinformatics*, **23**, 1792-1800.
121. Wang, X.D., Qi, Y.X. and Jiang, Z.L. (2011) Reconstruction of transcriptional network from microarray data using combined mutual information and network-assisted regression. *IET Syst Biol*, **5**, 95-102.
122. Xu, X., Wang, L. and Ding, D. (2004) Learning module networks from genome-wide location and expression data. *FEBS Lett*, **578**, 297-304.
123. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet*, **34**, 166-176.
124. Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, **95**, 14863-14868.
125. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. and Church, G.M. (1999) Systematic determination of genetic network architecture. *Nat Genet*, **22**, 281-285.
126. Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. and Barkai, N. (2002) Revealing modular organization in the yeast transcriptional network. *Nat Genet*, **31**, 370-377.
127. Barenco, M., Tomescu, D., Brewer, D., Callard, R., Stark, J. and Hubank, M. (2006) Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biol*, **7**, R25.

128. Barenco, M., Brewer, D., Papouli, E., Tomescu, D., Callard, R., Stark, J. and Hubank, M. (2009) Dissection of a complex transcriptional response using genome-wide transcriptional modelling. *Mol Syst Biol*, **5**, 327.
129. Ernst, J., Vainas, O., Harbison, C.T., Simon, I. and Bar-Joseph, Z. (2007) Reconstructing dynamic regulatory maps. *Mol Syst Biol*, **3**, 74.
130. Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M. *et al.* (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet*, **24**, 227-235.
131. Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D. and Brown, P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell*, **11**, 4241-4257.
132. Jelinsky, S.A., Estep, P., Church, G.M. and Samson, L.D. (2000) Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol Cell Biol*, **20**, 8157-8167.
133. Smith, A.D., Sumazin, P. and Zhang, M.Q. (2007) Tissue-specific regulatory elements in mammalian promoters. *Mol Syst Biol*, **3**, 73.
134. Moses, A.M., Chiang, D.Y., Pollard, D.A., Iyer, V.N. and Eisen, M.B. (2004) MONKEY: identifying conserved transcription-factor binding sites in multiple alignments using a binding site-specific evolutionary model. *Genome Biol*, **5**, R98.
135. Wasserman, W.W., Palumbo, M., Thompson, W., Fickett, J.W. and Lawrence, C.E. (2000) Human-mouse genome comparisons to locate regulatory sites. *Nat Genet*, **26**, 225-228.
136. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. *et al.* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, **31**, 374-378.
137. Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*, **34**, D108-110.
138. Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res*, **29**, 281-283.
139. Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I. and Schacherer, F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res*, **28**, 316-319.
140. Bryne, J.C., Valen, E., Tang, M.H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B. and Sandelin, A. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res*, **36**, D102-106.
141. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W. and Sandelin, A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res*, **38**, D105-110.
142. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. and Lenhard, B. (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res*, **32**, D91-94.

143. Vlieghe, D., Sandelin, A., De Bleser, P.J., Vleminckx, K., Wasserman, W.W., van Roy, F. and Lenhard, B. (2006) A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res*, **34**, D95-97.
144. Zambelli, F., Pesole, G. and Pavesi, G. (2009) Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res*, **37**, W247-252.
145. Zhu, Z., Pilpel, Y. and Church, G.M. (2002) Computational identification of transcription factor binding sites via a transcription-factor-centric clustering (TFCC) algorithm. *J Mol Biol*, **318**, 71-81.
146. Shamir, R., Maron-Katz, A., Tanay, A., Linhart, C., Steinfeld, I., Sharan, R., Shiloh, Y. and Elkon, R. (2005) EXPANDER--an integrative program suite for microarray data analysis. *BMC Bioinformatics*, **6**, 232.
147. Sharan, R., Maron-Katz, A. and Shamir, R. (2003) CLICK and EXPANDER: a system for clustering and visualizing gene expression data. *Bioinformatics*, **19**, 1787-1799.
148. Ulitsky, I., Maron-Katz, A., Shavit, S., Sagir, D., Linhart, C., Elkon, R., Tanay, A., Sharan, R., Shiloh, Y. and Shamir, R. (2010) Expander: from expression microarrays to networks and functions. *Nat Protoc*, **5**, 303-322.
149. Elkon, R., Linhart, C., Sharan, R., Shamir, R. and Shiloh, Y. (2003) Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res*, **13**, 773-780.
150. Kaina, B., Christmann, M., Naumann, S. and Roos, W.P. (2007) MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst)*, **6**, 1079-1099.

Chapter 2: A High-Throughput Survival Assay to Measure Drug-induced Cytotoxicity and Cell Cycle Effects

Contributions: Chandni R. Valiathan developed the flow cytometry protocol for suspension cells. Jose L. McFaline optimized parameters for adherent cell lines.

2.1 Abstract

We describe a high-throughput method to accurately measure the cytotoxicity induced in mammalian cells upon exposure to various drugs. Using this assay, we obtain survival data in a fraction of the time required to perform the traditional clonogenic survival assay, long considered the gold standard. The dynamic range of the assay allows sensitivity measurements on a multi-log scale allowing better resolution of comparative sensitivities. Moreover, the results obtained contain additional information on cell cycle effects of the drug treatment. Cell survival is obtained from a quantitative comparison of proliferation between drug-treated and untreated cells. During the assay, cells are treated with a drug and following a recovery period allowed to proliferate in the presence of bromodeoxyuridine (BrdU). Cells that synthesize DNA in the presence of BrdU exhibit quenched Hoechst fluorescence easily detected by flow cytometry; quenching is used to determine relative proliferation in treated versus untreated cells. Finally, this assay can be used in high-throughput format to simultaneously screen multiple cell lines and drugs for accurate measurements of cell survival and cell cycle effects after drug treatment.

ABBREVIATIONS

PI- propidium iodide; CEN - chicken erythrocyte nuclei; BCNU – 1,3-bis(2-chloroethyl)-1-nitrosourea

2.2 Introduction

Survival of cells upon exposure to toxic agents is an important phenotypic measure used to understand the biological importance of certain proteins and pathways in either preventing or enabling cell survival after toxic stress. For example, key proteins involved in DNA repair or the DNA damage response have been identified by measuring the effect of silencing or over-expressing these proteins on cell survival after DNA damage. The gold standard for assessing the survival of cells after drug treatment continues to be the clonogenic survival assay that is extremely sensitive and has a dynamic range of several orders of magnitude. Unfortunately, it suffers from being very low throughput as well as time and labor intensive. Typically, the clonogenic survival assay takes 10-14 days and requires a large number of cell culture plates, thus limiting its practical application to a few cell lines and to a limited number of doses or agents. Moreover, for cells grown in suspension or, those that fail to form colonies, the clonogenic survival assay is done by either monitoring growth from single cells or by following their ability to form colonies in soft agar. These approaches are even more time intensive taking up to 3-4 weeks to complete a single experiment (1). Currently, the only available high-throughput techniques for measuring sensitivity involve the correlation of viability to membrane permeability (trypan blue or propidium iodide exclusion) or measurement of metabolic activity (MTT assay). Unfortunately, membrane permeability only takes into account cells that undergo cell death after treatment and fails to identify sensitivity due to activation of a static program such as arrest or senescence, and metabolic activity primarily reflects mitochondrial function. Furthermore, changes in metabolic activity do not always correlate well with cell viability after treatment and also does not differentiate between cytotoxic and static effects after treatment (2). Moreover, these methods have an inherently limited dynamic range for detection of sensitivity,

generally less than a single order of magnitude versus three or four orders of magnitude for the clonogenic survival assay.

In this paper, we describe a high-throughput method for measuring the sensitivity of cells to a wide variety of agents with a dynamic range comparable to that of the clonogenic survival assay. In addition to obtaining survival information, it can also be used to deduce cell cycle effects of drug treatment. This method is based on the fact that the fluorescence of Hoechst, a dye that preferentially binds AT-rich regions in the DNA, is quenched when bromodeoxyuridine (BrdU), a thymine analog, is incorporated into AT-rich regions (3,4). Cells that have divided zero, one or two times in the presence of BrdU can be differentiated based on the level of quenched Hoechst fluorescence, thus giving a measure of cell proliferation (5). In a previous report, Poot et al. (6) took advantage of the Hoechst quenching property of BrdU to measure survival of cells after exposure to a DNA damaging agent. We have extensively modified the technique to be performed in a multi-well format (96-well plate for suspension cells and 24-well plate for adherent cells), drastically decreasing the setup time and reducing the number of cells required for a survival curve to as little 10^6 - 3×10^6 cells. Furthermore, we have broadened the scope of the assay to simultaneously be used with different cell types and different cytotoxic agents. The assay can be completed within half the time taken to perform a clonogenic survival assay while maintaining high sensitivity and a dynamic range of three to four orders of magnitude.

Figure 2.1 shows a concise representation of the steps involved in the assay. After cells are treated they are allowed to recover for the duration of two doubling times and subsequently allowed to proliferate (if they can) for the duration of another two doubling times in the presence

of BrdU. To make our approach amenable to screening multiple agents with diverse mechanisms of actions we allow cells two doubling times after treatment for drug toxicity to present itself. This allows for simultaneous detection of the toxicity of agents that act immediately versus those that require formation of intermediates to slowly build up in cells. As an example, various DNA damaging agents are dependent on replication for toxicity to occur, and more than one round of replication may be needed to generate an observable phenotypic response. The duration of time prior to BrdU addition can be optimized depending on the agents to be tested. Cells are then gently lysed to obtain nuclei which are stained with propidium iodide (PI) and Hoechst dye. Nuclei fluorescence is measured by flow cytometry to quantify the percentage of cells that have proliferated in the presence of BrdU. The relative proliferation rate of treated samples compared to untreated controls gives a measure of the sensitivity of cells to treatment. The ease, economy and efficiency of this assay will enable rapid progress in systematic approaches to understanding the biological importance of many proteins and pathways whose modulation leads to an observed phenotype after exposure to cytotoxic agents.

2.3 Materials and methods

2.3.1 Cell culture

The lymphoblastoid cell lines TK6 and two TK6 derivatives (MT1 and TK6+MGMT) were grown in suspension in RPMI medium supplemented with 10% equine serum, 1% l-glutamine and 1% penicillin-streptomycin. The adherent U87MG glioblastoma cell line was grown in DMEM medium supplemented with 15% fetal bovine serum, 1% l-glutamine and 1% penicillin-streptomycin.

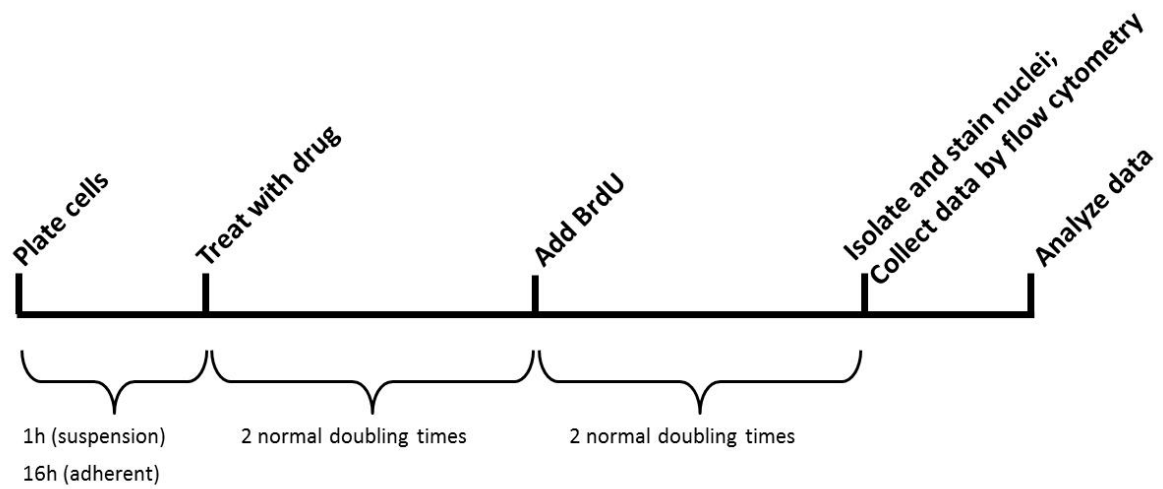


Figure 2.1: Experimental set up with anticipated timing
Time line showing key steps in the experimental procedure

2.3.2 Determining the optimal BrdU concentration

Cells were grown in the presence of different concentrations of BrdU (0-100 μ M) for a little longer than one doubling time, then lysed and stained as described below with Hoechst and propidium iodide for flow cytometry analysis. The optimal BrdU dose was determined as that which quenched Hoechst fluorescence of G1 cells by half after one doubling time; this dose allows the effective resolution of cells that have undergone one division after BrdU addition from those that have undergone none or two divisions after addition. The optimal dose was determined as 45 μ M for TK6 and the TK6 derivatives, and 20 μ M for the U87MG cell lines.

2.3.3 Cell cycle profile analysis by flow cytometry

BrdU is sometimes known to cause a G2/M arrest in cultured human cells. TK6, TK6 derivatives and U87MG cell lines were grown in the presence of the optimal BrdU concentration for at least two doubling times, during which samples were collected at multiple time points, washed with cold PBS and fixed overnight in cold 100% ethanol. Fixed cells were washed with PBS+1% BSA, resuspended in PBS+1%BSA containing propidium iodide (50 μ g/ml) and immediately analyzed by flow cytometry to obtain cell cycle profiles. None of the cell lines showed a G2/M arrest when grown in the presence of BrdU. If however such an arrest is observed, the effect can be overcome by adding deoxycytidine at an equal concentration as the added BrdU.

2.3.4 Drug treatment

Cell lines were treated in duplicate at multiple doses of BCNU (100mM in 100% ethanol). A maximum of six doses were assayed and therefore 12 wells were used for each survival curve. Treatment was performed in serum free media for one hour after which the drug was washed

away and the cells were grown in fresh serum-containing media. This treatment scheme can be accommodated for cytotoxic agents that require longer exposure times as well as short-lived agents that do not require to be washed away after treatment. The drug treatment procedure is described below for suspension or adherent cell lines.

Suspension cells: Before drug treatment, cells were grown to mid-log phase (6×10^5 cells/ml for TK6 and its derivatives). 270 μ l of cells at a density of 4.5×10^5 cells/ml were plated in each well of one row of a round-bottom 96-well plate. If multiple cell lines were assayed, each cell line was plated in one row of a 96-well plate in serum-free media. The drug was diluted to 10X of the final dose concentrations in serum-free media in another 96-well plate. 30 μ l of the 10X drug was transferred to each well in the 96-well plate containing cells, after which the cells were incubated at 37°C for 1h. After 1h, cells were spun down at 1200rpm for 5min, drug-containing media was removed using a multi-channel pipette and the cells were washed with 200 μ l warm 1XPBS per well. The cells were then resuspended in 300 μ l of warm fresh media containing serum, transferred to a flat-bottom 96-well plate and incubated at 37°C for the duration of two normal doubling times. If the drug is not to be washed away, the experiment is set up in a flat-bottom 96-well plate at a density of 1.7×10^5 cells/ml in serum containing media, and the washing step is omitted.

Adherent cells: Before drug treatment, U87MG cells were grown to 80% confluence, washed with warm 1XPBS, trypsinized and diluted to 4×10^4 cells/ml. 1ml of the diluted cells were plated in each well of a 24-well plate and cells were allowed to attach overnight at 37°C, 5% CO₂. One such 24-well plate can be set up for each cell line being assayed. At the time of treatment, the cells were removed from the incubator and the media replaced with 900 μ l of warm, fresh serum-free media. The drug was diluted to 10X the final dose concentrations and 100 μ l of the 10X drug

was added per well such that there were duplicate wells per dose and six doses, including the untreated control. The cells were returned to the incubator for the duration of 1h after which the drug-containing media was replaced with warm serum-containing media. The cells were returned to the incubator for a period of time equivalent to two doubling times for normally growing cells.

2.3.5 Bromodeoxyuridine Addition

After allowing cells to respond for two doubling times after drug treatment, the cells are incubated in the presence of BrdU for another two doubling times. The optimal BrdU concentration was determined to be 45 μ M for TK6 and its derivatives, and 20 μ M for U87MG cells. If the optimal BrdU concentration is greater than 20 μ M, then BrdU must be replenished every 12 hours. Therefore, after the first two doubling times, BrdU was added to TK6 and TK6 derivatives at a concentration of 45 μ M (from a 10mM stock) and replenished by simply adding BrdU to each well every 12 hours for the duration of two normal doubling times. Similarly BrdU was added to U87MG cells at a concentration of 20 μ M. Since cells become photosensitive upon BrdU addition, care must be taken to keep cells in the dark at all times post BrdU addition.

2.3.6 Nuclei isolation and staining for flow cytometry

At the end of four doubling times after drug treatment, cells were transferred to a v-bottom 96-well plate. For suspension cells, the cells were transferred directly to a v-bottom 96-well plate using a multi-channel pipette. For adherent cells, the cell plate was spun down at 1500rpm for 5min to pull down any unattached cells. The media was removed and replaced with 100 μ l of trypsin EDTA. Trypsin was quenched with 200 μ l of serum-containing media and the cells from one well in the 24-well plate were transferred to one well in the v-bottom 96-well plate. Once

transferred, both cell types were spun down at 1500rpm for 5min, media was removed with a multi-channel pipette and cells were washed with cold 1XPBS. The cells were then resuspended in 300 μ l of 1X lysis/staining buffer (0.1M Tris HCl pH 7.5, 0.1% Igepal CA-60, 1mM CaCl₂, 5mM MgCl₂, 0.2%BSA (w/v), 1.2 μ g/ml Hoechst 33258 and 1x10⁴ chicken erythrocyte nuclei (CEN)/ml) and incubated on ice for 15 minutes after which 6 μ l of 200 μ g/ml propidium iodide was added to each well using a multi-channel pipette. Two wells containing only the lysis/staining buffer and propidium iodide are also prepared as blanks. Samples were mixed well using a multi-channel pipette and analyzed on a BD LSR II flow cytometer equipped with a 96-well plate robot arm.

2.3.7 Data collection

Events were visualized on the side scatter vs. forward scatter plot to gate out debris (Figure 2.2a), and on the PI-height vs. PI-area plot to exclude doublets that fall below the diagonal (Figure 2.2b). 30,000 events that passed these two criteria were collected and viewed on a PI-Area vs. Hoechst-Area plot (Figure 2.2c). During data collection, the voltages for PI and Hoechst were adjusted to position CEN at the (30K, 30K) point to facilitate subsequent data analysis.

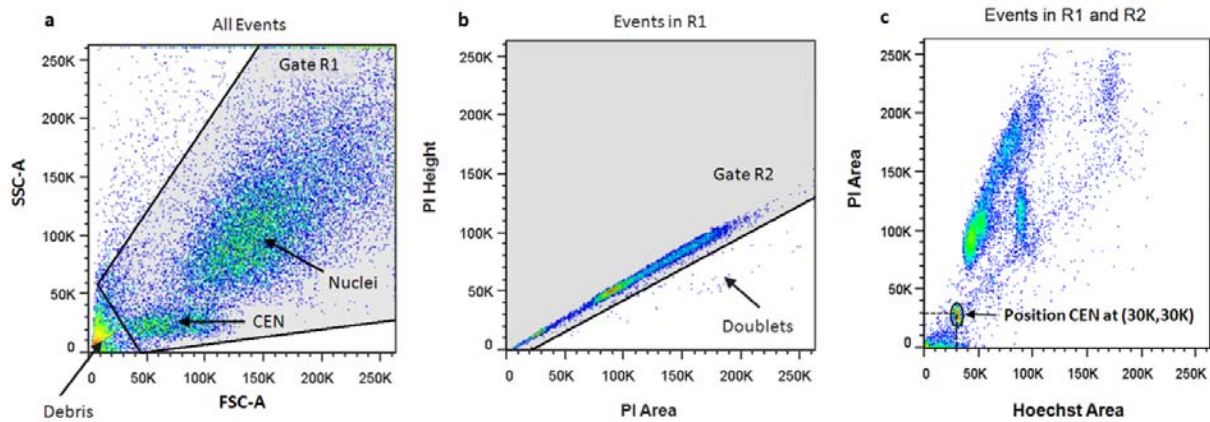


Figure 2.2: Data filtering and collection plots

a) Side scatter vs. forward scatter plot shows the position of nuclei, CEN and debris. All events in gate R1 are to be included in the data collection and analysis steps. b) PI-Height vs. PI-Area plot shows the doublet events below the diagonal that must be excluded. Events in gate R2 are included in the data collection and analysis steps. c) PI-Area vs. Hoechst-Area plot of events included in gates R1 and R2. PI and Hoechst channel voltages must be adjusted to position CEN at (30K, 30K) or any other practical coordinates for all samples.

2.4 Results

2.4.1 Detecting cells that have undergone zero, one or two cell divisions after drug treatment

BCNU is a DNA damaging agent commonly used in the clinic to treat glioblastoma despite the fact that it can have severe side effects on hematopoietic cells (7,8). We used the multi-well assay to test the BCNU sensitivity of the human lymphoblastoid cell lines TK6 and TK6 derivatives (MT1 and TK6+MGMT) that grow in suspension and the human U87MG glioblastoma cell lines that grow attached. All data were analyzed using FlowJo (TreeStar Inc). For each cell line and drug dose, the debris and doublets were gated out as described in Figure 2.2. The remaining events were observed on a PI-Area vs. Hoechst-Area plot. Figure 2.3 delineates the regions corresponding to cells that are in the first, second or third cell cycle after treatment, recovery and incubation with BrdU. As seen in Figure 2.3A, Hoechst fluorescence of cells decreases as they replicate their DNA in the presence of BrdU. Therefore, as the cells replicate and divide, they move from the region labeled 1st cell cycle leftwards to the region labeled 2nd cell cycle, and so on. For each sample, gates were drawn as in Figure 2.3C and the number of events in each gate was determined.

2.4.2 Calculating the fraction of proliferated cells after drug treatment

For each cell type and dose assayed, the number of events in each of the regions corresponding to the 1st, 2nd or 3rd cell cycle was used to calculate the number of proliferating cells in the sample as shown in formula 1.

Formula 1

$$\frac{1}{2} * (\#events \text{ in cell cycle } 2 / \#CEN) + \frac{1}{4} * (\#events \text{ in cell cycle } 3 / \#CEN)$$

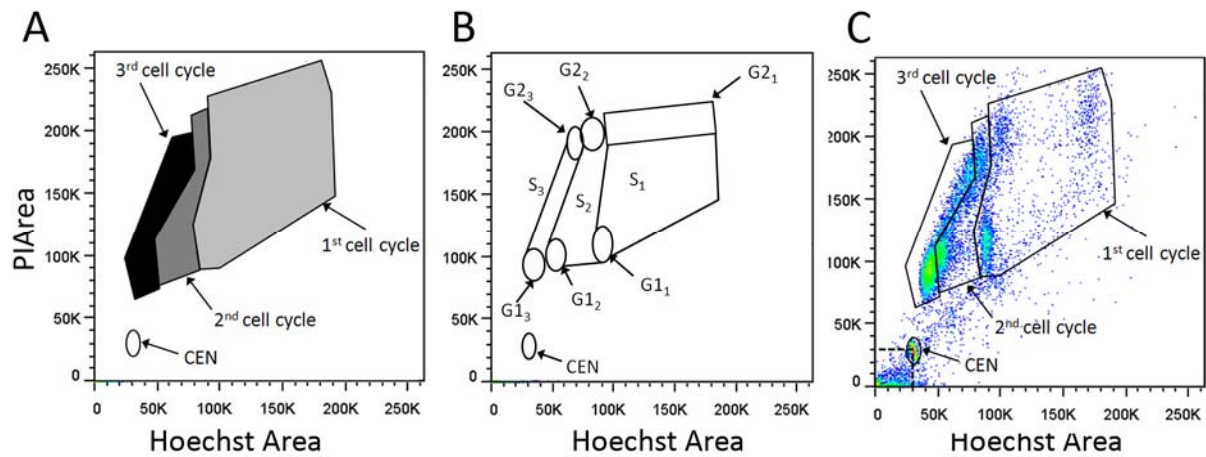


Figure 2.3: Gates delineating cell cycle regions for the data analysis

A) PI-Area vs. Hoechst-Area plot showing the regions that would contain nuclei in the first cell cycle (CC1)- light grey, second cell cycle (CC2)- dark grey, third cell cycle (CC3) black. B) This plot shows the G1, S and G2 populations for each cell cycle. G1₁, S₁, G2₁ are G1, S and G2 in cell cycle 1 and so on. C) Example of a sample for which the CEN and cell cycle gates have been drawn.

The formula calculates the total number of cells that have divided after drug treatment by halving the number of cells in the 2nd cell cycle (these cells have undergone one cell division) and dividing the number of cells in the 3rd cell cycle by four (these cells have divided twice). Since all samples of equal volume were spiked with the same number of CEN, the number of CEN counted per sample is proportional to the volume of sample used for flow cytometry data collection. Therefore the density of proliferating cells is calculated by dividing the total number of proliferating cells per sample by the number of CEN counted in that sample. Comparing the density of proliferated cells in a treated sample with that in an untreated sample gives the % control growth value that is used to plot a survival curve. Example calculations for the survival curves are shown in Table 2.1.

2.4.3 The multi-well assay has a large dynamic range, yielding log scale killing for suspension and adherent cell lines treated with a cytotoxic agent

BCNU's efficacy as a chemotherapeutic agent arises from its ability to generate extremely cytotoxic DNA inter-strand crosslinks (9). DNA crosslinks are formed in a multistep process, the first of which is the formation of O^6 -chloroethylguanine lesions (10). The O^6 -methylguanine methyl transferase (MGMT) protein is known to remove chloroethyl adducts from the O^6 position of guanine (11,12). Thus, MGMT provides protection against BCNU-induced cytotoxicity, and cells lacking MGMT are particularly sensitive to BCNU (13-15). With this in mind, we used our multi-well assay to measure BCNU sensitivity of cell lines either lacking or expressing MGMT to determine the range of sensitivity accurately measured by the assay.

Table 2.1 Example showing the process of calculating % control growth of treated samples.

Sample-dose(μM)	CEN	CC2	CC3	$A=CC2/CEN$	$B=CC3/CEN$	$N_d=A/2+B/4$	$P=100*(N_d/N_{d=0})$
MT1-0	2852	3461	19049	1.21353436	6.67917251	2.27656031	100
MT1-25	11075	602	74	0.05435666	0.00668172	0.02884876	1.26720818
MT1-50	8543	24	0	0.00280932	0	0.00140466	0.06170093
TK6+MGMT-0	2128	10463	8881	4.91682331	4.17340226	3.50176222	100
TK6+MGMT-25	2106	8950	8183	4.24976258	3.88556505	3.09627255	88.4204113
TK6+MGMT-50	2631	12822	8030	4.87343216	3.05207146	3.19973394	91.3749633

Data for survival curves shown in Figure 2.4. CEN: Chicken Erythrocyte Nuclei; CC2: number of events in cell cycle 2; CC3: number of events in cell cycle 3; N_d : number of cells that divided; P: percentage of proliferating cells in a treated versus untreated sample.

BCNU sensitivity of suspension (TK6 and MT1) and adherent (U87MG) cell lines that lack MGMT was measured using our multi-well assay. The TK6 and MT1 cell lines have previously been shown to be extremely sensitive to BCNU in the clonogenic survival assay as shown in Figure 2.4A (16). Results from our multi-well assay, shown in Figure 2.4B, are remarkably similar to those from the colony-forming assay. Additionally, we measured BCNU sensitivity of the TK6+MGMT cell line (TK6 cell line reconstituted with MGMT) using our multi-well assay and found that TK6+MGMT cells show extreme resistance to BCNU, as expected (Figure 2.4B). We also tested the assay on adherent cells by measuring BCNU sensitivity of the adherent U87MG glioblastoma cell line that lacks MGMT (17). Again, as expected the U87MG cells showed extreme BCNU sensitivity (Figure 2.4), also with multi-log-scale killing.

2.4.4 The multi-well assay detects cell-cycle effects of drug treatment

Flow cytometry plots obtained using our assay for U87MG cells treated with BCNU show that with increasing doses of BCNU, there is not only a decrease in the total number of cells but that this decrease is also accompanied by a steady increase in the fraction of cells in late S and G2/M phase of the first cell cycle (Figure 2.5). From these observations we can conclude that at higher BCNU doses, surviving cells are unable to divide during the BrdU pulse and remain in the region corresponding to the first cell cycle. Moreover, surviving cells are arrested at late S or G2/M for the entire duration after BrdU addition.

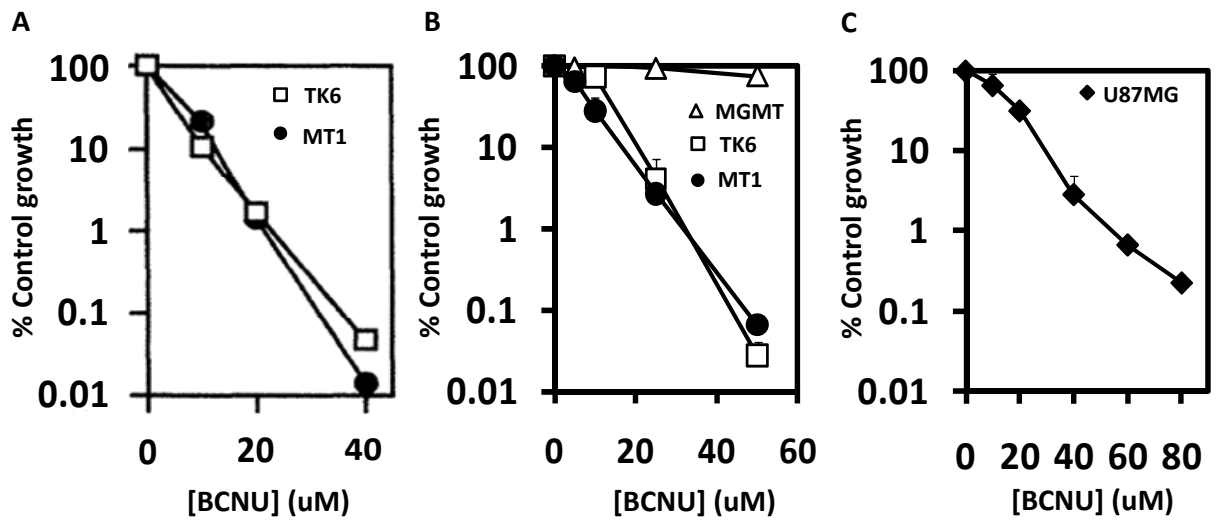


Figure 2.4: Example of data obtained using the assay

(A) Killing curves for BCNU treatment of TK6 and MT1 cell lines using the traditional clonogenic survival assay (reproduced from (13)); (B) Killing curves for BCNU treatment of the TK6, MT1 and TK6+MGMT cell lines using the multi-well assay; (C) Killing curves for BCNU treatment of the U87MG glioblastoma cell line using the multi-well assay.

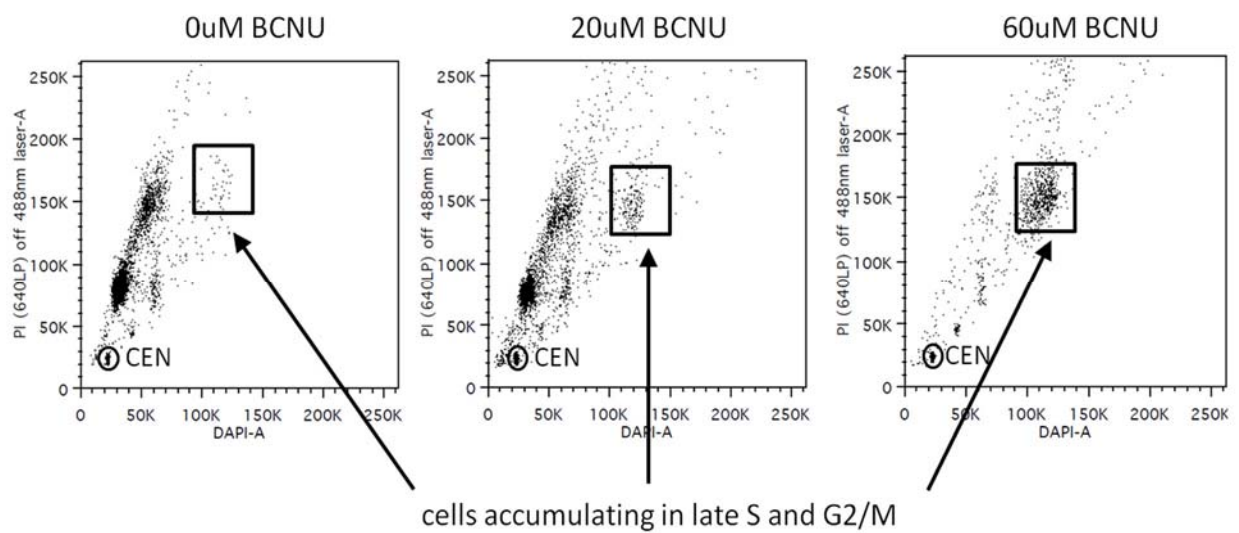


Figure 2.5: Cell cycle effects of BCNU on U87MG cells

U87MG cells show accumulation in late S or G2/M with increasing doses of BCNU. This information is obtained from the survival data with no additional experiments.

2.5 Discussion

The use of survival measurements is a key method implemented to determine the effect that a gene product has on the phenotypic outcome of cell death. The bottleneck in these studies is usually the method used to measure cell survival. The clonogenic survival assay, long considered the gold standard in the field, takes 10-14 days and can only be performed for a few cell lines and few doses at any one time due to the long and laborious setup procedure. Moreover, the assay becomes troublesome for cells that do not form colonies or that grow in suspension, where cells must be plated at single cell densities or in soft agar to count colonies (1). To get around this time-consuming assay, people have resorted to other higher-throughput techniques such as measuring membrane permeability or metabolic (mitochondrial) activity; while these assays are quicker, they have limited dynamic range such that cells that with drastically different sensitivities on a log scale may be considered as only moderately different on the linear scale.

The multi-well assay we describe here produces survival measurements on a multi-log scale comparable to the clonogenic survival assay. As described in the results section, BCNU sensitivities of the suspension cell lines TK6 and MT1 are remarkably similar to that measured by the clonogenic survival assay. Moreover, the TK6+MGMT cell line is identified as being resistant to BCNU treatment as expected (13,15). These results highlight three important aspects of our assay: i) The assay can measure sensitivity and resistance equally well, thus yielding accurate results that reflect those obtained from the clonogenic survival assay; ii) the multi-well assay is capable of measuring cell sensitivity on a multi-log scale, thus showing the large dynamic range the assay has in contrast to other available quick assays; iii) the assay can be used to measure cell survival of both suspension and adherent cell lines to cytotoxic agents, thus

increasing its applicability. The assay has a reduced setup time and requires fewer cells allowing one to simultaneously assay a large number of cell lines, doses and agents. This greatly improves cell survival measurement efficiency to yield results in a fraction of the time and in a less labor-intensive manner as compared to the clonogenic survival assay.

In addition to inducing cell death, many cytotoxic agents affect regular cell cycle progression, with cells undergoing arrest either in the G1, S or G2/M phases of the cell cycle. The nature of the arrest changes with the drug or cell line used, and provides insight into mechanisms of drug action and possible ways to modify cell sensitivity to a particular drug. Traditional colony-forming survival assays as well as more recent high-throughput survival assays yield no information on possible cell cycle arrest from the drug treatment. Any such cell cycle effects have been determined by separate and additional experiments such as cell cycle profile measurements. In comparison, results from our multi-well assay are extremely rich in cell-cycle information, and therefore provide added insight into the long-term cell cycle effects of the drug treatment.

The data obtained using our assay showed, as seen in Figure 2.5, that the U87MG glioblastoma cell lines show a decrease in total number of cells, accompanied by a concomitant arrest in late S/G2 phase of the cell cycle. Previous studies have shown that U87MG cells show an accumulation of cells in late S or G2/M after BCNU treatment (18) (assayed by PI staining and cell cycle profile analysis by flow-cytometry). However, it is important to note that after determining cell sensitivity to BCNU using the colony forming assay, additional experiments were needed to determine the cell cycle effect of BCNU in U87MG cells. In contrast, our assay

yields the same observations in addition to the survival measurements with no additional time and effort.

In conclusion, we have presented a high throughput method that takes advantage of current flow cytometry technology and properties of proliferating cells to measure the sensitivity of both suspension and adherent cells to different cytotoxic agents. The flexibility of the method, its large dynamic range and its broad applicability makes it a powerful tool with great potential in many different applications including both small and large scale screening of sensitivity of numerous cell lines to numerous toxic agents.

2.6 References

1. Liao, W.T., Jiang, D., Yuan, J., Cui, Y.M., Shi, X.W., Chen, C.M., Bian, X.W., Deng, Y.J. and Ding, Y. (2011) HOXB7 as prognostic factor and mediator of colorectal cancer progression. *Clin Cancer Res*.
2. Nikkhah, G., Tonn, J.C., Hoffmann, O., Kraemer, H.P., Darling, J.L., Schachenmayr, W. and Schonmayr, R. (1992) The MTT assay for chemosensitivity testing of human tumors of the central nervous system. Part II: Evaluation of patient- and drug-specific variables. *J Neurooncol*, **13**, 13-24.
3. Rabinovitch, P.S., Kubbies, M., Chen, Y.C., Schindler, D. and Hoehn, H. (1988) BrdU-Hoechst flow cytometry: a unique tool for quantitative cell cycle analysis. *Exp Cell Res*, **174**, 309-318.
4. Poot, M., Hoehn, H., Kubbies, M., Grossmann, A., Chen, Y. and Rabinovitch, P.S. (1994) Cell-cycle analysis using continuous bromodeoxyuridine labeling and Hoechst 33358-ethidium bromide bivariate flow cytometry. *Methods Cell Biol*, **41**, 327-340.
5. Ormerod, M.G. and Kubbies, M. (1992) Cell cycle analysis of asynchronous cell populations by flow cytometry using bromodeoxyuridine label and Hoechst-propidium iodide stain. *Cytometry*, **13**, 678-685.
6. Poot, M., Silber, J.R. and Rabinovitch, P.S. (2002) A novel flow cytometric technique for drug cytotoxicity gives results comparable to colony-forming assays. *Cytometry*, **48**, 1-5.
7. Durando, X., Lemaire, J.J., Tortochaux, J., Van-Praagh, I., Kwiatkowski, F., Vincent, C., Bailly, C., Verrelle, P., Irthum, B., Chazal, J. *et al.* (2003) High-dose BCNU followed by autologous hematopoietic stem cell transplantation in supratentorial high-grade malignant gliomas: a retrospective analysis of 114 patients. *Bone Marrow Transplant*, **31**, 559-564.
8. Hochberg, F.H., Parker, L.M., Takvorian, T., Canellos, G.P. and Zervas, N.T. (1981) High-dose BCNU with autologous bone marrow rescue for recurrent glioblastoma multiforme. *J Neurosurg*, **54**, 455-460.
9. Gonzaga, P.E., Potter, P.M., Niu, T.Q., Yu, D., Ludlum, D.B., Rafferty, J.A., Margison, G.P. and Brent, T.P. (1992) Identification of the cross-link between human O6-methylguanine-DNA methyltransferase and chloroethylnitrosourea-treated DNA. *Cancer Res*, **52**, 6052-6058.
10. Tong, W.P., Kirk, M.C. and Ludlum, D.B. (1982) Formation of the cross-link 1-[N3-deoxycytidyl],2-[N1-deoxyguanosinyl]ethane in DNA treated with N,N'-bis(2-chloroethyl)-N-nitrosourea. *Cancer Res*, **42**, 3102-3105.
11. Kaina, B., Fritz, G., Mitra, S. and Coquerelle, T. (1991) Transfection and expression of human O6-methylguanine-DNA methyltransferase (MGMT) cDNA in Chinese hamster cells: the role of MGMT in protection against the genotoxic effects of alkylating agents. *Carcinogenesis*, **12**, 1857-1867.
12. Gerson, S.L. (2004) MGMT: its role in cancer aetiology and cancer therapeutics. *Nat Rev Cancer*, **4**, 296-307.
13. Maze, R., Carney, J.P., Kelley, M.R., Glassner, B.J., Williams, D.A. and Samson, L. (1996) Increasing DNA repair methyltransferase levels via bone marrow stem cell transduction rescues mice from the toxic effects of 1,3-bis(2-chloroethyl)-1-nitrosourea, a chemotherapeutic alkylating agent. *Proc Natl Acad Sci U S A*, **93**, 206-210.
14. Meikrantz, W., Bergom, M.A., Memisoglu, A. and Samson, L. (1998) O6-alkylguanine DNA lesions trigger apoptosis. *Carcinogenesis*, **19**, 369-372.

15. Wu, M., Kelley, M.R., Hansen, W.K. and Martin, W.J., 2nd. (2001) Reduction of BCNU toxicity to lung cells by high-level expression of O(6)-methylguanine-DNA methyltransferase. *Am J Physiol Lung Cell Mol Physiol*, **280**, L755-761.
16. Hickman, M.J. and Samson, L.D. (1999) Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proc Natl Acad Sci U S A*, **96**, 10764-10769.
17. Roos, W.P., Batista, L.F., Naumann, S.C., Wick, W., Weller, M., Menck, C.F. and Kaina, B. (2007) Apoptosis in malignant glioma cells triggered by the temozolomide-induced DNA lesion O6-methylguanine. *Oncogene*, **26**, 186-197.
18. Xu, G.W., Mymryk, J.S. and Cairncross, J.G. (2005) Pharmaceutical-mediated inactivation of p53 sensitizes U87MG glioma cells to BCNU and temozolomide. *Int J Cancer*, **116**, 187-192.

Chapter 3: BCNU induced transcriptional changes identify NF-Y as a putative regulator of mitotic entry

Contributions: Chandni Valiathan designed and performed the experiments and bioinformatics analysis. Prajit Limsirichai performed time series measurements of Annexin V/7AAD and performed immunoblot analysis for total, phosphorylated and acetylated p53.

3.1 Abstract

In the presence of DNA lesions generated by exogenous and endogenous agents, transcriptional changes are extremely important for the proper control of cell cycle progression and cell death/survival decisions. To better understand temporal kinetics and control of transcription after exposure to a chemotherapeutic agent BCNU, differences in expression kinetics between BCNU sensitive and BCNU resistant cell lines were studied. This led to the identification of a transcriptional signature that correlated with the observed cellular phenotype in BCNU sensitive and resistant cell lines. Furthermore, NF-Y was identified as a putative regulator of mitotic genes that were down-regulated in BCNU sensitive but not resistant cell lines, and thus implicated in the observed stalling of entry into mitosis.

3.2 Introduction

Constant damage to the DNA in our cells from exogenous and endogenous agents elicits DNA damage response (DDR) mechanisms that include the induction of DNA repair, initiation of cell cycle checkpoints, chromatin remodeling and activation of transcription. The prominent role of the transcriptional response to DNA damage is seen by the significant genome-wide transcriptional changes induced upon exposure to DNA damaging agents in yeast and mammalian cells, not only for genes involved in DNA damage related functions like DNA repair, cell cycle arrest and apoptosis, but also in other cellular processes such as protein degradation and metabolism (1-4). In addition to the increased incidence of mutated transcription factors in cancer (5), the increased sensitivity to DNA damaging agents of yeast strains that were silenced in genes involved in transcription regulation (6,7) shows the importance of transcriptional control in rescuing a cell from DNA damage.

Although cell-wide responses have been observed transcriptionally and are known to be important for cell decision processes after DNA damage, the mechanisms by which these decisions are made is not understood well. Moreover, the transcriptional control of these wide spread gene expression changes is still an area of active research.

In this study, we set out to identify a transcriptional signature that was differentially induced or repressed over time in cells with extreme sensitivity or resistance to the chemotherapeutic agent BCNU. We went a step further to confirm that the signature gene set we identified indeed correlated with the cellular phenotype. We also attempted to identify a mechanism for the transcriptional regulation of the signature set in the presence of DNA damage.

3.3 Materials and Methods

3.3.1 Cell culture

Lymphoblastoid suspension cell lines were grown in RPMI medium (Invitrogen) supplemented with 15% FBS, supplemented with penicillin-streptomycin and L-glutamine. All cell lines were in the mid-log phase of growth prior to treatment.

3.3.2 Proliferation assay

Survival curves for the panel of 24 cell lines were obtained using the proliferation assay described in Chapter 2. Briefly, cells were plated in a 96-well plate and treated with 0, 10, 20, 40, 60 and 80 μ M BCNU for one hour in serum-free media. After this, the drug was washed away and the cells were resuspended in warm serum-containing media. The cells were allowed to recover for two normal doubling times and then grown in the presence of BrdU (Sigma-Aldrich B5002) for another two doubling times. At the end of this, the cells were lysed in lysis buffer (0.1M Tris HCl pH 7.5, 0.1% Igepal CA-60, 1mM CaCl₂, 5mM MgCl₂, 0.2%BSA (w/v), 1.2 μ g/ml Hoechst 33258, 1x10⁴ chicken erythrocyte nuclei/ml), nuclei were stained with 100 μ g/ml propidium iodide (Sigma-Aldrich P4170) and data was collected by flow cytometry.

3.3.3 Cell cycle profile analysis

Cells were spun down at 1500 rpm for 5min, washed with cold 1XPBS and fixed in 100% cold ethanol while vortexing. Cells were fixed overnight at 4⁰C, after which they were washed twice with 1XPBS+1%BSA and stained in 1XPBS containing 50 μ g/ml propidium iodide and 1mg/ml RNaseA (Invitrogen). Samples were protected from light and analyzed by flow cytometry.

3.3.4 Drug treatment

Cells were grown to mid log-phase prior to drug treatment. On the day of treatment, cells were spun down at 1500rpm for 5min and diluted to 4.5×10^5 cells/ml in warm serum free RPMI medium. BCNU (100mM stock in 100% ethanol; Sigma-Aldrich C0400) was added directly to cells at a concentration of 40 μ M. Mock-treated samples were set up in the same manner as treated samples, except that 100% ethanol was added to cells instead of BCNU. After an hour-long BCNU exposure, the drug was washed away and cells were re-suspended in fresh RPMI medium supplemented with 15% FBS. Cells were collected at the each time-point following the appropriate protocol for each assay.

3.3.5 RNA isolation and hybridization

Cell lines 4, 5, 13 and 16 were treated with 40 μ M BCNU as described. At each time point, two million cells were collected on ice and lysed in TRIzol[®] Reagent (Invitrogen). Samples were stored at -80⁰C until all samples for the time-course were collected. RNA was isolated using the Qiagen RNeasy Mini kit and checked for integrity on the Agilent BioAnalyzer 2100. All samples that passed quality control were prepared using the NuGEN sample preparation procedure, and hybridized on Affymetrix HG-U133 plus 2.0 arrays. RNA quality control, sample preparation and array hybridization were performed at the BiomicroCenter, MIT.

3.3.6 Microarray data analysis

Normalization: Affymetrix .CEL files were uploaded to the GenePattern suite (8) and standard RMA normalization was used to normalize and extract expression values from the data set. Only probe-sets that were present in at least one sample were included in the analysis. The array data

was formatted into a three dimensional matrix in MATLAB (Mathworks). Similarly three dimensional matrices were used to store relevant gene information such as the time-point, drug dose and whether the expression value was measured in a sensitive or resistant cell line.

Gene-set selection: ANOVA analysis was performed using the *anovan* function in MATLAB using the three dimensional matrices for gene expression values, BCNU dose, time-point and sensitivity. P-values for each of the variables (treatment, time and sensitivity) and for the different combinations of the variables (e.g treatment X time) were obtained. Probe-sets that passed a p-value cutoff of $p < 0.01$ for the treatment X sensitivity or treatment X sensitivity X time variables were selected for further analysis.

Clustering: The hierarchical clustering module from the GenePattern suite was used to cluster along the rows of log-normalized data (\log_2 (gene value at time t/gene value at time 0)). Clustered data were exported and visualized using MATLAB. Similarly, the transcriptional signature was clustered using CLICK within the EXPANDER platform (9-12) to identify clusters of up-regulated and down-regulated genes. These clusters were used for the Ingenuity Pathway Analyses, Gene Ontology Enrichment and transcription factor binding site enrichment.

Ingenuity Pathway Analysis: The set of 984 probe-sets as well as the up-regulated and down-regulated clusters were analyzed using Ingenuity Pathways Analysis (Ingenuity® Systems, www.ingenuity.com) for network connectivity and canonical pathway enrichment analysis. Eligibility settings were chosen to be stringent, only allowing interactions and findings seen in human cell lines.

Gene Ontology Enrichment: Gene ontology enrichment for biological processes was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6.7; (13)). The standard GO-FAT enrichment within DAVID was used, which focuses on enrichment of

more specific GO terms within the provided gene set by filtering out broader terms that might over-shadow the lower-level terms.

Gene Set Enrichment Analysis (GSEA): The GSEA algorithm (14) available from the GenePattern Suite (Broad Institute) was used to check enrichment of genes that are repressed by p53 acetylation within the set of 984 differentially expressed probe-sets. Gene expression values at 48hours were used as the input and gene-wise permutations were performed for the enrichment analysis.

3.3.7 Annexin/7AAD assay for measuring cell death

At each time point, cells were collected (1500rpm, 5min) and washed once in cold Annexin buffer (10mM HEPES, 140mM NaCl, 2.5mM CaCl₂). Cell pellets were then incubated with 7AAD (50µg/ml; Sigma A9400) and PE-AnnexinV (1:20 dilution; Invitrogen A35111) for 15 minutes at room temperature, protected from light. Cells were immediately analyzed by flow-cytometry.

3.3.8 Phospho-histone H3 for measuring the mitotic fraction

At each time point, cells from mock-treated and treated samples were collected (1500rpm, 5minutes) and washed once with cold 1xPBS. Cells were then fixed in 4% formaldehyde (prepared in 1XPBS), at room temperature for 15 minutes. Cells were spun down and washed once with 1XPBS, after which cell pellets were resuspended in 100% methanol and stored at -20°C until analysis by flow cytometry. To prepare cells for flow cytometry, samples were spun down to remove methanol, and washed once in 1XPBS + 1% BSA. Cells were permeabilized by incubation in 1XPBS + 0.25% Triton for 15 minutes at room temperature. Cells were washed

once with 1XPBS + 1%BSA, and incubated with 0.75 μ g of anti-phospho-Histone H3 antibody (Upstate #06-570) for three hours at room temperature. After this incubation, cells were washed twice with 1XPBS + 1%BSA, and incubated with goat anti-rabbit IgG Alexa Fluor 647 (1:30 dilution in 1xPBS + 1%BSA; Molecular Probes A21244) for 30 minutes at room temperature and protected from light. Samples were washed twice with 1XPBS and resuspended in 1XPBS containing 50 μ g/ml propidium iodide and 1mg/ml RNaseA. Cells were analyzed by flow cytometry.

3.3.9 Immunoblot analysis

At each time point, mock-treated and treated samples were collected (1500rpm, 5minutes), washed with 1XPBS and frozen in liquid nitrogen. Cell pellets were stored at -80⁰C until lysis. Cells were lysed for 30 minutes on ice in lysis buffer (10mM Tris-HCl, pH8.0; 137mM NaCl; 10% glycerol; 1% NP-40; 10mM EDTA; protease inhibitor cocktail; 10mM NaF; 1mM DTT; 1mM sodium orthovanadate). The lysate was sonicated (3 times, 2 seconds each at 20% amplitude) to dissociate chromatin-bound proteins. Debris was pelleted by centrifuging samples for 10 minutes at maximum speed at 4⁰C. Whole cell lysates were size-separated on 4-12% Bis-Tris polyacrylamide gels (Invitrogen) and transferred onto nitrocellulose membrane (BioRad) for immunoblot analysis. The membrane was probed with the following primary antibodies: total p53 (Santa Cruz, sc-263), p-p53ser20 (R&D systems, AF2286), p-p53ser15 (R&D systems, AF1043), acetyl-p53lys373 (Millipore 06-916), beta-actin (Sigma A5441), and vinculin (Sigma V9131). Appropriate IRDye-conjugated secondary antibodies (Rockland) were used and the immunoblots were scanned and quantified using the Odyssey system (Licor Biosciences).

Positive controls were prepared from irradiated cells (10Gy) and cells treated with 50 μ M etoposide (Sigma E1383) and 16.5nM TSA (Sigma T1953) for 12 hours.

3.3.10 Chromatin immunoprecipitation followed by sequencing and peak calling

Chromatin immunoprecipitation was performed using the Agilent mammalian protocol. Briefly, cells were cross-linked in 10% formaldehyde for 10 minutes at room temperature after which Glycine was added to quench cross-linking activity. Cells were washed with 1XPBS, aliquoted into tubes (10^8 cells per aliquote) and stored at -80°C . Prior to immunoprecipitation, cells were permeabilized in lysis buffer (50mM Hepes-KOH, pH7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X and protease inhibitors) for 10 minutes at 4°C , nuclei were washed and resuspended in sonication buffer (50mM Tris-HCl pH 8, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate). Nuclei were sonicated to obtain chromatin fragments $<500\text{bp}$ (36 cycles of 20 seconds on and 1 minute off at 45% amplitude). 50 μ l of the sonicated chromatin was saved as input control and the rest was incubated overnight with IgG (Sigma) or anti-NF-YA antibody (Rockland, 200-401-100) bound ProteinA-Dynal beads (Invitrogen). Beads were washed thrice with sonication buffer and one last time with a LiCl wash (20mM Tris pH 8, 1mM EDTA, 250mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and eluted in elution buffer(50mM Tris-HCl, pH 8, 10mM EDTA, 1% SDS) at 65°C with vortexing. Crosslinks were reversed at 65°C for 6 hours, and samples were treated with RNase and proteinase K. Following this, DNA was isolated by phenol-chloroform extraction.

Sample preparation, sequencing and preliminary data extraction was performed at the Biomicro Center, MIT on the Illumina sequencing platform. Aligned sequences from two NF-Y IPs and

two IgG control IPs were combined for Model-based Analysis of ChIP-Seq (MACS) (15) using a p-value cut-off of 10^{-10} . The 200 most significant peaks were analyzed using THEME (16) to identify the enriched motif within this set.

3.4 Results

3.4.1 Determining the optimal BrdU dose for lymphoblastoid cells

The first step in using the proliferation assay (described in Chapter 2) to obtain the dose-response curve of a cell line to a particular DNA damaging agent was to establish the optimal BrdU concentration for the cell lines being tested. This optimal concentration is defined as the concentration of BrdU that quenches Hoechst fluorescence of cells in the G1 phase of the cell cycle by one half after cells complete one doubling in the presence of BrdU (17-19). Half-quenching of Hoechst fluorescence in the first cell cycle is necessary for optimal resolution of cells that have divided once, twice or thrice in the presence of BrdU.

The optimization was performed in the lymphoblastoid cell line TK6 with the assumption that all lymphoblastoid cell lines would require similar BrdU concentrations to quench G1 Hoechst fluorescence by half. A range of BrdU doses between 0-100 μM was tested by growing TK6 cells in the presence of BrdU for a little more than one doubling time (16 hours). Figure 3.1 shows the flow cytometry plots obtained for the four BrdU concentrations tested. As is seen in the figure, 45 μM BrdU was enough to quench Hoechst fluorescence by almost half (from ~90K to ~50K) and any further increase in the BrdU concentration did not significantly increase quenching. Moreover, doses of BrdU lower than 45 μM did not yield enough Hoechst quenching. Therefore, 45 μM BrdU was defined as the optimal concentration for the panel of 24 cell lines.

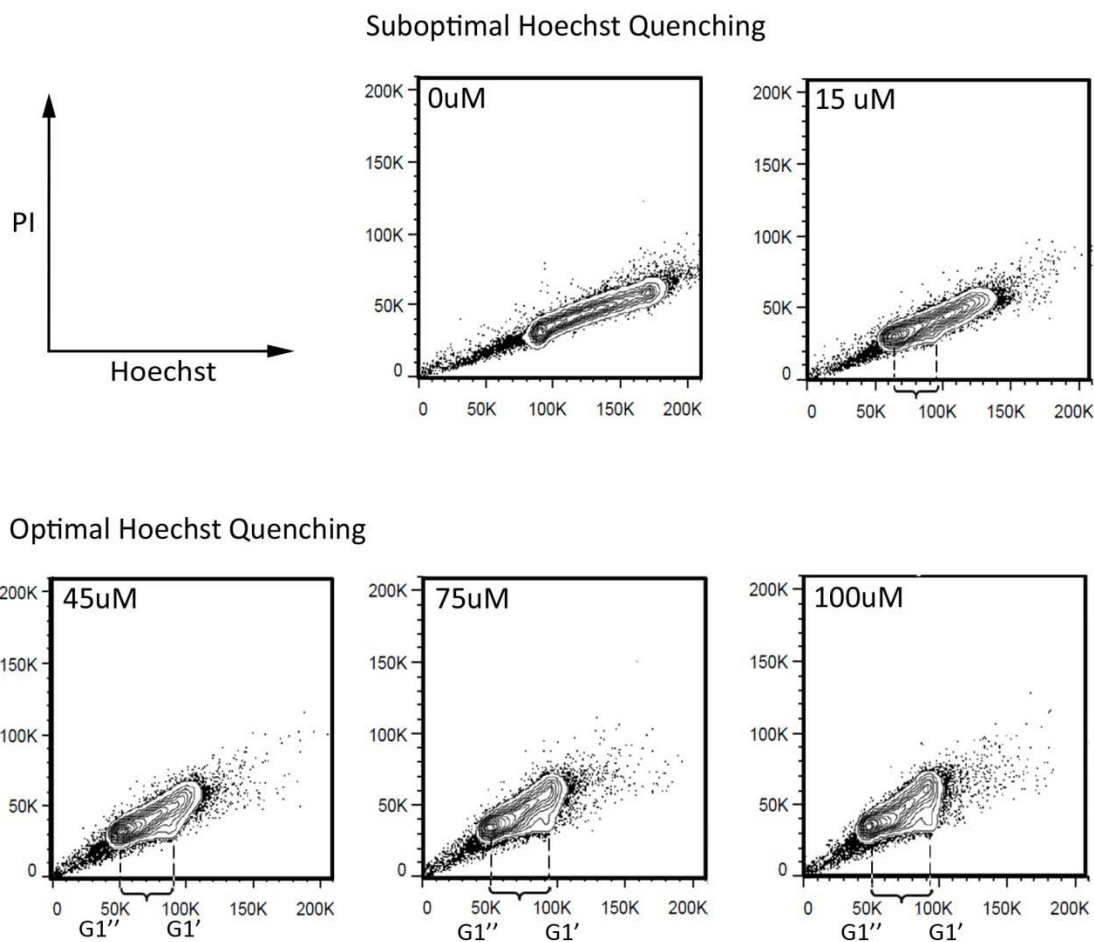


Figure 3.1: Determining the optimal BrdU concentration for lymphoblastoid cell lines

Flow cytometry scatter plots of PI fluorescence vs. hoechst fluorescence showing hoechst quenching for various doses of BrdU. The difference in hoechst fluorescence between G1 cells of the first (G1') and second cell cycle (G1'') is marked along the x-axis.

3.4.2 Cell lines show no G2/M arrest after BrdU addition

Although most diploid cells show no cell cycle effects from short term BrdU incorporation (20,21), in some cell types, BrdU is known to induce a G2/M arrest (22). Therefore, a fraction of the twenty-four cell lines were tested to see whether lymphoblastoid cell lines undergo a G2/M arrest when grown in 45 μ M BrdU. For this experiment, cells were grown in the presence of BrdU for 48 hours. At various time intervals during the BrdU incubation, cells were collected, fixed in 100% ethanol, stained with propidium iodide (as described in the Materials and Methods section) and analyzed by flow-cytometry to obtain cell cycle profiles. Figure 3.2 to Figure 3.7 show the cell cycle profiles, fraction of cells in G1 and G2 over time, and the growth of cells in the presence of BrdU for the cell lines tested. These show that no discernible G2/M arrest or growth retardation is seen in these lymphoblastoid cell lines grown for a 48 hours in the presence of BrdU.

3.4.3 The genetically varied panel of cell lines show a wide range of sensitivities to BCNU

With the optimal BrdU concentration defined as 45 μ M and the knowledge that this dose did not induce a G2/M arrest in the lymphoblastoid cell lines, the panel of cell lines were ready to be screened for BCNU sensitivity. It is important to note that the multi-well proliferation assay used to ascertain BCNU sensitivity is dependent on the doubling times of the cell lines being assayed. Based on previously determined doubling times (Samson lab, unpublished data), the panel of cell lines was divided into four sub-groups. The four sub-groups are shown in Table 3.1, accompanied by the approximate doubling times defined for each group.

Cell line 6

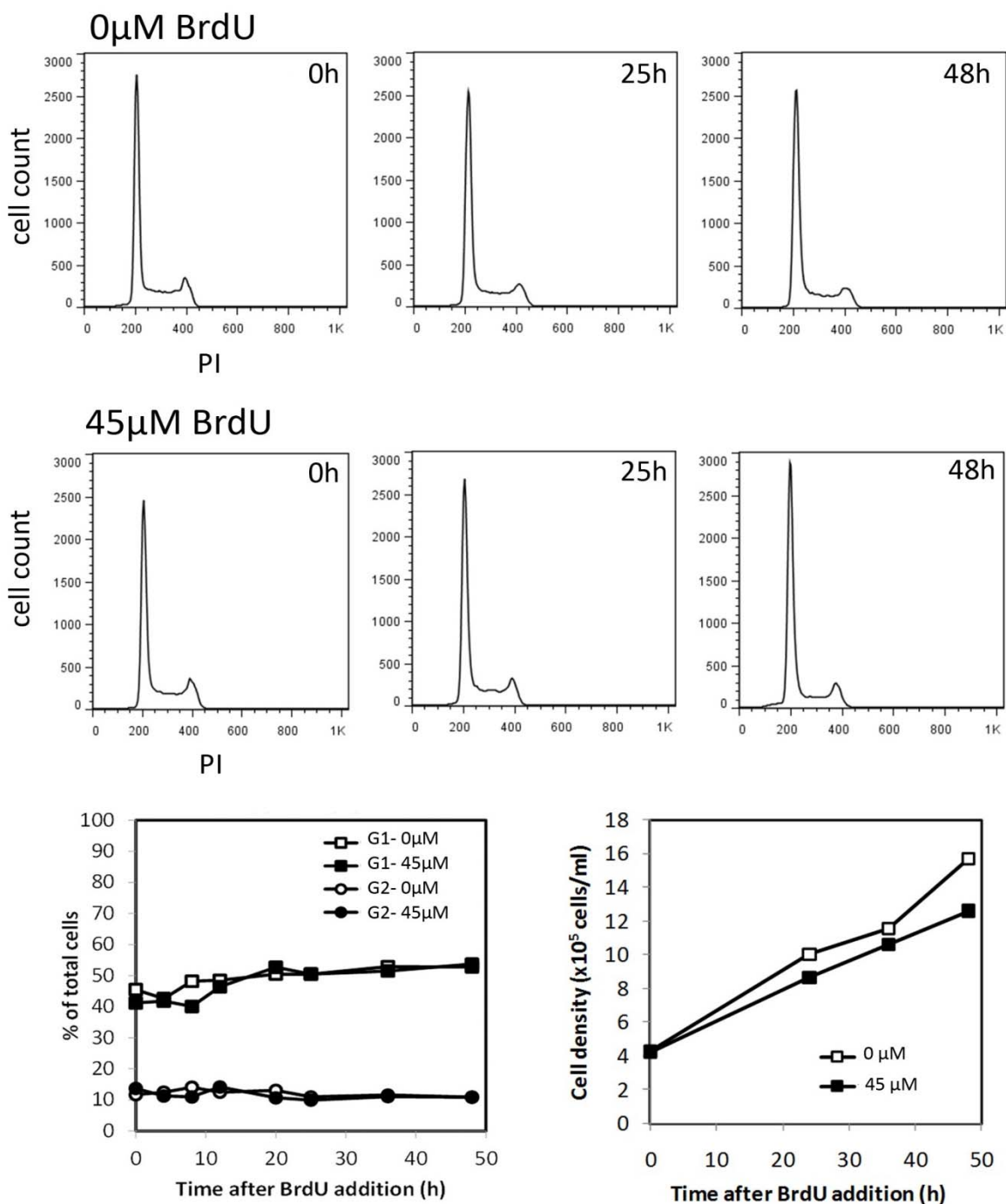


Figure 3.2: Checking for BrdU dependant G2/M arrest in Cell line 6

Cell cycle profiles for cells grown in the absence (top panel) or presence (middle panel) of 45 μ M BrdU for 0, 25 and 48 hours. The bottom, left panel shows fraction of cells in G1 (squares) and G2 (circles) for cells grown with 0 (open) or 45 μ M (filled) BrdU. The bottom, right panel shows cell growth for cells grown in 0 (open) or 45 μ M (filled) BrdU.

Cell line 14

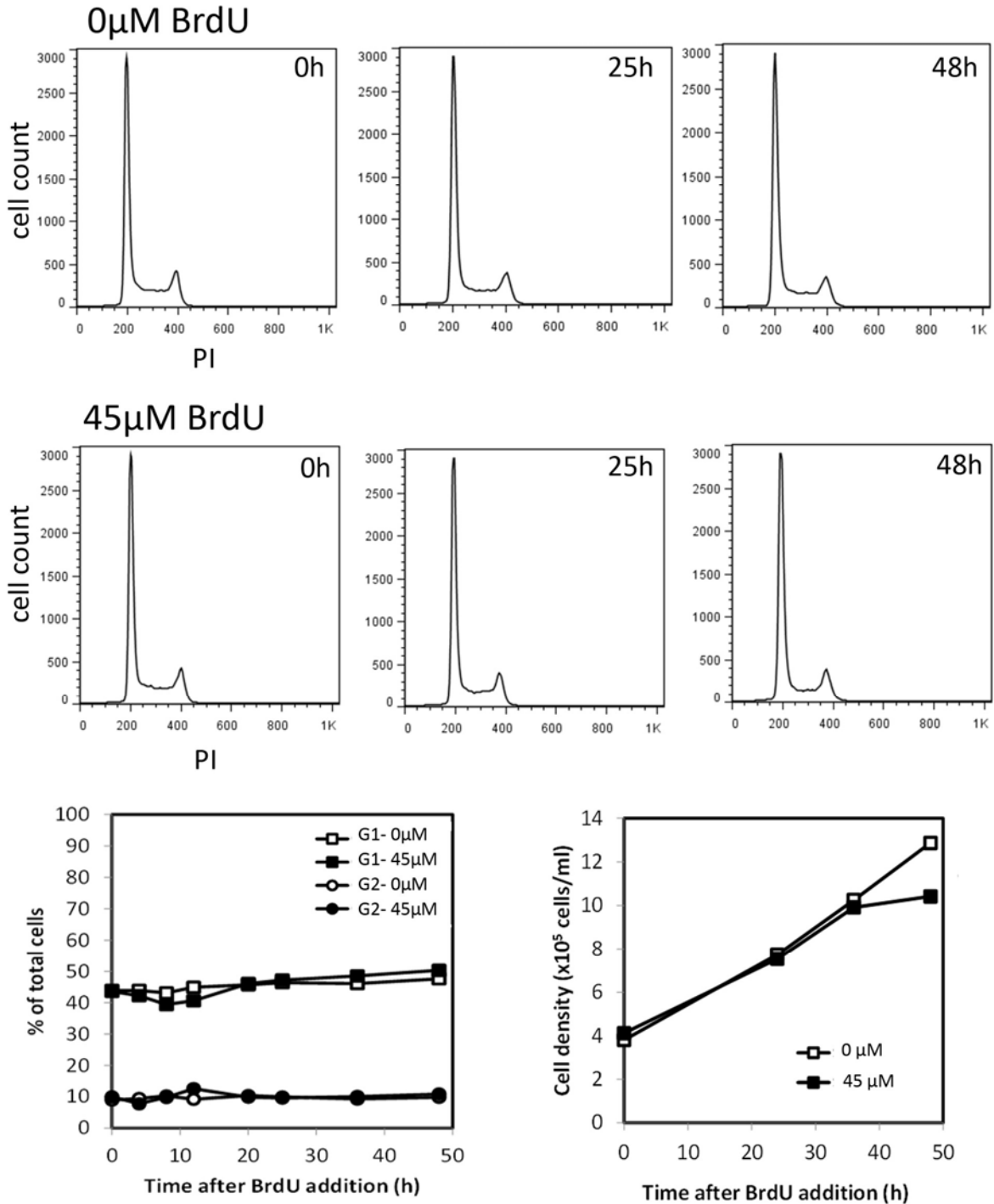


Figure 3.3: Checking for BrdU dependant G2/M arrest in Cell line 14

Cell cycle profiles for cells grown in the absence (top panel) or presence (middle panel) of 45 μ M BrdU for 0, 25 and 48 hours. The bottom, left panel shows fraction of cells in G1 (squares) and G2 (circles) for cells grown with 0 (open) or 45 μ M (filled) BrdU. The bottom, right panel shows cell growth for cells grown in 0 (open) or 45 μ M (filled) BrdU.

Cell line 16

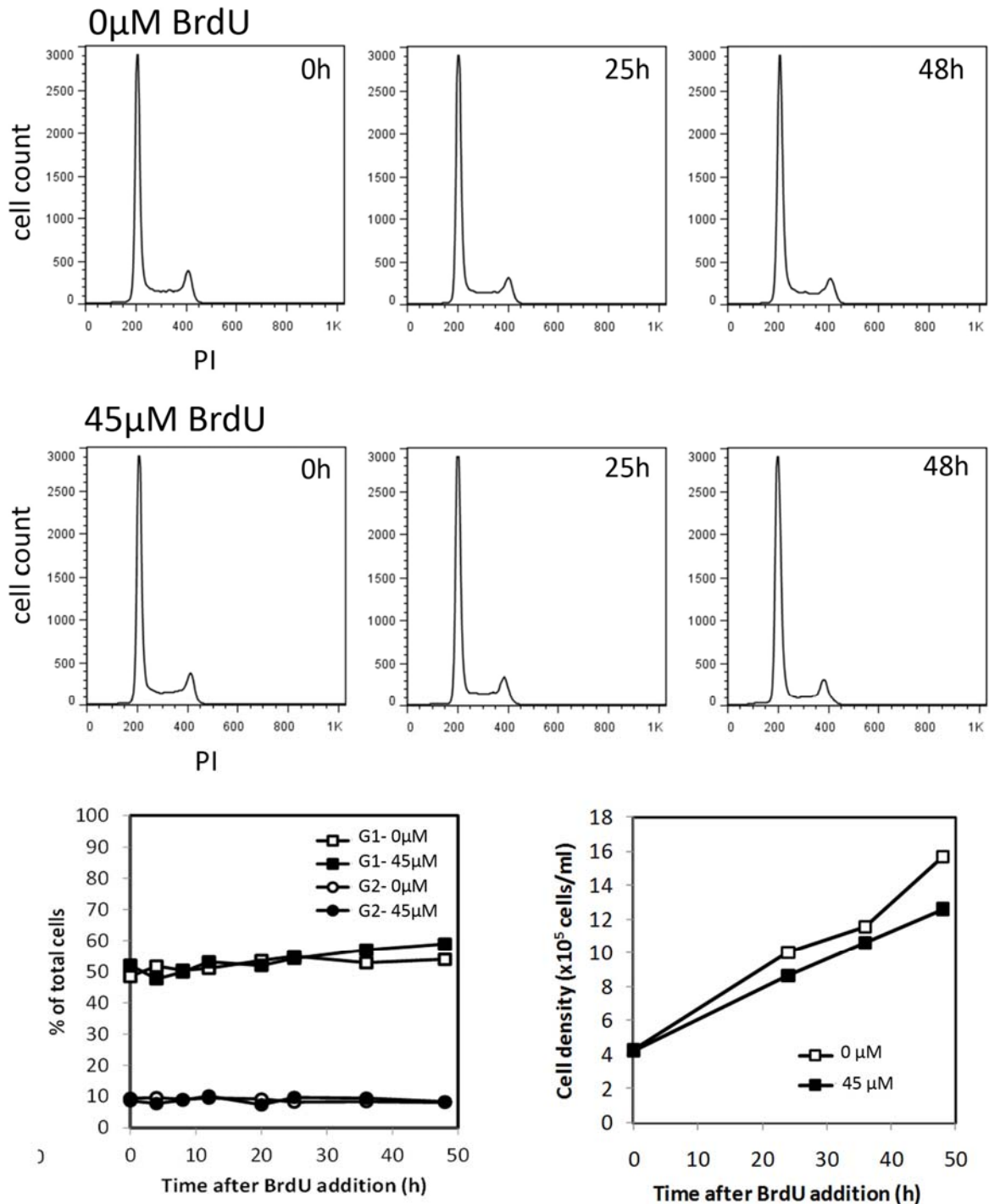


Figure 3.4: Checking for BrdU dependent G2/M arrest in Cell line 16

Cell cycle profiles for cells grown in the absence (top panel) or presence (middle panel) of 45 μ M BrdU for 0, 25 and 48 hours. The bottom, left panel shows fraction of cells in G1 (squares) and G2 (circles) for cells grown with 0 (open) or 45 μ M (filled) BrdU. The bottom, right panel shows cell growth for cells grown in 0 (open) or 45 μ M (filled) BrdU.

Cell line 21

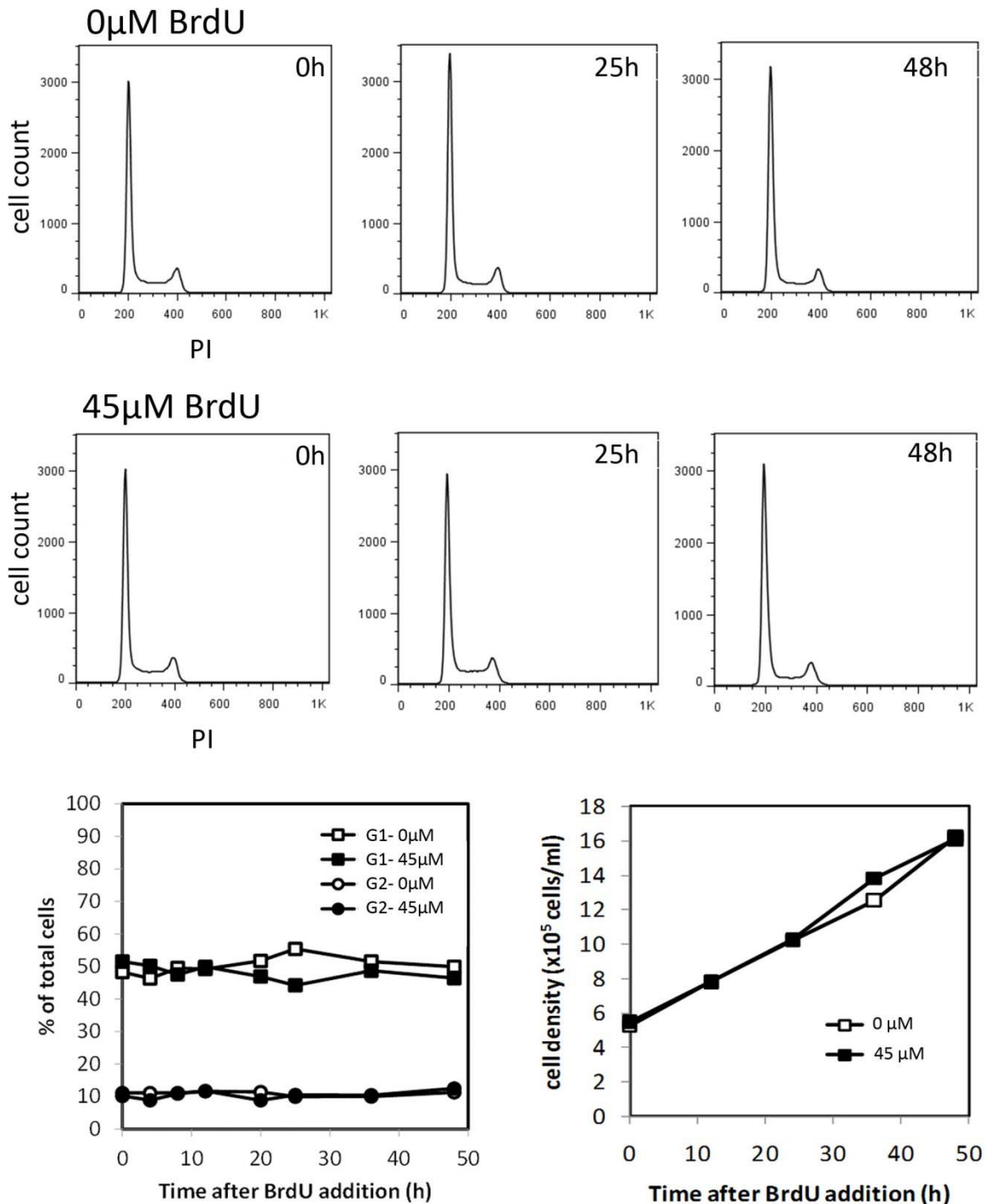


Figure 3.5: Checking for BrdU dependent G2/M arrest in Cell line 21

Cell cycle profiles for cells grown in the absence (top panel) or presence (middle panel) of 45 μ M BrdU for 0, 25 and 48 hours. The bottom, left panel shows fraction of cells in G1 (squares) and G2 (circles) for cells grown with 0 (open) or 45 μ M (filled) BrdU. The bottom, right panel shows cell growth for cells grown in 0 (open) or 45 μ M (filled) BrdU.

Cell line 22

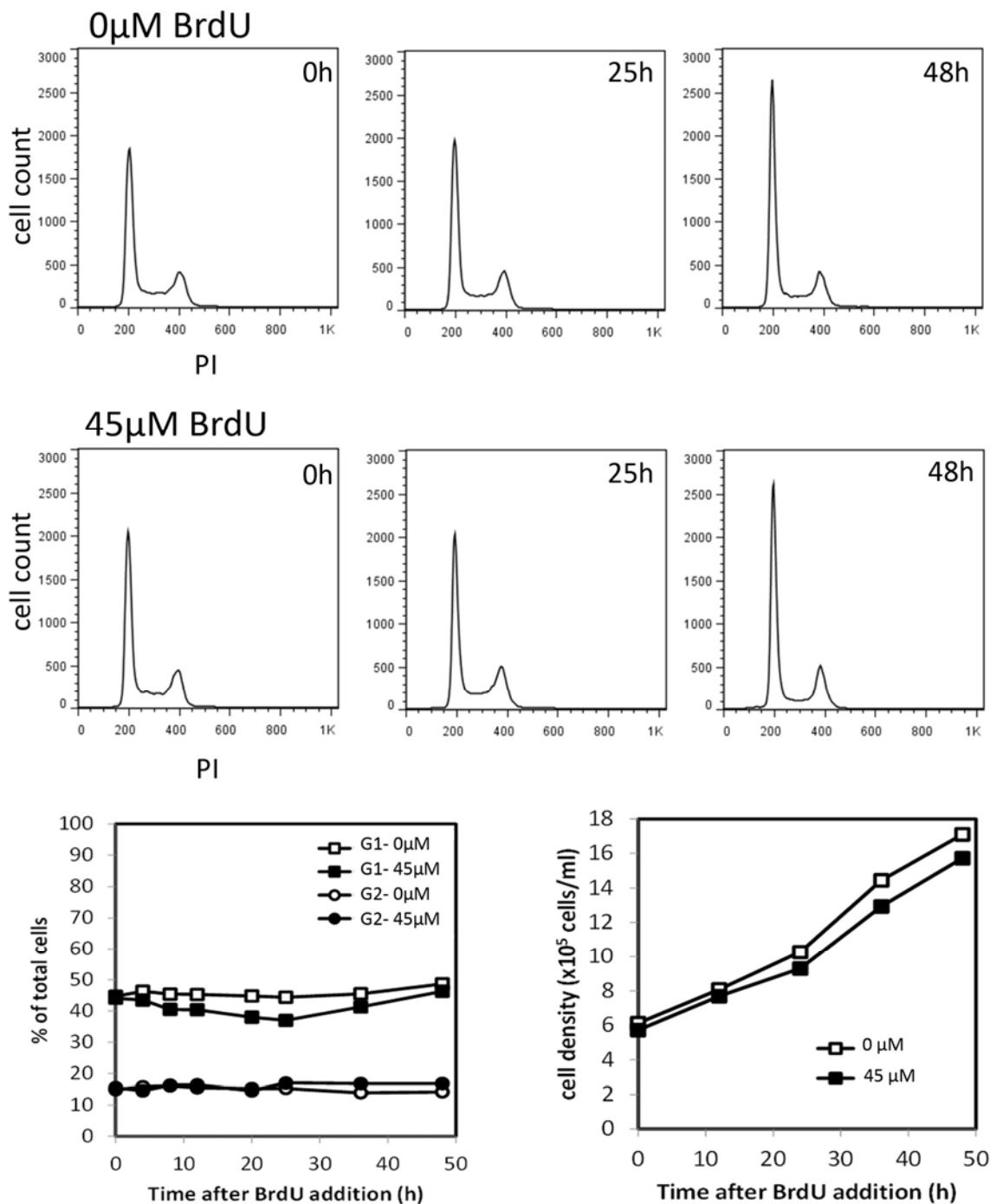


Figure 3.6: Checking for BrdU dependant G2/M arrest in Cell line 22

Cell cycle profiles for cells grown in the absence (top panel) or presence (middle panel) of 45 μM BrdU for 0, 25 and 48 hours. The bottom, left panel shows fraction of cells in G1 (squares) and G2 (circles) for cells grown with 0 (open) or 45 μM (filled) BrdU. The bottom, right panel shows cell growth for cells grown in 0 (open) or 45 μM (filled) BrdU

Cell line 24

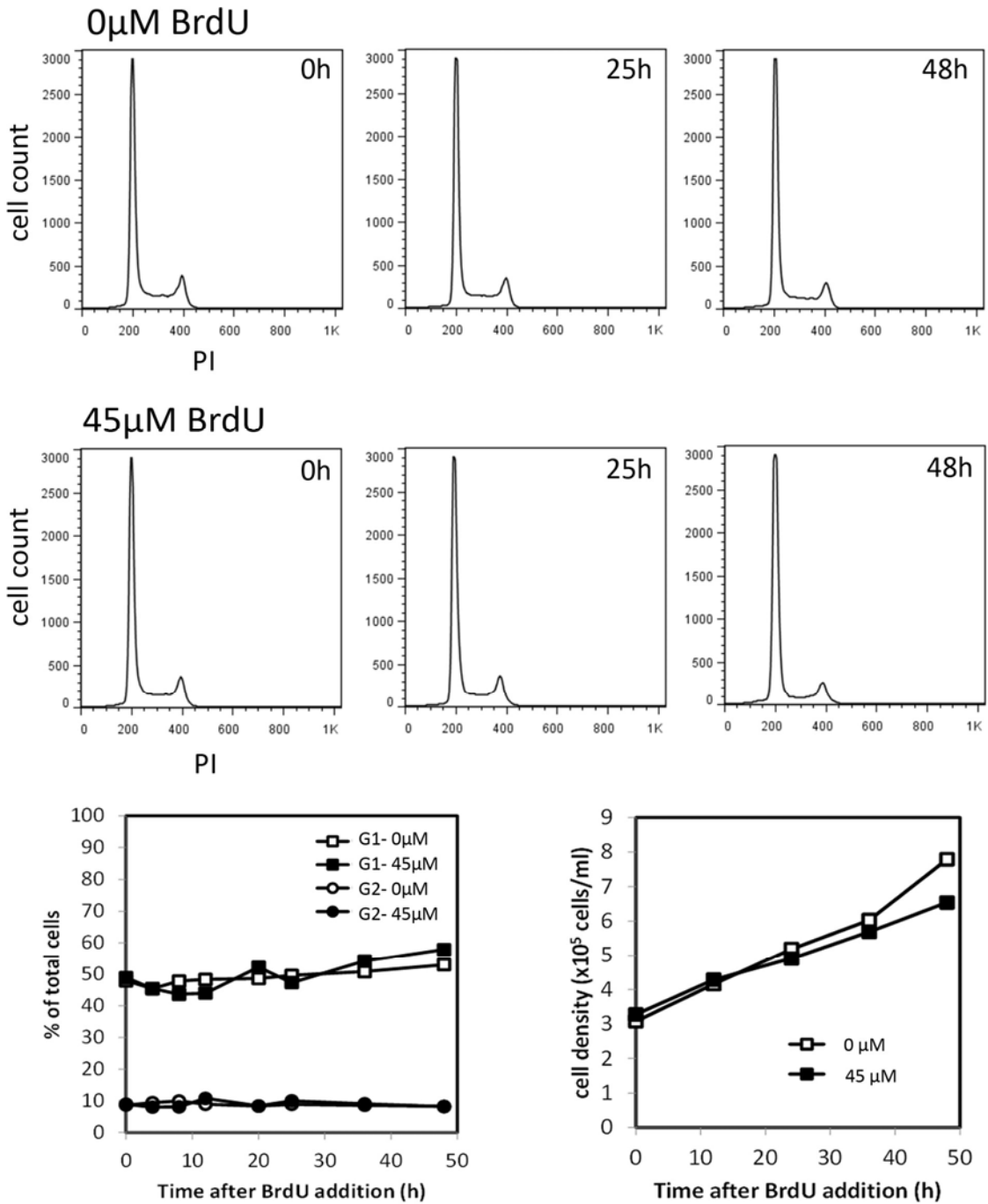


Figure 3.7: Checking for BrdU dependant G2/M arrest in Cell line 24

Cell cycle profiles for cells grown in the absence (top panel) or presence (middle panel) of 45 μ M BrdU for 0, 25 and 48 hours. The bottom, left panel shows fraction of cells in G1 (squares) and G2 (circles) for cells grown with 0 (open) or 45 μ M (filled) BrdU. The bottom, right panel shows cell growth for cells grown in 0 (open) or 45 μ M (filled) BrdU

Table 3.1: The partition of the panel of 24 cell lines based on their doubling times

Cell Line	CCR Catalog No.	Normal doubling time (h)	Group No.	Approx. doubling time (h)
8	GM15510	17	1	18
22	GM15324	19	1	18
21	GM15268	20	2	22
20	GM15242	20	2	22
16	GM15072	21	2	22
24	GM15061	21	2	22
6	GM15224	22	2	22
14	GM15038	22	2	22
1	GM15029	22	2	22
17	GM15144	22	2	22
3	GM15215	23	2	22
5	GM15245	24	2	22
15	GM15056	24	2	22
23	GM15386	24	2	22
12	GM15385	25	3	27
13	GM15590	25	3	27
9	GM15213	26	3	27
19	GM15226	27	3	27
7	GM15236	27	3	27
2	GM13036	28	3	27
4	GM15223	30	3	27
11	GM15227	30	3	27
18	GM15216	35	4	37
10	GM15221	40	4	37

The 24 cell lines were ordered based on their normal doubling time, divided into four groups. Cell lines in each group were assayed based on the approximate doubling time for the group.

Based on preliminary experiments, the doses chosen for BCNU treatment were 0, 10, 20, 40, 60 and 80 μ M BCNU. Cell lines with similar doubling times that belonged to the same sub-group were assayed on one 96-well plate. The survival curves were obtained using the proliferation assay (see Materials and Methods). Once the data was collected, all flow cytometry data was analyzed using FlowJo (TreeStar Inc.) as described in the Chapter 2. Gates were drawn appropriately and the percentage of proliferating cells was calculated to obtain survival curves for the 24 cell lines. The survival curves obtained in this manner for the panel of 24 cell lines are shown in Figure 3.8.

From Figure 3.8, one can see that there is a wide range of sensitivities for the panel of cell lines exposed to BCNU. The panel of genetically varied cell lines was also previously shown to have a wide range of sensitivities to other alkylating agents MNNG (23) and MMS (unpublished data). To identify any correlation between the sensitivities of the cell lines for the three agents, a correlation coefficient was calculated between the three possible pairs of treatments: (BCNU, MMS), (BCNU, MNNG) and (MNNG, MMS). As seen in Figure 3.9, the correlation between the sensitivity values for the three DNA damaging agents is weak, ($R^2= 0.15, 0.28$ and 0.23 respectively) thus suggesting that the factors that affect the sensitivity of the cell lines to the three agents are different.

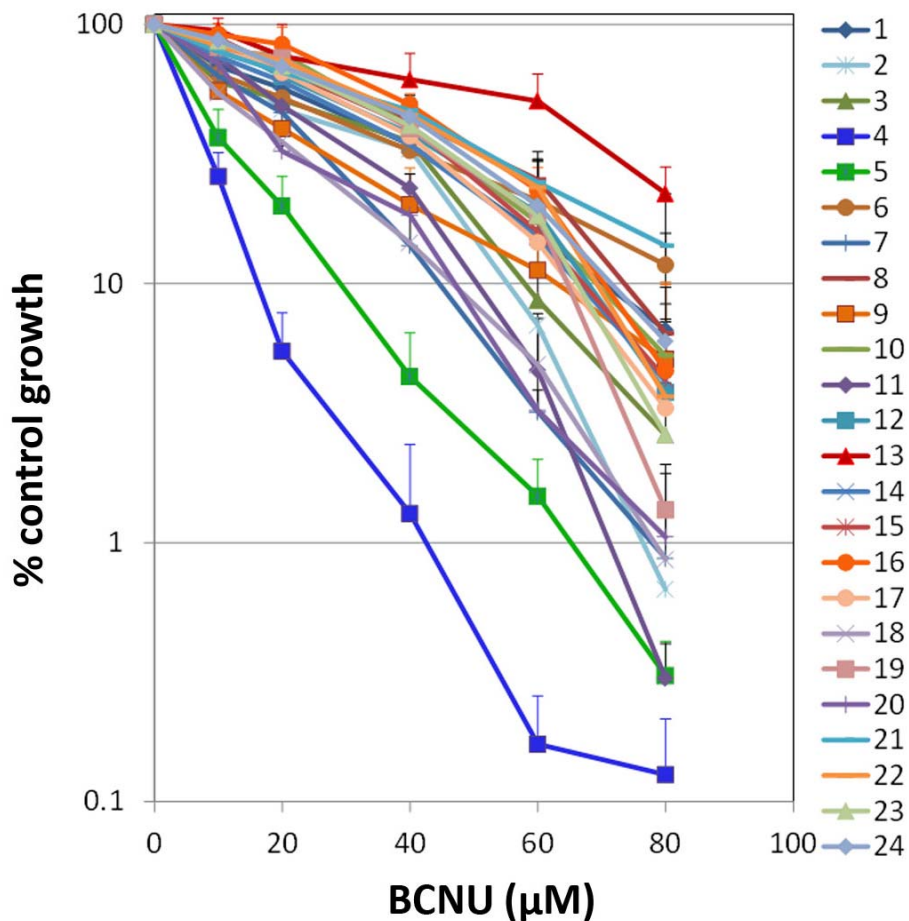


Figure 3.8: Survival curves for the 24 cell lines after BCNU treatment

Survival curves for the panel of 24 genetically varied cell lines for BCNU doses of 0, 10, 20, 40, 60 and 80 μM. Graphs were obtained using the proliferation assay described in Chapter 2

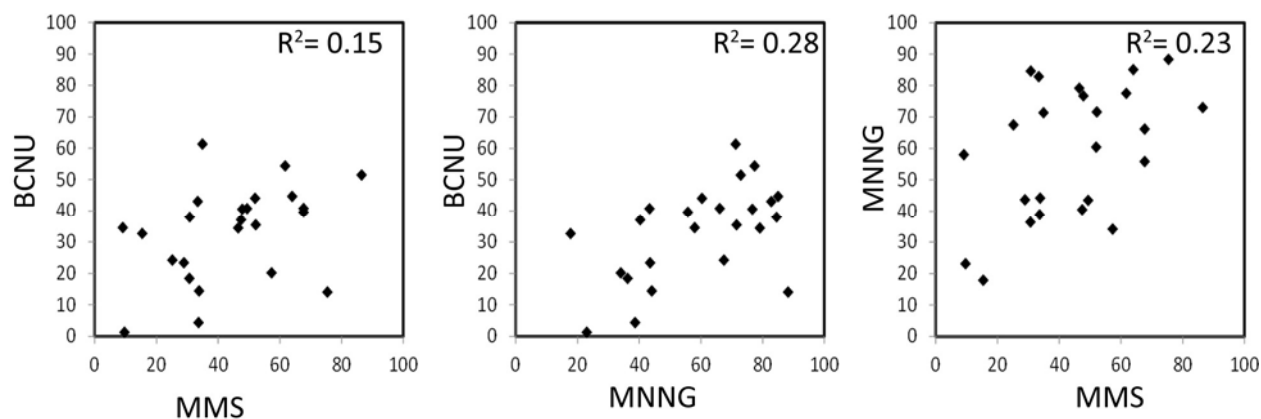


Figure 3.9: Correlation between sensitivities of the 24 cell lines to MMS, MNNG and BCNU

The R^2 for correlation between the sensitivity of the cell lines for the MMS (0.4mM), MNNG (0.5 μ g/ml) and BCNU (40 μ M) was calculated for each of the pairs (BCNU, MMS), (BCNU,MNNG) and (MNNG, MMS). The values for R^2 obtained are shown on each plot.

To identify genes whose basal expression values correlate well with BCNU sensitivity, and possibly affect the BCNU sensitivity of the cell lines, previously obtained basal expression for the twenty four cell lines from (23) was used. These basal gene expression values were correlated with the BCNU sensitivity values of the cell lines at the 40 μ M dose. Those genes that had an $r > 0.6$ or $r < -0.6$ with a p-value < 0.05 of correlation were selected. 123 probe-sets corresponding to 94 genes passed these criteria and are shown in Table 3.2 and a heat-map representation of their expression values is shown in Figure 3.10A. Of these genes, only one gene, *O*⁶-methylguanine DNA methyltransferase (MGMT) is positively correlated with BCNU sensitivity. This is reassuring since MGMT is known to remove the *O*⁶-chloroethyl adduct generated by BCNU and confer resistance to BCNU treatment in the clinic. All of the other genes included within this list are negatively correlated with BCNU resistance and most have not yet been associated with BCNU sensitivity or resistance. Gene ontology enrichment (see Materials and Methods) shows that the 94 genes are enriched for molecular processes involved in protein catabolism (see Figure 3.10B), a process yet to be studied in the context of BCNU exposure.

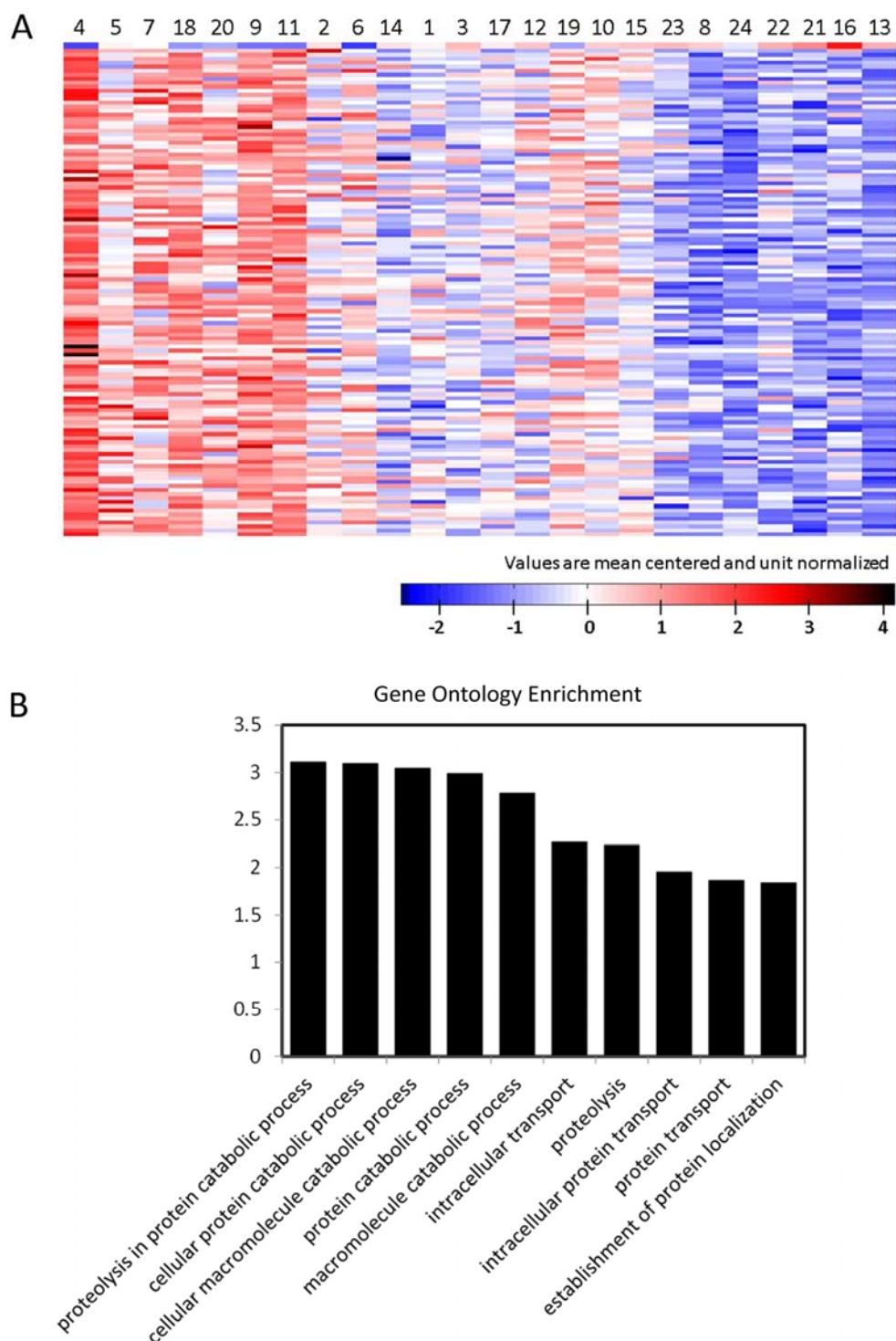


Figure 3.10: Genes whose basal expression correlates with BCNU sensitivity

(A) Heat-map of 123 probesets (corresponding to 94 genes) with high positive or negative correlation with BCNU sensitivity ($r > 0.6$ or $r < -0.6$ with p -value < 0.05) ordered according to their correlation. The top-most gene is MGMT and is the only positively correlated gene; (B) Gene ontology enrichment within the set of 94 genes obtained using GO-FAT from DAVID (v7.6)

Table 3.2: 123 probesets with high positive or negative correlation to BCNU sensitivity

probeset	Gene ID	Gene name
204880_at	MGMT	O-6-methylguanine-DNA methyltransferase
223299_at	SEC11C	SEC11 homolog C (<i>S. cerevisiae</i>)
225674_at	BCAP29	B-cell receptor-associated protein 29
201823_s_at	RNF14	ring finger protein 14
209174_s_at	QRICH1	glutamine-rich 1
203102_s_at	MGAT2	mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
223305_at	TMEM216	transmembrane protein 216
236254_at	VPS13B	vacuolar protein sorting 13 homolog B (yeast)
213374_x_at	HIBCH	3-hydroxyisobutyryl-Coenzyme A hydrolase
1552472_a_at	ACAP2	ArfGAP with coiled-coil, ankyrin repeat and PH domains 2
1555399_a_at	DUSP16	dual specificity phosphatase 16
223835_x_at	OTP	orthopedia homeobox
218487_at	ALAD	aminolevulinate, delta-, dehydratase
204313_s_at	CREB1	cAMP responsive element binding protein 1
239288_at	TNIK	TRAF2 and NCK interacting kinase
242943_at	ST8SIA4	ST8 alpha-N-acetylneuraminide alpha-2,8-sialyltransferase 4
201732_s_at	CLCN3	chloride channel 3
232520_s_at	NSFL1C	NSFL1 (p97) cofactor (p47)
203099_s_at	CDYL	chromodomain protein, Y-like
218172_s_at	DERL1	Der1-like domain family, member 1
1552302_at	TMEM106A	hypothetical LOC728772; transmembrane protein 106A
218088_s_at	RRAGC	Ras-related GTP binding C
223569_at	PPAPDC1B	phosphatidic acid phosphatase type 2 domain containing 1B
201988_s_at	CREBL2	cAMP responsive element binding protein-like 2
225677_at	BCAP29	B-cell receptor-associated protein 29
230462_at	NUMB	numb homolog (<i>Drosophila</i>)
223568_s_at	PPAPDC1B	phosphatidic acid phosphatase type 2 domain containing 1B
201735_s_at	CLCN3	chloride channel 3
200917_s_at	SRPR	signal recognition particle receptor (docking protein)
222729_at	FBXW7	F-box and WD repeat domain containing 7
212542_s_at	PHIP	pleckstrin homology domain interacting protein
228041_at	AASDH	aminoadipate-semialdehyde dehydrogenase
203758_at	CTSO	cathepsin O
225306_s_at	SLC25A29	solute carrier family 25, member 29
229018_at	C12orf26	chromosome 12 open reading frame 26
207124_s_at	GNB5	guanine nucleotide binding protein (G protein), beta 5
235705_at	TRIO	triple functional domain (PTPRF interacting)
204562_at	IRF4	interferon regulatory factor 4
209913_x_at	KIAA0415	KIAA0415
229419_at	FBXW7	F-box and WD repeat domain containing 7
210653_s_at	BCKDHB	branched chain keto acid dehydrogenase E1, beta polypeptide
230029_x_at	UBR3	ubiquitin protein ligase E3 component n-recognin 3 (putative)
235196_at	Cdc73	cell division cycle 73, Paf1/RNA polymerase II complex component, homolog (<i>S. cerevisiae</i>)
213373_s_at	CASP8	caspase 8, apoptosis-related cysteine peptidase
241364_at	TMEM57	transmembrane protein 57
201989_s_at	CREBL2	cAMP responsive element binding protein-like 2
225644_at	CCDC117	coiled-coil domain containing 117
224453_s_at	ETNK1	ethanolamine kinase 1
213483_at	PPWD1	peptidylprolyl isomerase domain and WD repeat containing 1
228711_at	ZNF37A	zinc finger protein 37A
222496_s_at	RBM47	RNA binding motif protein 47
218588_s_at	FAM114A2	family with sequence similarity 114, member A2
240410_at	--	--

Table 3.2 (contd.)

probeset	Gene ID	Gene name
217906_at	KLHDC2	kelch domain containing 2
227980_at	ZNF322A	zinc finger protein 322A
203567_s_at	TRIM38	tripartite motif-containing 38
226897_s_at	ZC3H7A	zinc finger CCCH-type containing 7A
209207_s_at	SEC22B	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)
215930_s_at	CTAGE4	CTAGE family, member 5 pseudogene; CTAGE family member; CTAGE family, member 4; CTAGE family, member 5
213009_s_at	TRIM37	tripartite motif-containing 37
222669_s_at	SBDS	Shwachman-Bodian-Diamond syndrome pseudogene; Shwachman-Bodian-Diamond syndrome
221857_s_at	TJAP1	tight junction associated protein 1 (peripheral)
203486_s_at	ARMC8	armadillo repeat containing 8
228242_at	N4BP2	NEDD4 binding protein 2
220980_s_at	ADPGK	ADP-dependent glucokinase
203447_at	PSMD5	proteasome (prosome, macropain) 26S subunit, non-ATPase, 5
201358_s_at	COPB1	coatamer protein complex, subunit beta 1
206958_s_at	UPF3A	UPF3 regulator of nonsense transcripts homolog A (yeast)
235670_at	STX11	syntaxin 11
225927_at	MAP3K1	mitogen-activated protein kinase kinase kinase 1
201990_s_at	CREBL2	cAMP responsive element binding protein-like 2
206925_at	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4
218827_s_at	CEP192	centrosomal protein 192kDa
1554260_a_at	FRYL	FRY-like
226719_at	--	--
227915_at	ASB2	ankyrin repeat and SOCS box-containing 2
203098_at	CDYL	chromodomain protein, Y-like
233857_s_at	ASB2	ankyrin repeat and SOCS box-containing 2
201659_s_at	ARL1	ADP-ribosylation factor-like 1
204566_at	PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform
201384_s_at	NBR1	neighbor of BRCA1 gene 1
212006_at	UBXN4	UBX domain protein 4
201098_at	COPB2	coatamer protein complex, subunit beta 2 (beta prime)
203610_s_at	TRIM38	tripartite motif-containing 38
218456_at	CAPRN2	caprin family member 2
201559_s_at	CLIC4	chloride intracellular channel 4
200762_at	DPYSL2	dihydropyrimidinase-like 2
210385_s_at	ERAP1	endoplasmic reticulum aminopeptidase 1
209011_at	TRIO	triple functional domain (PTPRF interacting)
218035_s_at	RBM47	RNA binding motif protein 47
226529_at	TMEM106B	transmembrane protein 106B
221881_s_at	CLIC4	chloride intracellular channel 4
202809_s_at	INTS3	integrator complex subunit 3
210609_s_at	TP53I3	tumor protein p53 inducible protein 3
210041_s_at	PGM3	phosphoglucomutase 3
202318_s_at	SENPA6	SUMO1/sentrin specific peptidase 6
230836_at	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4
218302_at	PSENE1	presenilin enhancer 2 homolog (C. elegans)
222787_s_at	TMEM106B	transmembrane protein 106B
243507_s_at	C20orf196	chromosome 20 open reading frame 196
1552303_a_at	TMEM106A	hypothetical LOC728772; transmembrane protein 106A
203833_s_at	TGOLN2	trans-golgi network protein 2
212993_at	NACC2	NACC family member 2, BEN and BTB (POZ) domain containing
239175_at	AFTPH	aftiphilin
212239_at	PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)

Table 3.2 (contd.)

probeset	Gene ID	Gene name
216986_s_at	IRF4	interferon regulatory factor 4
209497_s_at	RBM4B	RNA binding motif protein 4B
205120_s_at	SGCB	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)
224832_at	DUSP16	dual specificity phosphatase 16
224336_s_at	DUSP16	dual specificity phosphatase 16
212890_at	SLC38A10	solute carrier family 38, member 10
209234_at	KIF1B	kinesin family member 1B
204000_at	GNB5	guanine nucleotide binding protein (G protein), beta 5
233300_at	--	--
218930_s_at	TMEM106B	transmembrane protein 106B
214323_s_at	UPF3A	UPF3 regulator of nonsense transcripts homolog A (yeast)
204872_at	TLE4	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)
212249_at	PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
212240_s_at	PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
214688_at	TLE4	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)
217993_s_at	MAT2B	methionine adenosyltransferase II, beta
237034_at	--	--
226112_at	SGCB	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)

3.4.4 Cell cycle profiles for cell lines with extreme BCNU sensitivity/resistance define the time frame for microarray measurements

From the wide range of sensitivities observed in the panel of 24 cell lines after BCNU treatment, we chose the two cell lines that were most sensitive to BCNU (cell lines 4 and 5) and the two that were most resistant (cell lines 13 and 16) for further analyses. Although the proliferation assay gives a measure of how sensitive a cell line is to BCNU, it provides no information about the time frame within which key cellular decisions are made after BCNU treatment. To better gauge the duration after DNA damage within which interesting transcriptional changes might occur, we measured some phenotypic properties of the four cell lines to gain insight into the changes that occur in these cell lines after BCNU treatment. Informative time points at which to measure transcripts were identified. We also chose to use 40 μ M BCNU for all of the subsequent experiments since this dose had only a slight effect on the resistant cell lines, while inducing significant death in the sensitive cell lines, but not so much death that it would be impossible to harvest cells for RNA isolation.

One well studied property of cells after DNA damage is the initiation of cell cycle checkpoints in the presence of damage. BCNU has previously been shown to induce late S or G2/M arrest in glioblastoma cell lines (24). We reasoned that if such an arrest occurs with the set of four cell lines used here, the time points prior to initiation of the cell cycle arrest, and all the way up to resolution of the arrest or initiation of cell death would yield informative transcriptional data.

Cell cycle profiles measured at various time points after DNA damage provide information on whether cell cycle arrest occurs in treated versus mock-treated cells, how long after treatment

such an arrest occurs, the duration and nature of the arrest, and whether cells are able to resolve the arrest and resume normal cell cycle progression. The final piece of information that can be obtained from cell cycle profile measurements is the time, post-treatment, when a sub-G1 population becomes visible in the sensitive cell lines, thus indicating the possible time point at which DNA fragmentation begins.

Therefore, with the goal of identifying time points for global transcriptional profiling, cell cycle profiles were measured for the four cell lines after BCNU treatment. The four cell lines were treated with BCNU (see Materials and Methods). Cells were collected at multiple time-points all the way up to 96 hours post treatment for cell cycle profile measurements by flow cytometry (as described in the Methods section).

Figure 3.11 and Figure 3.12 show cell cycle profiles of the sensitive cell lines 4 and 5 respectively at multiple time-points post BCNU treatment. These cell cycle profiles four important features - i) there is an initial accumulation of cells in the S-phase of the cell cycle slightly visible at 12 hours and more prominent at 24 and 36 hours post BCNU treatment; ii) there is slow progression of cells through S-phase towards 4N DNA content all the way up to the 48 hour time point, accompanied by a decrease in the G1 population; iii) there is an increase in the sub-G1 population beginning at 24 hours for the most sensitive cell line (cell line 4) and at 48 hours for the other sensitive cell line 5, and increasing all the way up to the 96 hour time point; iv) the accumulation of cells in G2/M is minimal in the sensitive cells. In contrast to the sensitive cell lines, the cell cycle profiles of the resistant cell lines 16 and 13 (seen in Figure 3.13 and Figure 3.14) show no obvious differences between treated and mock-treated samples.

Cell line 4

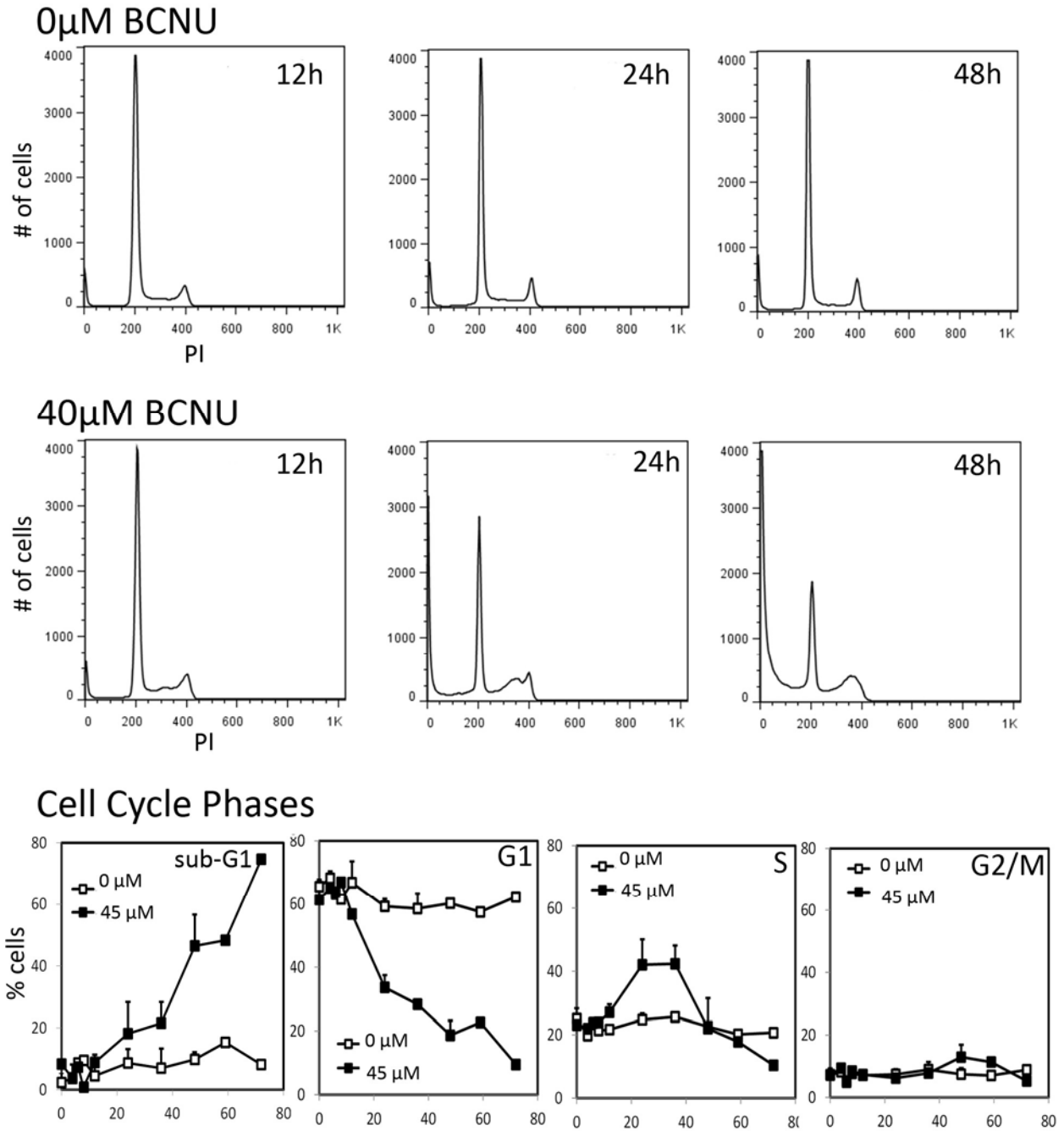


Figure 3.11: Cell cycle progression of the BCNU sensitive cell line 4

Cell cycle profiles for cell line 4 either mock-treated (top panel) or BCNU treated (middle panel) at 12, 24 and 48 hours. The bottom panel shows quantification of sub-G1, G1, S and G2 cell cycle phases at each time point for mock-treated (open) and BCNU treated (filled) samples.

Cell line 5

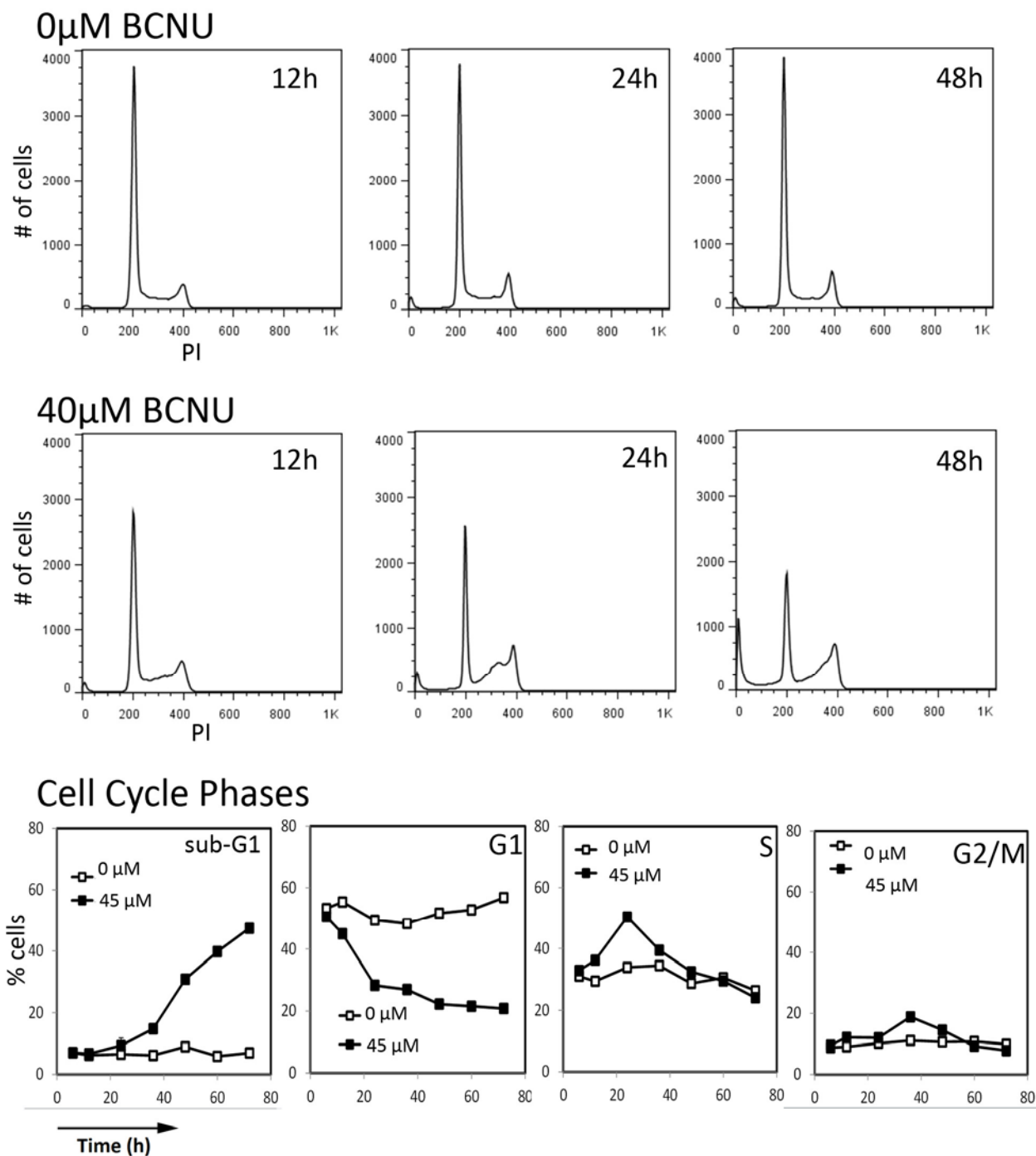


Figure 3.12: Cell cycle progression of the BCNU sensitive cell line 5

Cell cycle profiles for cell line 5 either mock-treated (top panel) or BCNU treated (middle panel) at 12, 24 and 48 hours. The bottom panel shows quantification of sub-G1, G1, S and G2 cell cycle phases at each time point for mock-treated (open) and BCNU treated (filled) samples.

Cell line 16

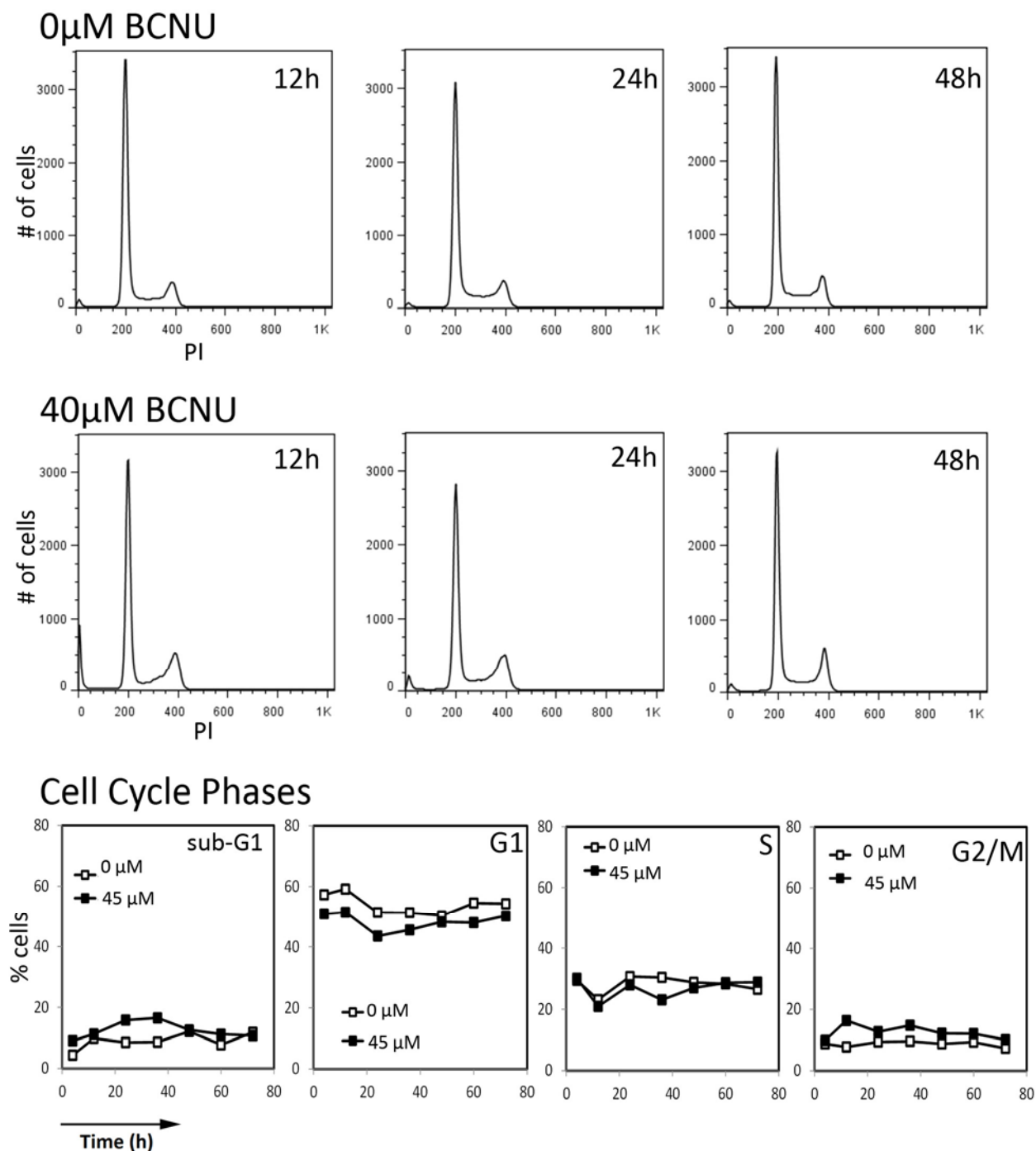


Figure 3.13: Cell cycle progression of the BCNU resistant cell line 16

Cell cycle profiles for cell line 13 either mock-treated (top panel) or BCNU treated (middle panel) at 12, 24 and 48 hours. The bottom panel shows quantification of sub-G1, G1, S and G2 cell cycle phases at each time point for control (open) and BCNU treated (filled) samples.

Cell line 13

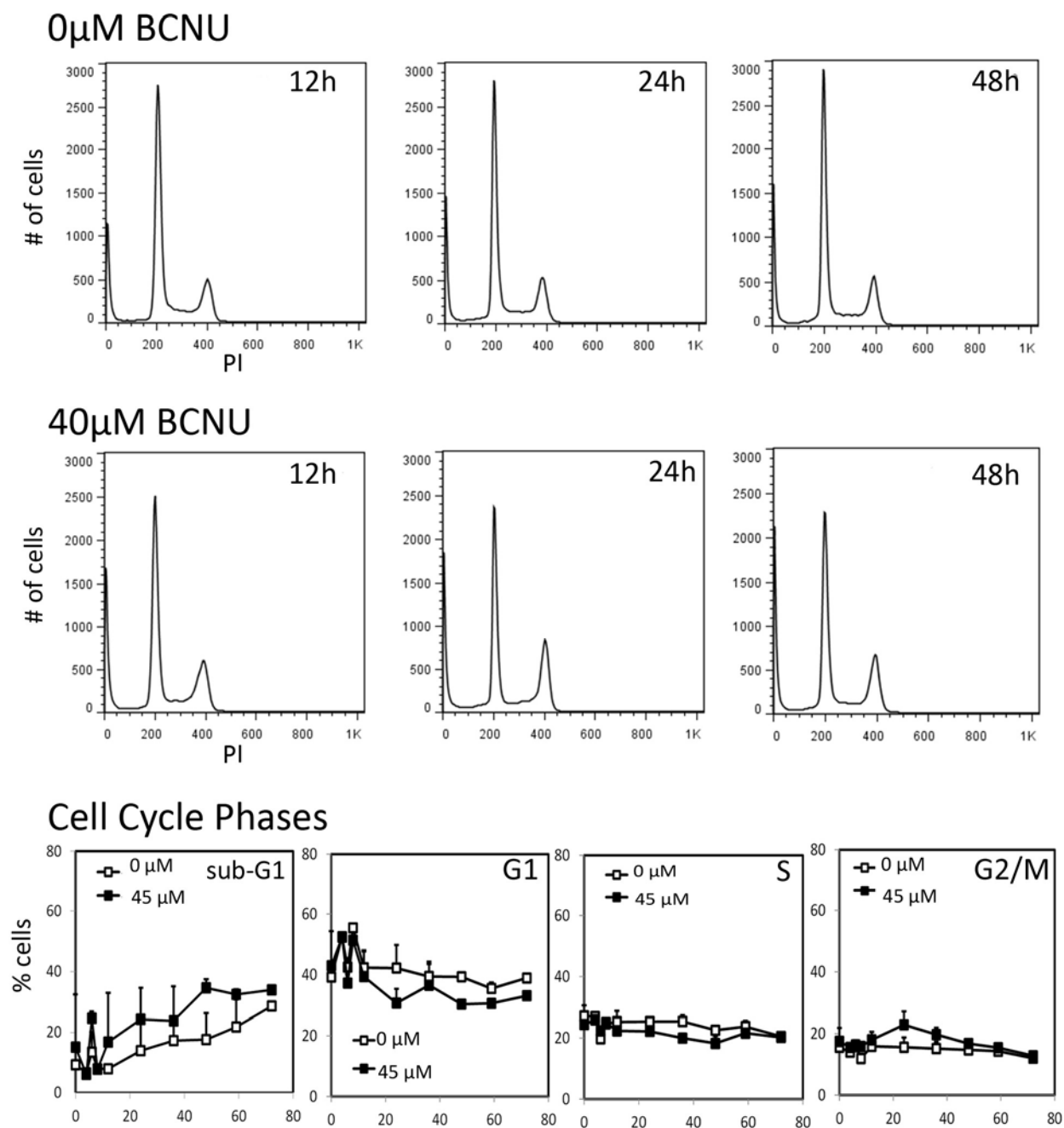


Figure 3.14: Cell cycle progression of the BCNU resistant cell line 13

Cell cycle profiles for cell line 13 either mock-treated (top panel) or BCNU treated (middle panel) at 12, 24 and 48 hours. The bottom panel shows quantification of sub-G1, G1, S and G2 cell cycle phases at each time point for control (open) and BCNU treated (filled) samples.

3.4.5 Viable cell number measurements for BCNU sensitive/resistant cell lines show growth inhibition for sensitive cell lines but not resistant cell lines

To complement the cell cycle profile measurements, viable cell numbers were measured by trypan blue exclusion to determine the relative growth rates between treated and mock-treated samples. Figure 3.15 shows the viable cell numbers for the four cell lines. The figure shows that the two sensitive cell lines show strong growth inhibition after BCNU exposure as compared to mock-treated samples. In contrast, the viable cell measurements of the resistant cell lines only showed a slight lag in cell growth. The fact that the cell cycle profile measurements of the resistant cell lines showed no discernible stalling of cells in any one phase of the cell cycle suggests that the lag in cell growth in resistant cell lines is not due to a strong cell cycle arrest but could be due to a fleeting arrest that was either missed or undetectable by the assay used.

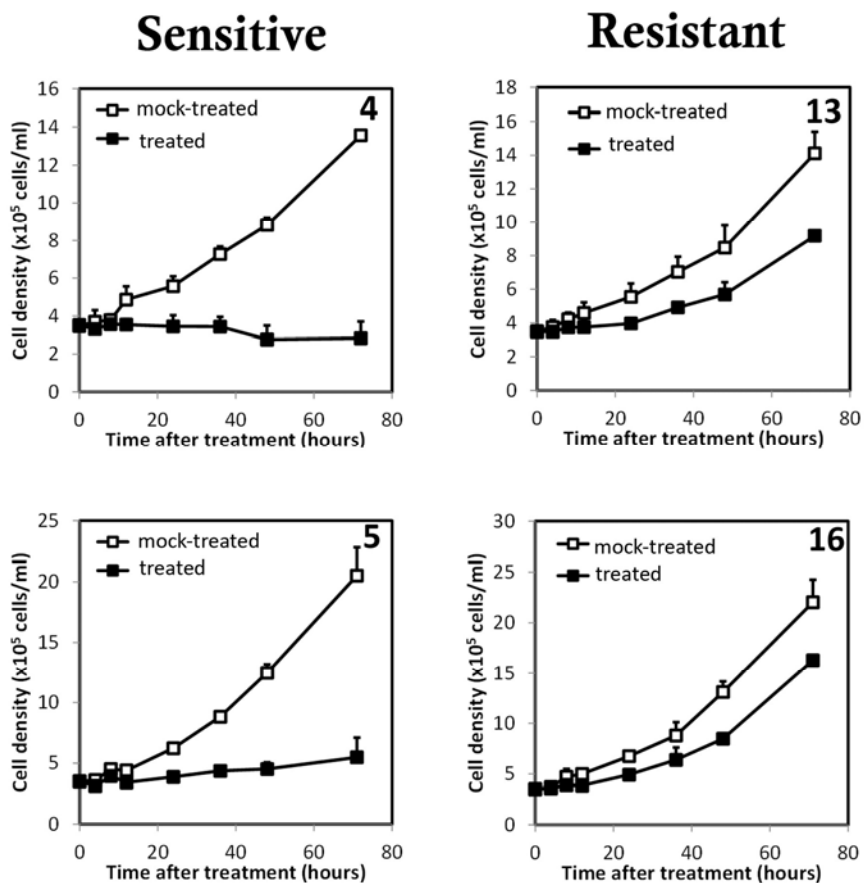


Figure 3.15: Growth inhibition of cell lines showing extreme BCNU sensitivity/resistance
Cell density as measured by trypan blue exclusion for the two sensitive cell lines (left panels) and two resistant cell lines (right panels) for control (open) and BCNU treated (filled) samples. All cell lines were treated with 40 μ M BCNU

3.4.6 Designing and generating a multi-dimensional transcriptional data set

Taken together, the cell cycle profile and viability measurements for the two most sensitive and two most resistant cell lines defined the time points at which global transcription profiles should be measured. The first signs of S-phase accumulation occurred in sensitive cell lines at the 12 hour time-point. However, transcription programs take time to be activated in order to change the state of the cell. Moreover, any transcript changes that could explain the slight lag in growth of the resistant cell lines would most likely be observed early in the time course. Therefore, early time points including 4 hours, 8 hours and 12 hours were chosen to capture any transcript changes that might be involved in the accumulation of sensitive cells in the S-phase of the cell cycle and the slower growth of the resistant cell lines. As observed from the late increase in the sub-G1 population in sensitive cell lines, cell death was a late event, with an increase in the sub-G1 population only seen after 24-48 hours. Therefore, additional late time-points were measured, including 24, 36, 48 and 72 hours to capture transcripts that might provide insight into the observed cell death.

Once the time points for microarray measurements were determined, there was one more decision to be made regarding the design of the experiment. Since the BCNU treatment involves an hour-long incubation of cells in serum-free media, we anticipated that serum starvation would induce a stress response as was observed in (25). It was therefore important to measure transcript changes in mock-treated samples that also underwent an hour-long incubation in serum-free media minus BCNU, to ensure correct identification of DNA damage related transcriptional changes. Therefore, transcript changes were measured in both mock-treated and BCNU treated samples to discern the BCNU treatment related expression changes.

With all the parameters, time points and conditions determined, time-series experiments were performed. Cell lines 4, 5, 13 and 16 were treated with 40 μ M BCNU as described in the Materials and Methods section. Cells were collected at 0, 4, 8, 12, 24, 36, 48 and 72 hours post BCNU treatment and RNA was isolated. The RNA was tested for integrity and hybridized onto Affymetrix HG-U133 plus 2.0 chips (see Materials and Methods).

3.4.7 Exploring the transcriptional data space

As a first step towards understanding the transcriptional changes after BCNU treatment, approximately 10,000 probe-sets that varied over time, treatment or cell line were visualized. These probe-sets are shown in Figure 3.16 and reveal some interesting properties of the data set. Many probe-sets that show induction or repression in treated samples relative to the 0h time-point also show similar induction or repression in mock-treated samples over time (examples are highlighted in the box). Therefore gene expression changes are seen in both mock-treated and treated samples showing that these changes are not necessarily BCNU dependent.

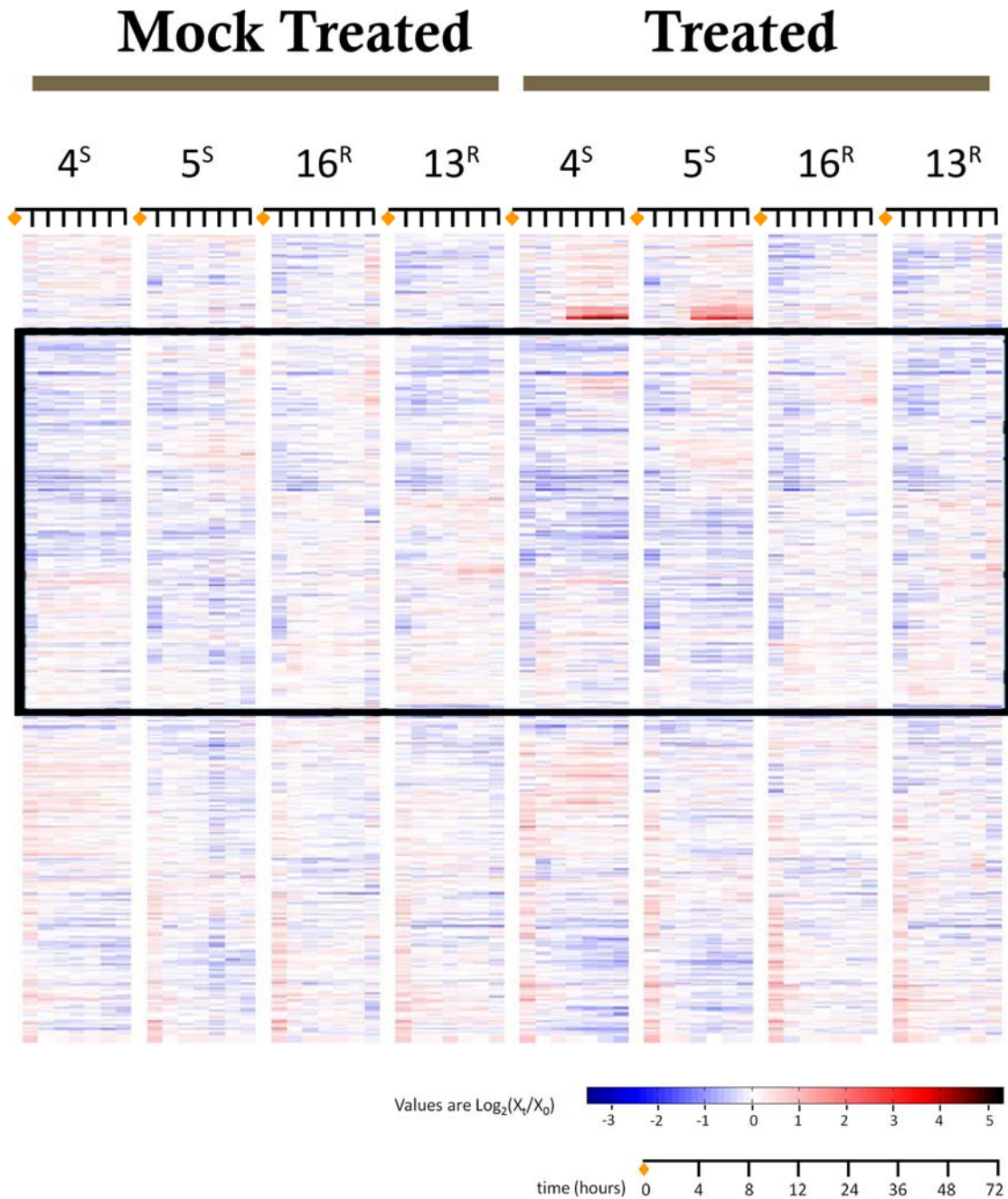


Figure 3.16: Transcripts that vary with time, cell line and treatment

The heat-map shows probe-set values (rows) for each array (column). Values are log₂ ratios as compared to the 0h time-point. Mock-treated samples are on the left and treated samples are on the right. The black box marks examples of probe-sets that show expression changes in both mock-treated and treated samples over time.

One reason for such varying gene expression in mock-treated samples might be the inherent expression changes that occur as cells progress from the early log-phase of growth to mid-log and finally to the stationary phase of growth (26). Other studies have shown that gene-expression changes with change in proliferation rate, which changes as cells move along the growth curve (27). In addition to growth-rate induced gene-expression changes, there is another possible cause for the observed gene expression changes within mock-treated samples in this particular experiment, namely serum-starvation. Even short exposure to serum-free media has been shown to induce transcriptional change in certain cell types (25). Regardless of the cause of gene expression changes over time in the mock-treated samples, we know that these changes are not related to or in response to BCNU treatment. Therefore, we took the approach of identifying transcripts that are differentially induced or repressed in the treated versus mock-treated samples over time to obtain true BCNU-induced transcriptional changes.

A second observation from visualizing the data set in Figure 3.16 is that gene induction and repression patterns even in the mock-treated samples differ between the two sensitive cell lines (compare the first and second columns within the box in Figure 3.16). Therefore, there was no guarantee that the two sensitive cell lines would have similar gene expression responses to BCNU treatment. However, if we were to identify such a set of genes that behaved similarly in the two sensitive cell lines but differently between sensitive and resistant cell lines, this signature could be a strong indicator of key transcriptional changes that affect cell sensitivity to BCNU.

3.4.8 Identifying a general transcriptional signature for cell death/survival after DNA damage

From our data set, we wanted to identify genes that were induced or repressed differentially between sensitive and resistant cell lines but similarly in the two sensitive cell lines, or in the two resistant cell lines upon BCNU treatment. For example genes induced after BCNU treatment to a greater extent in the two sensitive cell lines but to a lesser extent in the resistant cell lines might help us describe transcriptional programs that are induced when cells are destined to die after DNA damage. Conversely, genes that are induced to a greater extent in treated resistant cell lines as compared to the sensitive cell lines, might give us insight into any transcripts that protect cells against DNA damaging agents. Moreover, requiring gene expression changes to follow similar expression patterns in the two sensitive or two resistant cell lines eliminates any cell line specific gene expression changes.

In our data set, each gene expression value is associated with three discrete variables – treatment (0 or 40 μ M BCNU), time (0, 4, 8, 12, 24, 36, 48, 72 hours) and sensitivity (sensitive or resistant cell line). Using 3-way ANOVA in MATLAB, genes that were significantly different in the treatment and sensitivity variables either at all time-points or at a subset of time-points were identified (details of the analysis are described in the Materials and Methods section). This yielded 984 probe-sets representing 706 genes that were differentially induced or repressed between the sensitive and resistant cell lines. Within this set of 984 probe-sets, differential gene expression values between sensitive and resistant cell lines at any of the eight time points ranged from 4.2 to -1.7 when calculated as a \log_2 ratio of sensitive to resistant gene expression value.

3.4.9 Visualizing the DNA damage induced transcriptional signature

As a first step towards understanding the biological meaning and relevance of the 984 probe-sets extracted using ANOVA, the probe-sets were clustered to identify striking expression patterns present within the gene set. A heat-map representing the gene set clustered using hierarchical clustering (GenePattern suite, Broad Institute) is shown in Figure 3.17.

The heat-map shows that the 984 probe-sets (representing 706 genes) can be divided into two groups – those that are induced, and those that are repressed to a greater extent in the sensitive treated cells compared to the resistant treated cells. The concerted induction and repression of these genes more so in the two sensitive cell lines as compared to the two resistant cell lines suggests that these genes are most likely relevant for the decision to die rather than survive after DNA damage. To further investigate the biological context of these 706 genes and the cellular pathways they are involved in, network analysis was performed on the selected gene sets.

3.4.10 Network and canonical pathway analyses reveal the biological relevance of the transcriptional signature

The network connectivity within the set of 984 probe-sets was obtained through the use of Ingenuity Pathways Analysis (Ingenuity® Systems, www.ingenuity.com). Out of the 706 genes, 507 were eligible for functional and pathway analysis and 295 were eligible for network analysis (only those genes and interactions that were experimentally identified in human cell lines were considered eligible). The top four significant networks are involved in cell death, cell cycle, cancer and cell growth proliferation (Figure 3.18A).

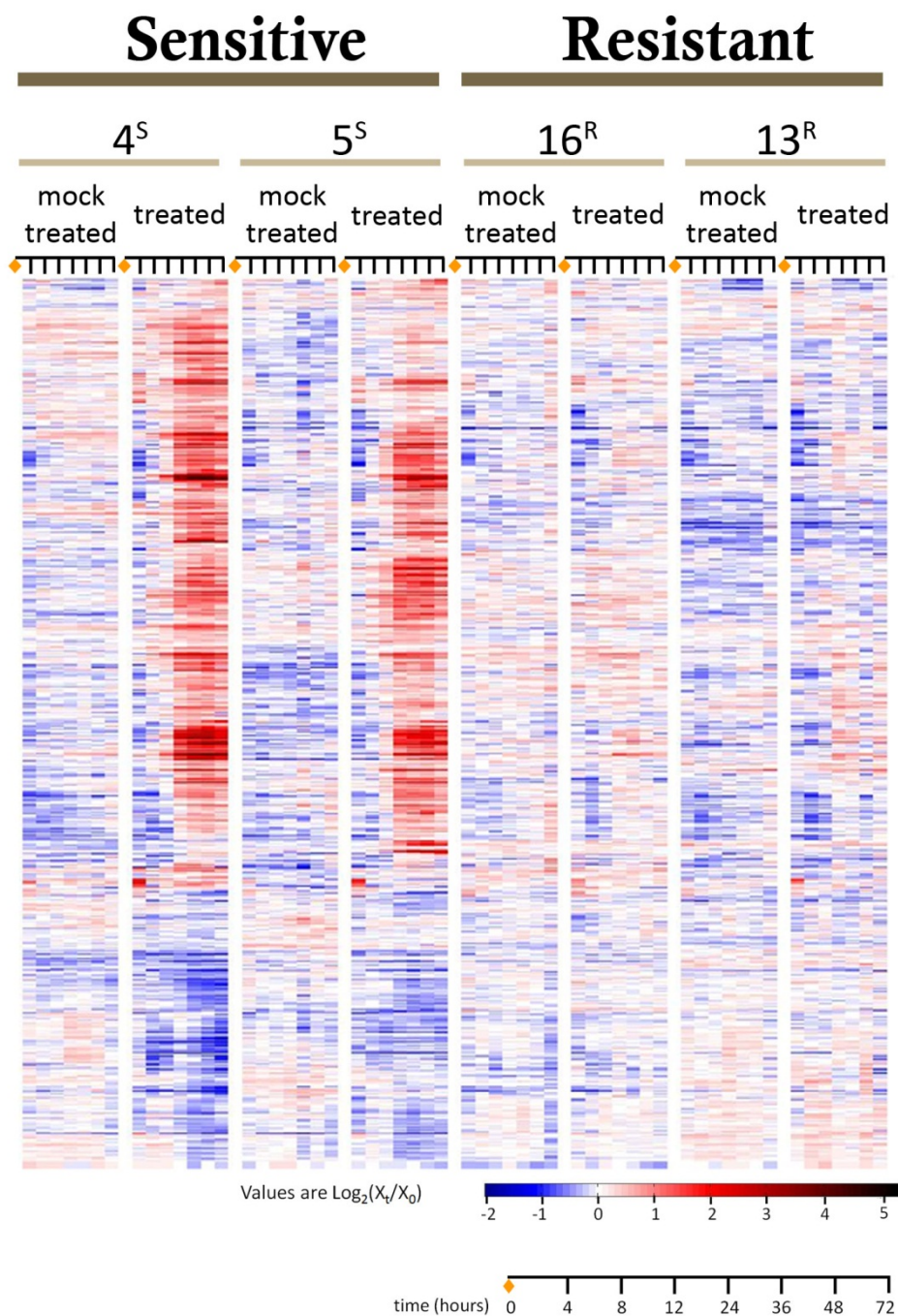


Figure 3.17: Hierarchical clustering of the BCNU transcriptional signature

Each row is a probe-set, and each column one time-point. Values are log ratios as compared to the 0h time-point. Mock-treated and treated samples for each cell-line are shown next to each other.

Focus molecules, indicated in bold, are the molecules in our gene set also present in the network. These four networks are interconnected and form one large network containing 273 of the 295 network eligible genes (Figure 3.18B). What this shows is that a majority of the network eligible genes from our signature set form interconnected signaling and interaction modules that are involved in cell death, cell cycle, growth and proliferation.

Looking at sub-networks within the larger network obtained from our gene set, we see examples of signaling modules that could provide some insight into the pathways utilized by the sensitive cell lines in the response to BCNU treatment. Figure 3.19 shows an example of a sub-network containing the tumor suppressor and DNA damage response modulator p53 as the hub node regulating targets from our selected gene set that are involved in DNA replication, repair and cell cycle. Identifying p53 related signaling modules and the pathways and networks that control cell cycle, cell death and proliferation increases our confidence in the selected gene set as a DNA damage response gene set and provides some insight into possible signaling modules involved in cell death/survival decisions after BCNU treatment.

In addition to network analyses, canonical pathways that were enriched within our data set were identified. Canonical pathways, as defined by Ingenuity, are predefined and well studied pathways within the database. Figure 3.20 shows the canonical pathways that are enriched within our transcriptional signature. p53 signaling, G2/M arrest, aryl-hydrocarbon receptor signaling, cell cycle regulation and mitotic roles of Polo-like kinase are the most significantly enriched canonical pathways within the entire gene set. We infer that the transcriptional signature genes

are involved in key DNA damage response pathways with the exception of the aryl-hydrocarbon pathway that has not yet been implicated in DNA damage response.

Both the network and canonical pathway analyses confirm that among the set of 706 genes identified from the ANOVA analysis, a large fraction are meaningful in the response of a cells to a DNA damaging agent and that they contribute to cell death, survival and cell cycle control. However, because these analyses were performed with the entire gene set consisting of both up-regulated and down-regulated genes, the resulting networks were complex and hard to comprehend. Instead, if it were the case that the up-regulated or down-regulated genes were involved in functionally separate pathways, the system and transcriptional response might be easier to understand.

A

Molecules in Network	Score	Focus Molecu	Top Functions
265 Proteasome, ↑ ADRB2, Ait, Ap1, ↑ APAF1, ↑ AREG/AREGB, ↑ ARL6IP5, ↑ ASH1L, ↑ AURKA, ↑ AUTS2, ↑ AXL1, ↑ BTG2, ↑ BTRC, ↑ BUB1B, ↑ CS, ↑ CALCOO1, ↑ CASP1, Caspase, ↑ CNA2, ↑ CCNB1, ↑ CND1, ↑ CD2, ↑ CD44, ↑ CDC20, ↑ CDKN1A, ↑ CDKN1C, ↑ CENPE, Collagen(s), Creb, ↑ CSNK2A1, ↑ CSP4, ↑ CTNND1, Cyclin A, ↑ DAB2IP, ↑ DPYSL2, E2f, ↑ E2F2S1, ↑ E2F4, ↑ E2F4L1, ↑ E2F4L2, ↑ E2F4L3, ↑ ERK, ERK1/2, Estrogen Receptor, ↑ ETS2, ↑ F2R, ↑ FAS, ↑ FCER1A, ↑ FGFR2, ↑ FGFR3, Fibrinogen, ↑ FOSL1, FSH, ↑ GAB1, ↑ GADD45A, ↑ GDF15, ↑ GK, ↑ GORASP2, ↑ GRM1, ↑ GRN, hCG, ↑ HDAC2, ↑ HIST4H4 (includes others), Histone h3, Histone h4, ↑ HMMR, Hsp70, Hsp90, ↑ HSP90A1, ↑ IGFBP3, ↑ IGFBP4, ↑ IL1RN, Interferon alpha, ↑ IRF1, ↑ ITGA3, ↑ ITGA2B, Jnk, ↑ KAT2B, ↑ KITLG, ↑ KRT17, ↑ LAMP3, LDL, ↑ LGALS8, Lh, ↑ LIF, ↑ LIPA, ↑ LOC81691, Mapk, ↑ MAPK1, ↑ MDM2, Mediator, ↑ NBPF11 (includes others), ↑ NF2, NFkB (complex), ↑ NOTCH1, P38 MAPK, p85 (pi3k), ↑ PADI4, PDGF BB, ↑ PDSSA, ↑ PGF, PI3K (complex), Plk, Plk(s), ↑ PLK1, ↑ PLK2, ↑ PML, ↑ PODXL, ↑ PPP2CA, ↑ PRF1, ↑ PTAFR, ↑ PTPRF, ↑ PYHIN1, ↑ RBI, ↑ RBL2, ↑ RGS12, ↑ RGS16, RNA polymerase II, ↑ RNF144B, ↑ SERPINA1, ↑ SERPINE1, ↑ SKP2, ↑ SMAD5, ↑ SMC3, ↑ SMC4, ↑ STK10, ↑ STK11, ↑ SYTL1, ↑ TACR1, ↑ TFDP2, ↑ TNFRSF10B, ↑ TNFRSF10D, ↑ TP53B, ↑ TPX2, ↑ TRIM22, ↑ TUBB2C, ↑ TYMS, ↑ UBE2C, ↑ YLPM1	125	108	Cell Death, Cell Cycle, Cancer
ABCB1, ↑ AIMP2, ↑ ANK1, ↑ ASPM, ATF3, ATM, ↑ BBX3, BCL2L1, BIRC5, ↑ BTG1, ↑ BUB1, ↑ BUB1B, C10orf119, ↑ CAPRINI, CASP2, CASP3, CASP8, CBL, ↑ CCNA2, CCNE2, CCNT1, ↑ CD24, ↑ CDC20, ↑ CDC14B, CDC25C, ↑ CDKN1A, CDKN1B, CDT1, ↑ CEACAM1 (includes others), ↑ CENPE, ↑ CENPF, ↑ CHEK1, CHEK2, CIT, COL1A1, CRKL, CSF2, Cyclin B, ↑ CYP4F2, ↑ DDDB2, ↑ DDX54, ↑ DEPDC1, ↑ DFFA, ↑ EEF1E1, ERBB2, ERC3, ESR1, ↑ EXOSC2, F Actin, ↑ FADS3, FOSL2, FOXO1, FOXO3, ↑ GADD45A, GADD45B, ↑ GDF15, ↑ GMEB1, ↑ GNG4, ↑ GTSE1, H2AFX, HCK, HGF, ↑ HMMR, ↑ HNRNPAB, HNRNPD, ↑ HNRNPDL, HRAS, ↑ IGFBP4, ↑ IGHG1, JAG1, ↑ KAT2B, KHSPR, ↑ KIF14, ↑ KIF23, ↑ KIF18A, ↑ KITLG, LDLR, ↑ LIF, LMNA, ↑ MAD1L1, MAD2L1, ↑ MCM4, ↑ MCM6, MCM7, ↑ MDM2, MDM4, ↑ MFGE8, MMP2, ↑ MSH6, NCAFG, NDC80, ↑ NEK2, ↑ NUPR1, ↑ PADI4, ↑ PBK, ↑ PHLDA3, ↑ PIK3CA, ↑ PIK3IP1, PIK3R1, PLAT, ↑ PLK1, ↑ PLK2, ↑ PLXNB1, ↑ PLXNB2, PMS1, ↑ POLH, ↑ PPM1D, PRC1, PRKAA2, ↑ PRKAB1, ↑ PROCR, ↑ PRODH, ↑ PTPN22, RBC1C, Rsk, SATB1, ↑ SBF2, SFN, SHC1, ↑ SHISA5, SMARCA4, ↑ SMC3, SMPD1, SP1, SPHK2, SREBF1, ↑ SRSF3, STMN1, ↑ TNFRSF10B, TOPBP1, TP53, TP73, ↑ TP53B, ↑ TRIM22, ↑ TRIM32, ↑ USP6, Vegf, WTL, WWOX, ↑ XPC	59	70	Cell Cycle, Cell Death, Cellular Growth and Proliferation
↑ ADCK3, AKT1, ↑ ALDH1A3, ↑ AMD1, ↑ AP15, ↑ APL1, APP, ↑ ARNTL2, ATF1, ↑ ATG16L1, ATG4B, ATR1B1, ↑ ATXN3, B2M, ↑ BAG3, BCL2L1, ↑ BUB3, ↑ C10orf110, ↑ CAPN2, ↑ CASP1, CASP6, CAV1, ↑ CCDC90B, CCL2, CCL4, ↑ CCT4, CD47, CEACAM3, CELF1, CHIBL1, ↑ CLASP1, Collagen(s), CRP, ↑ DDR1, ↑ DKC1, ↑ DKK1, ETS1, ↑ ETS2, ↑ F2R, FADD, FANCC, FCER1G, FGF2, ↑ FLOT2, ↑ GABARAP, GADD45B, GAR1, ↑ GCHL, ↑ GEMINS, GEMIN6, GEMIN7, ↑ GM2A, ↑ GPR109B, GPX4, Hsp90, IDO1, IFI6, ↑ IFNA8, IFNB1, IKKKG, IL2, IL13, IL27, ↑ IL13RA1, IL2RA, IL2RB, IL2RG, ↑ IRF1, ITGAV, ↑ ITIM2B, JAG1, JUNB, KAT5, ↑ KITLG, ↑ KLRCL1, ↑ KPNA1, LAMA3, LAMC2, ↑ LRDD, LTF, Mapk, mir-122, MMP1, MX1, ↑ NCAM1, NFKB2, NFKBIA, NFKBIE, ↑ NID1, NR3C1, ↑ PAPOLA, ↑ PAPPA, ↑ PCLO, PIK3CB, ↑ PLEKHF1, ↑ PLXNC1, ↑ PML, PPARG, PRKCA, PRKCB, PRKCE, PSMB9, ↑ PSMD11, PTPN11, ↑ RABEPK, RAN, REL, ↑ RHCE/RHD, RIPK2, RUNX1, SAT1, ↑ SDCI1, ↑ SERPINE1, SMN1/SMN2, SMPD1, ↑ SMPD3, SNCA, ↑ SOX4, SP100, SPINT2, STAT1, STAT5B, ↑ STOM, ↑ SYNCRIP, TERC, TERT, TIMP3, TNF, ↑ TNFRSF10C, TNFSF13, TNFSF14, TNIP1, ↑ TOR1A, ↑ TP53B, ↑ TRIM45, ↑ TYH3, ↑ UNC13A, ↑ USP2, VCP, ↑ WDR12	44	59	Cell Death, Cellular Development, Hematological System Development and Function
Actin, ACTL6A, Alpha tubulin, ↑ ANKHD1, ↑ ANLN, ANXA2, BARD1, BEX2, BLOC1S1, ↑ BLOC1S2, BLOC1S3, ↑ BUB1B, CAND1, ↑ CCNA2, ↑ CCNB1, ↑ CCNB2, CCNE2, CCNE2, ↑ CCT8, ↑ CDC43, CDH1, CDH2, CDH11, CDT1, ↑ CEACAM1 (includes others), ↑ CENPA, ↑ CENPE, CHEK2, CKS1B, CLDN4, CLRN1, CTNNB1, CUL1, DDX11/DDX12, DIO2, DISC1, ↑ DLGAP5, DR1, DRAP1, DTNBP1, E2F4, ↑ ECT2, EEF1A1, ↑ EPHA2, ↑ EPSBL2, ↑ EWSR1, ↑ F11R, ↑ FBXW2, FBXW7, FLNA, Gamma tubulin, ↑ GTPBP4, H2AFX, ↑ HIST1H1B, ↑ HJURP, ↑ HMMR, ↑ HSP90B1, ITGAV, ↑ JRK, KAT5, KDM5B, KDM6A, ↑ KPNA2, LDB1, LMO2, ↑ MINA, ↑ MK167, ↑ MKI67P, MLL, MLL2, MMP2, MMP3, MMP7, ↑ MPZL2, MUTED, MYBL2, MYC, MYCN, NBN, NDC80, ↑ NEK2, ↑ NHLH2, NOP58, ↑ NUFIP1, PACRG, PAXIP1, ↑ PDE4C, PLDN, PNN, POLR2A, PP2A, PPARG, PRC1, ↑ PSME3, ↑ PTPRF, RACGAP1, RAD51, RAN, ↑ RASGRF1, RBX1, RNF4A, RPAP3, RUVBL1, ↑ SDCI1, ↑ SKP2, ↑ SLC6A2, SNAPIN, ↑ SOCS2, SOX9, ↑ SPOCK1, ↑ SPTBN1, ↑ SRSF2, ↑ SSB, ↑ STRAP6, ↑ STX6, SUZ12, TAF1A, ↑ TAF1B, TAF1C, ↑ TCEB1, ↑ TCPL1, TFAP2A, TGFBI, ↑ TMPO, TOP2A, ↑ TP53BP1, ↑ TP53TG1, ↑ TRAK1, TUBGCP2, ↑ TUBGCP3, ↑ TUBGCP4, TUBGCP5, TUBGCP6, UXT, VAMP3, ↑ WDR5, ↑ XPOS, ↑ XYL1L1, ZNF143, ZNF217	44	59	Cell Cycle, Cancer, Cellular Assembly and Organization

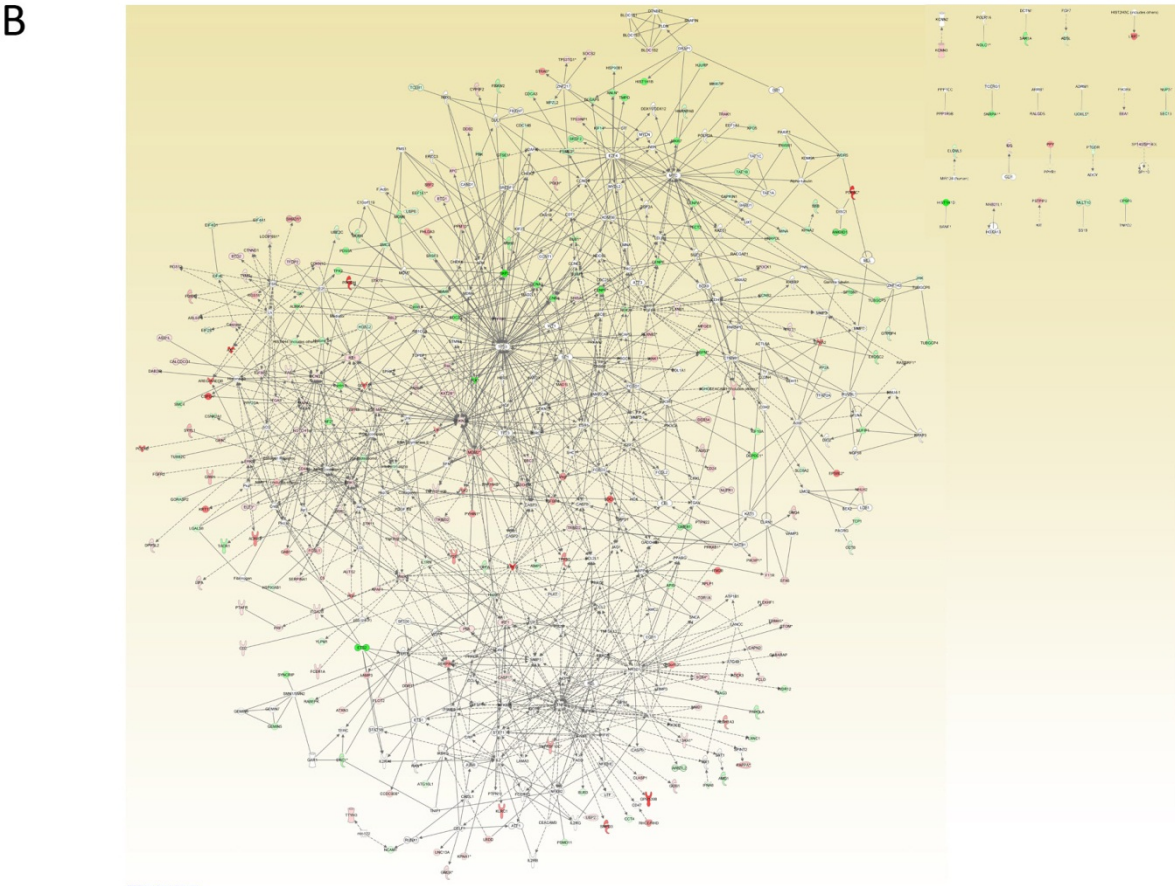


Figure 3.18: Network connectivity of the transcriptional signature

A) Top four networks obtained from the 295 network eligible genes show enrichment of cell death, cell cycle and DNA repair functions, Focus molecules (bold) and up-regulated (red) or down-regulated (green) molecules are marked; B) The four top networks are interconnected and contain 273 of the 295 genes

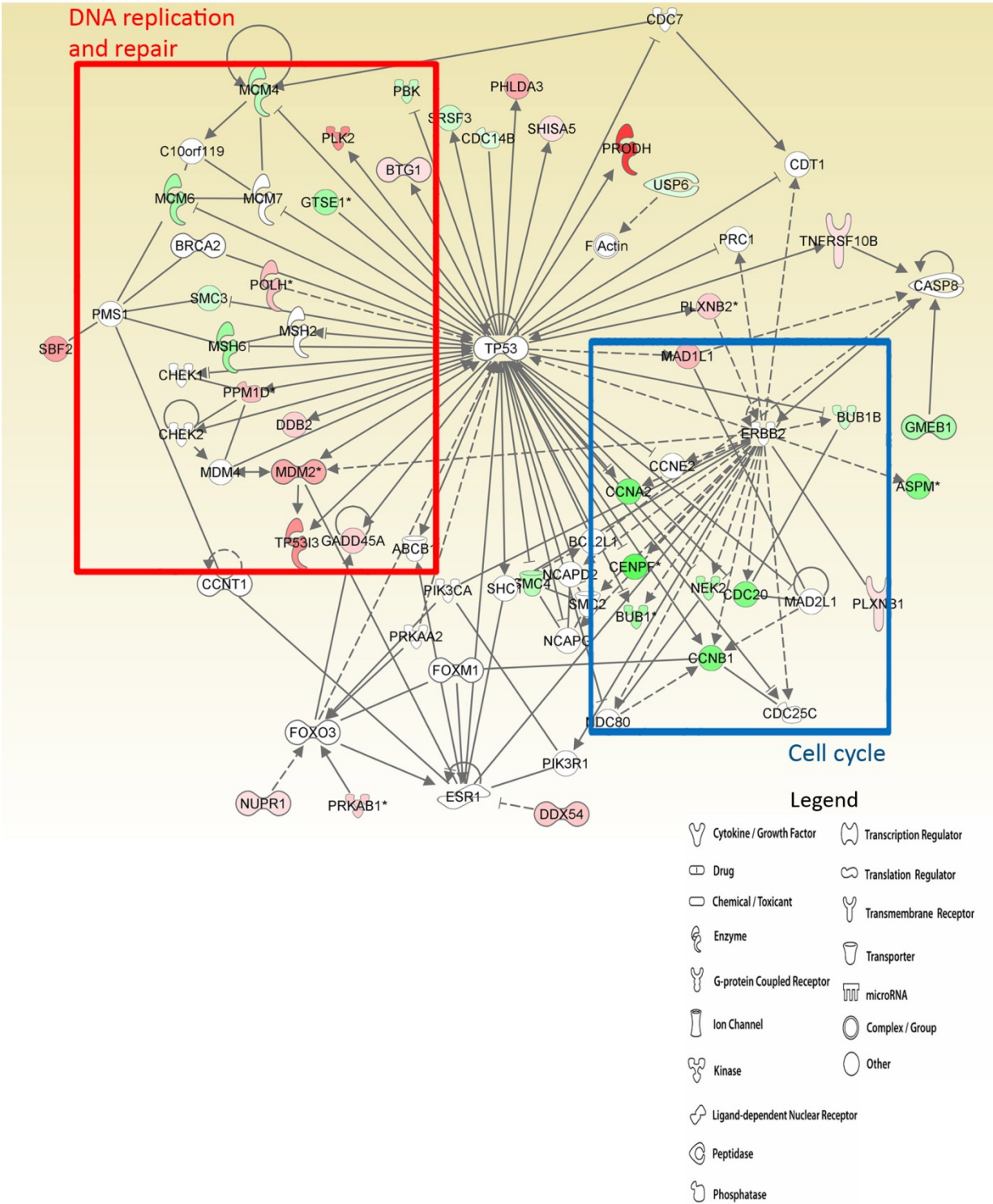


Figure 3.19: A p53-centric network with DNA replication, repair and cell cycle genes

A sub-network within the large interconnected network showing a p53-centric network consisting of up-regulated (red) genes mainly involved in DNA repair and the down-regulated genes (green) mainly involved in DNA replication and cell cycle.

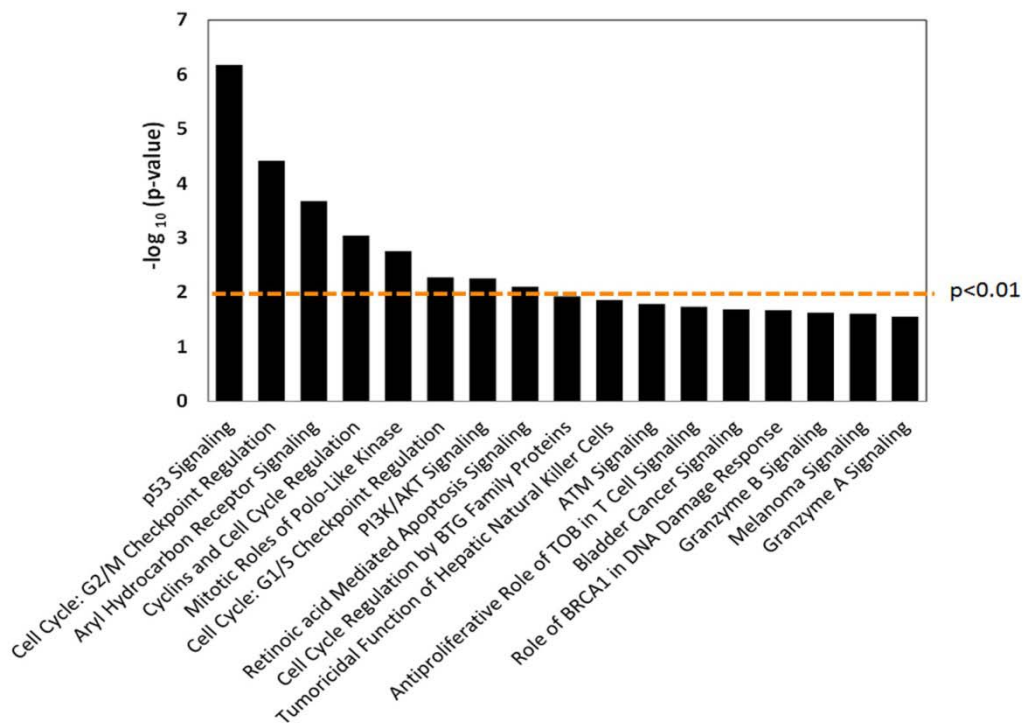


Figure 3.20: Canonical pathways significantly enriched in the transcriptional signature
 Canonical pathways defined in the Ingenuity database that are enriched within the 507 eligible genes from the transcriptional signature set.

3.4.11 The transcriptional signature can be partitioned into two functionally meaningful subsets

Gene ontology enrichment was performed separately on the set of up-regulated and down-regulated genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6.7) as described in the methods section. The Gene Ontology (GO) biological processes enriched within the up-regulated and down-regulated genes in the transcriptional signature are shown in Figure 3.21. The GO enrichment for the up-regulated genes specifically identified apoptotic cell death rather than any other cell death mechanism. Similarly, the repressed subset of genes showed enrichment in genes involved in the mitotic phase of the cell cycle.

Network analyses were performed separately for the induced and repressed gene sets. The resulting canonical pathway enrichment is shown in Figure 3.22. p53 signaling is the most significantly enriched canonical pathway for the induced genes, followed by enrichment in various cancer signaling pathways. Cell cycle as well as protein and RNA metabolism related canonical pathways are enriched in the repressed genes. It is interesting to note that the canonical pathways enriched within the combined set of up and down-regulated genes were distributed among, rather than appearing in both the induced and repressed gene sets.

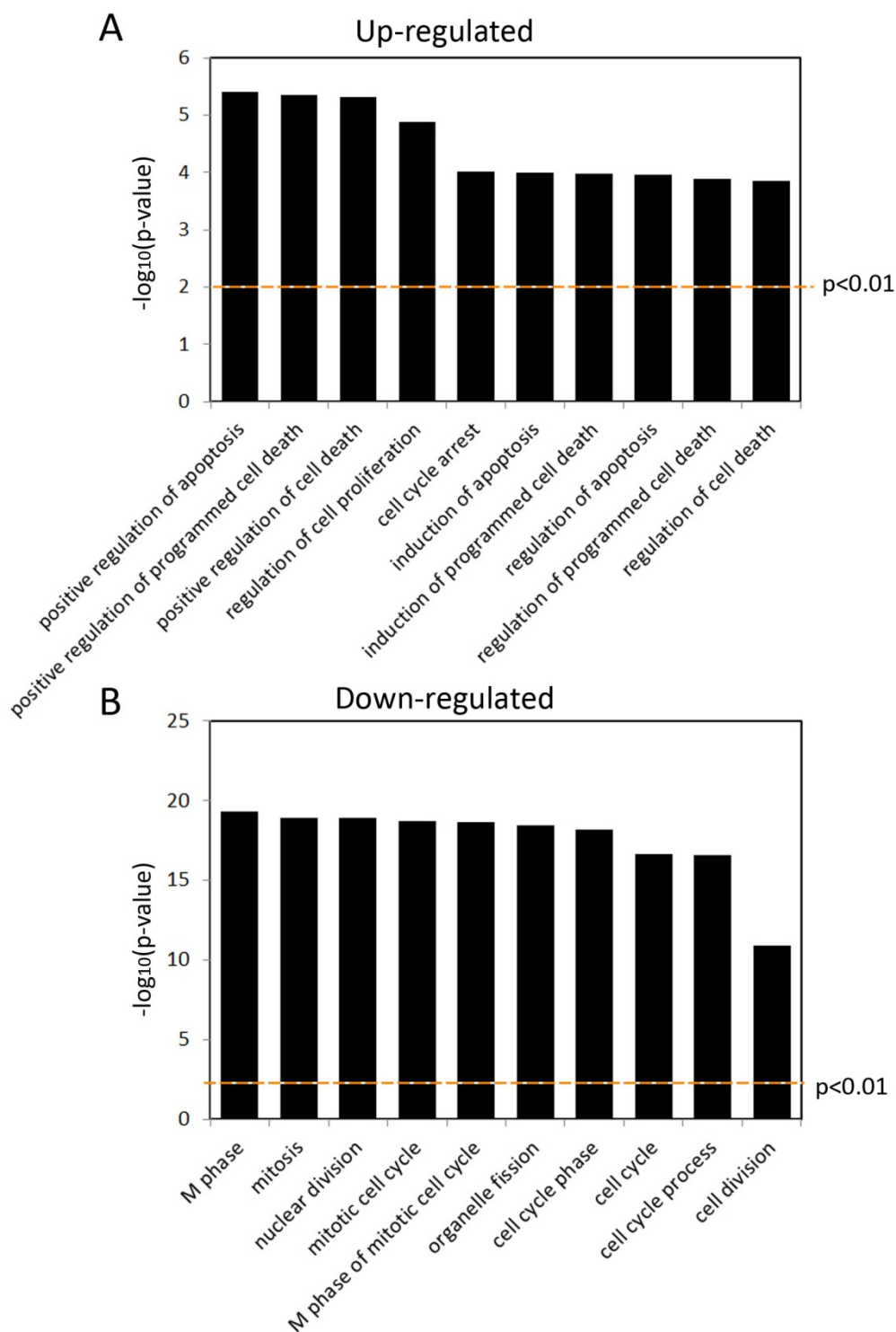


Figure 3.21: Gene Ontology enrichment of the up-regulated and down-regulated gene sets

GO enrichment for the up- and down-regulated genes from the transcriptional signature performed using DAVID(v.6.7) and the standard GO-FAT algorithm which highlights lower level terms present within the data set..

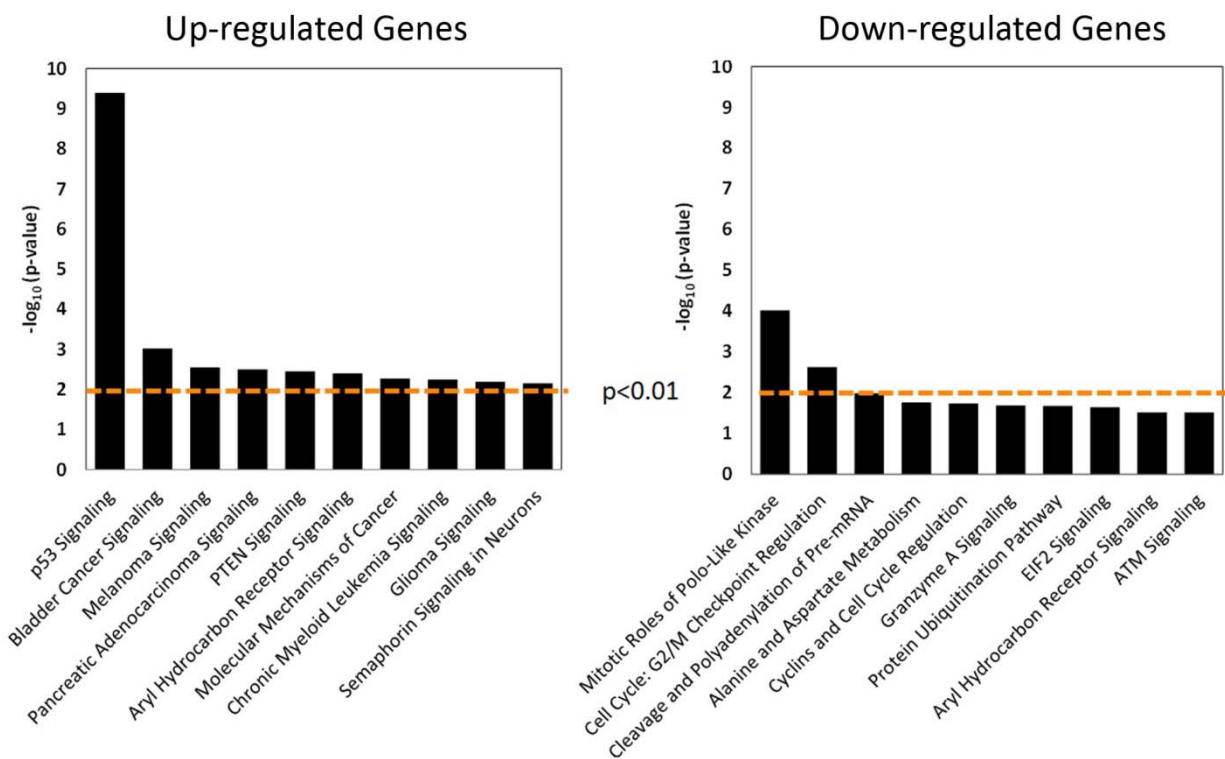


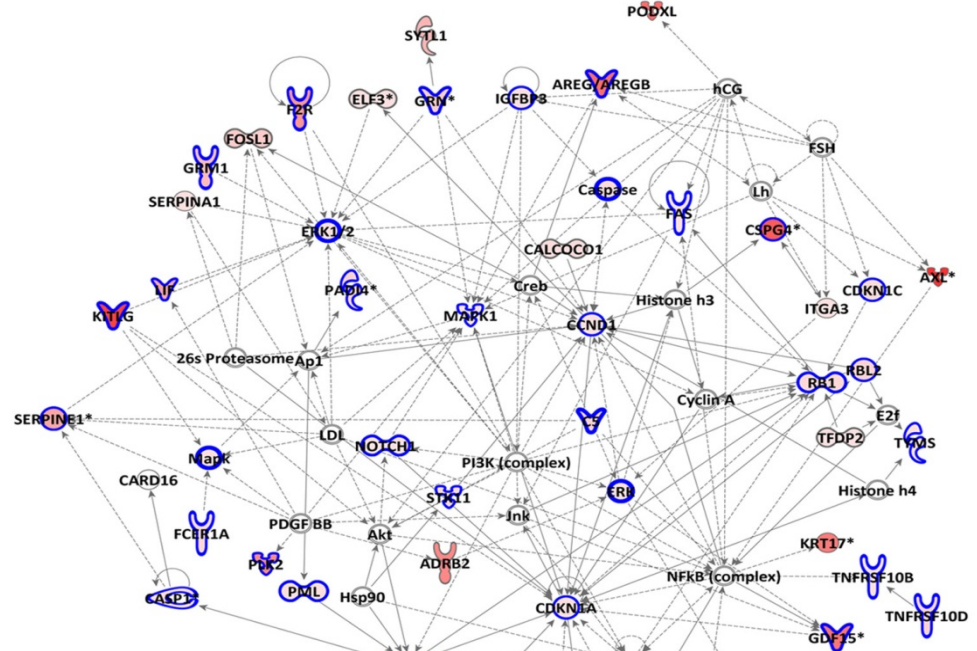
Figure 3.22: Canonical pathway enrichment of the up-regulated and down-regulated gene sets
 Canonical pathway enrichment performed separately on the up-regulated and down-regulated genes from the transcriptional signature.

Two of the top networks obtained for the up-regulated genes are shown in Figure 3.23. The top function associated with these networks was cell death ($p < 5 \times 10^{-10}$) in concordance with GO enrichment for the up-regulated genes. The nodes within the sub-networks shown in Figure 3.23 that are annotated in the Ingenuity database as being involved in cell death are highlighted in blue. These sub-networks show that 35/48 focus molecules in Figure 3.23A and 16/31 focus molecules in Figure 3.23B are involved in cell death. The network shown in Figure 3.23A is also enriched for cell growth and proliferation ($p < 5 \times 10^{-10}$; 26/48 focus molecules) and cellular development ($p < 5 \times 10^{-10}$; 23/48 focus molecules). For the network shown in Figure 3.23B no other functions besides cell death are enriched past the significance cutoff. In these networks, nodes are colored based on their expression value in cell line 4, with red nodes showing genes that are induced, with color intensity representing expression value. (Table 3.3 corresponds to the networks shown in Figure 3.23)

Figure 3.24 shows two of the top sub-networks obtained for the down-regulated genes. These two sub-networks are significantly enriched in cell cycle processes ($p < 5 \times 10^{-10}$), again in concordance with GO enrichment. Nodes that are associated with cell cycle functions in the Ingenuity database are highlighted in blue. There are 18/40 focus molecules in Figure 3.24A and 13/33 focus molecules in Figure 3.24B that are involved in cell cycle functions. The network in Figure 3.24B is also enriched for cellular assembly and organization (7/33 focus molecules), and DNA replication, recombination, and repair (7/33 focus molecules). Again the network is colored based on the expression values of the nodes in cell line 4. (Table 3.4 corresponds to the networks shown in Figure 3.24)

Gene Ontology enrichment, network connectivity and canonical pathway analysis highlight the fact that the transcriptional signature identified here can be split into two coherent groups; a group of genes differentially up-regulated in BCNU treated sensitive cell lines compared to resistant cell lines that are suggested to be involved in cell death functions and p53 signaling; and a second group of genes differentially down-regulated in the sensitive cell lines compared to the resistant cell lines that are suggested to be involved in cell cycle pathways.

A



B

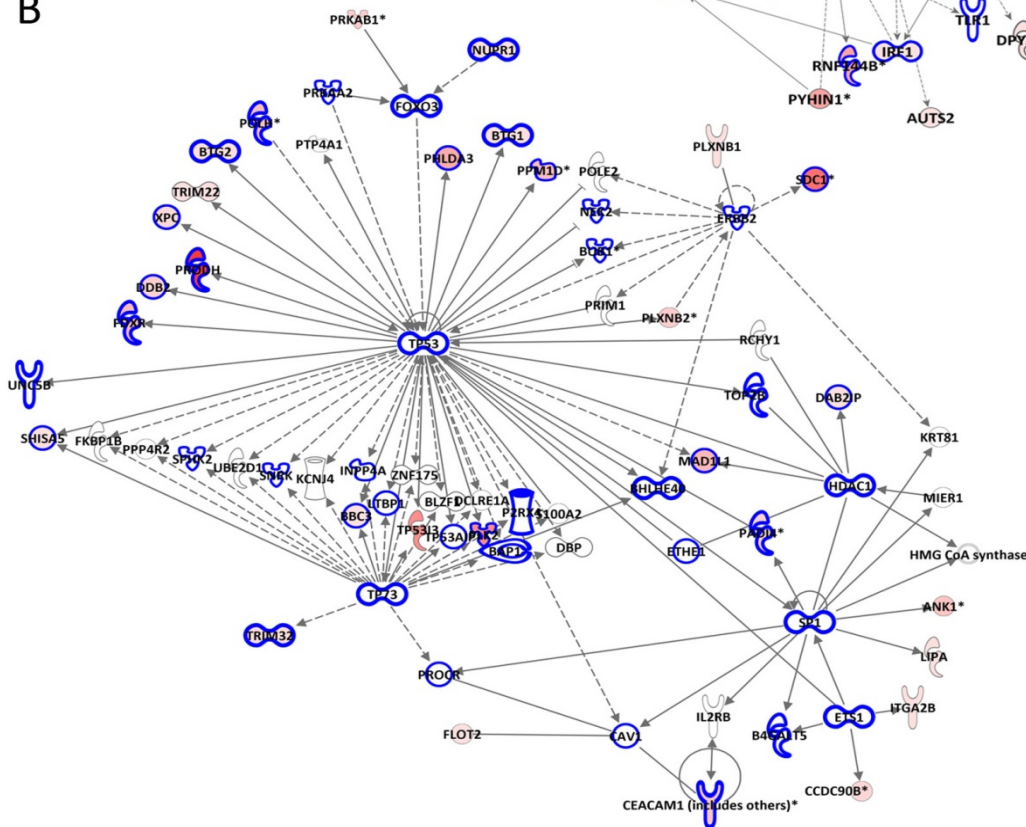


Figure 3.23: Top networks for the set of up-regulated genes are involved in cell death

Nodes highlighted in blue are annotated in the ingenuity database as being involved in cell death. Members of the up-regulated set are colored based on their expression value, red indicating up-regulation, and the intensity indicating expression value. The top network (A) and the second network (B) are shown (Legend same as for Figure 3.19)

Table 3.3: Molecules in network A and B from Figure 3.23 for the up-regulated gene set

ID	Molecules in Network	Score	Focus Molecules	Top Functions
A	26s Proteasome, ADRB2, Akt, Ap1, AREG/AREGB, AUTS2, AXL, C5, CALCOCO1, CARD16, CASP1, Caspase, CCND1, CDKN1A, CDKN1C, Creb, CSPG4, Cyclin A, DPYSL2, E2f, ELF3, ERK, ERK1/2, F2R, FAS, FCER1A, FGFR3, FOSL1, FSH, GDF15, GRM1, GRN, hCG, Histone h3, Histone h4, Hsp90, IGFBP3, Interferon alpha, IRF1, ITGA3, Jnk, KITLG, KRT17, LDL, Lh, LIF, Mapk, MAPK1, MDM2, NFkB (complex), NOTCH1, PADI4, PDGF BB, PI3K (complex), PLK2, PML, PODXL, PYHIN1, RB1, RBL2, RNF144B, SERPINA1, SERPINE1, STK11, SYTL1, TFDP2, TLR1, TNFRSF10B, TNFRSF10D, TYMS	61	48	Cell Death, Cellular Growth and Proliferation, Cellular Development
B	ANK1, B4GALT5, BAP1, BBC3, BHLHE40, BLZF1, BTG1, BTG2, BUB1, CAV1, CCDC90B, CEACAM1 (includes others), DAB2IP, DBP, DLCRE1A, DDB2, ERBB2, ETHE1, ETS1, FDXR, FKBP1B, FLOT2, FOXO3, HDAC1, HMG CoA synthase, IL2RB, INPP4A, ITGA2B, KCNJ4, KRT81, LIPA, LTBP1, MAD1L1, MIER1, NEK2, NUPR1, P2RX4, PADI4, PHLDA3, PLK2, PLXNB1, PLXNB2, POLE2, POLH, PPM1D, PPP4R2, PRIM1, PRKAA2, PRKAB1, PROCR, PRODH, PTP4A1, RCHY1, S100A2, SDC1, SHISA5, SNRK, SP1, SPHK2, TOP2B, TP53, TP73, TP53AIP1, TP53I3, TRIM22, TRIM32, UBE2D1, UNC5B, XPC, ZNF175	31	31	Cell Death

Focus molecules are in bold. Significantly ($p < 5 \times 10^{-10}$) enriched functions are labeled.

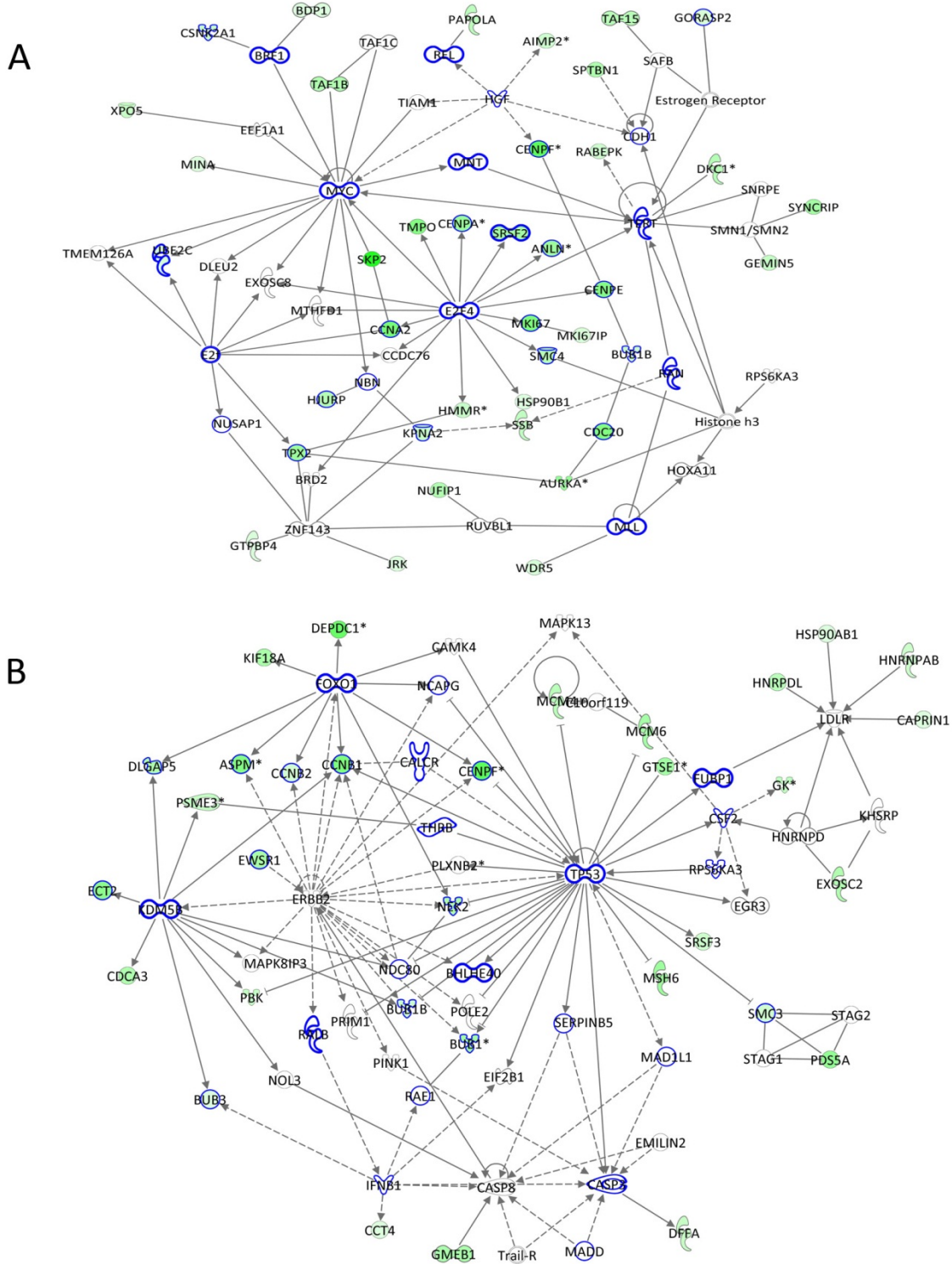


Figure 3.24: Top two networks for the down-regulated genes are involved in cell cycle
 Nodes highlighted in blue are annotated in the Ingenuity database as being involved in cell cycle. Members of the down-regulated set are colored based on their expression value, green indicating down-regulation, and the intensity indicating expression value. The top network (A) and the second network (B) are shown (Legend same as for Figure 3.19)

Table 3.4: Molecules in network A and B from Figure 3.24 for the down-regulated gene set

ID	Molecules in Network	Score	Focus Mo	Top Functions
A	AIMP2, ANLN, AURKA, BDP1, BRD2, BRF1, BUB1B, CCDC76, CCNA2, CDC20, CDH1, CENPA, CENPE, CENPF, CSNK2A1, DKC1, DLEU2, E2f, E2F4, EEF1A1, Estrogen Receptor, EXOSC8, GEMIN5, GORASP2, GTPBP4, HGF, Histone h3, HJURP, HMMR, HOXA11, HSP90B1, JRK, KPNA2, MINA, MKI67, MKI67IP, MLL, MNT, MTHFD1, MYC, NBN, NUFIP1, NUSAP1, PAPOLA, RABEPK, RAN, REL, RPS6KA3, RUVBL1, SAFB, SKP2, SMC4, SMN1/SMN2, SNRPE, SPTBN1, SRSF2, SSB, SYNCRIP, TAF15, TAF1B, TAF1C, TERT, TIAM1, TMEM126A, TMPO, TPX2, UBE2C, WDR5, XPO5, ZNF143	54	40	Cell Cycle
B	ASPM, BHLHE40, BUB1, BUB3, BUB1B, C10orf119, CALCR, CAMK4, CAPRIN1, CASP3, CASP8, CCNB1, CCNB2, CCT4, CDCA3, CENPF, CSF2, DEPDC1, DFFA, DLGAP5, ECT2, EGR3, EIF2B1, EMILIN2, ERBB2, EWSR1, EXOSC2, FOXO1, FUBP1, GK, GMEB1, GTSE1, HNRNPAB, HNRNPD, HNRPDL, HSP90AB1, IFNB1, KDM5B, KHSRP, KIF18A, LDLR, MAD1L1, MADD, MAPK13, MAPK8IP3, MCM4, MCM6, MSH6, NCAPG, NDC80, NEK2, NOL3, PBK, PDS5A, PINK1, PLXNB2, POLE2, PRIM1, PSME3, RAE1, RALB, RPS6KA3, SERPINB5, SMC3, SRSF3, STAG1, STAG2, THRB, TP53, Trail-R	41	33	Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair

Focus molecules are in bold. Significantly ($p < 5 \times 10^{-10}$) enriched functions are labeled.

3.4.12 BCNU sensitive cell lines induce cell death by apoptosis

The functional enrichment of the transcriptional signature showed a strong induction of genes involved in apoptosis in the sensitive cell lines 4 and 5 upon BCNU treatment. Previous studies showed that BCNU induces growth inhibition and cell killing in glioblastoma and other cancerous cell lines (24,28,29). The few studies that explored BCNU induced cell death mechanisms in human cells showed that some glioblastoma cell lines induce apoptosis upon BCNU treatment (30-32) whereas in other cases BCNU had a caspase inhibitory activity (33). Additionally, in the lymphoblastoid cell line TK6, BCNU was seen to induce apoptosis (34).

Although we knew that BCNU induced strong growth inhibition in cell lines 4 and 5 as seen from the proliferation assay and viability measurements (Figure 3.8 and Figure 3.15), we did not know the mechanisms of cell death induction in these particular cell lines. However, the induction of apoptotic genes within our transcription data, combined with previous evidence that BCNU induced cell death by apoptosis in TK6 cell lines suggested that the sensitive cell lines 4 and 5 most likely induce apoptosis when exposed to BCNU.

To confirm this, we used the AnnexinV/7AAD assay to measure cell death in cell lines 4, 5, 16 and 13 after BCNU treatment. During early apoptosis, phosphatidylserine (PS), usually located in the cytoplasmic side of the cell membrane, is flipped and exposed on the external surface of the cell (35). AnnexinV can bind PS with high affinity (36). Therefore, AnnexinV tagged with a fluorophore (such as phycoerythrin –PE) can be used to detect cells that have initiated apoptosis by flow cytometry. However, PE-AnnexinV can also permeate and stain cells that are already dead and that have compromised cell membrane integrity. Therefore, to distinguish AnnexinV

staining in early apoptotic versus dead cells, an additional dye, 7AAD, that only stains dead cells with compromised cell membranes is used. This yields flow cytometry plots as seen in Figure 3.25 with cells falling within one of four quadrants – i) the lower left quadrant contains AnnexinV and 7AAD negative live cells with intact cell membranes and PS on the cytoplasmic side of the membrane; ii) the lower right quadrant contains AnnexinV positive and 7AAD negative early apoptotic cells with exposed PS but intact membranes; iii) the upper right quadrant contains AnnexinV and 7AAD positive dead cells that have lost cell membrane integrity and iv) the upper left quadrant contains AnnexinV negative and 7AAD positive cells that induce cell death by mechanisms other than apoptosis, or dead cells that have lost all PS.

The four cell lines with extreme BCNU sensitivity/resistance were treated with 40 μ M BCNU as described in the methods section. At each of 12, 24, 48 and 72 hours, cells were collected, incubated with 7AAD and PE-AnnexinV (see Materials and Methods) and immediately analyzed by flow cytometry. Figure 3.25 shows an example of plots obtained for the four cell lines at 48hours post BCNU treatment. We see that there is a visible increase in the early apoptotic population after BCNU treatment in the two sensitive cells lines as compared to the resistant cells lines. Figure 3.26 shows the quantified data from all time-points for the four quadrants. The data show a significant increase, over time, in the early apoptotic and dead cell populations and a significant decrease in the live cell population for the two sensitive cell lines. Note that we do not see direct migration of cells from live cell quadrant to the upper left quadrant thus confirming that the major path to cell death after BCNU treatment occurs through apoptosis. Another point to note is that among the two sensitive cell lines, the extent of apoptosis and cell death is greater in the most sensitive cell line 4 and lesser in the cell line 5, as is expected from the survival data.

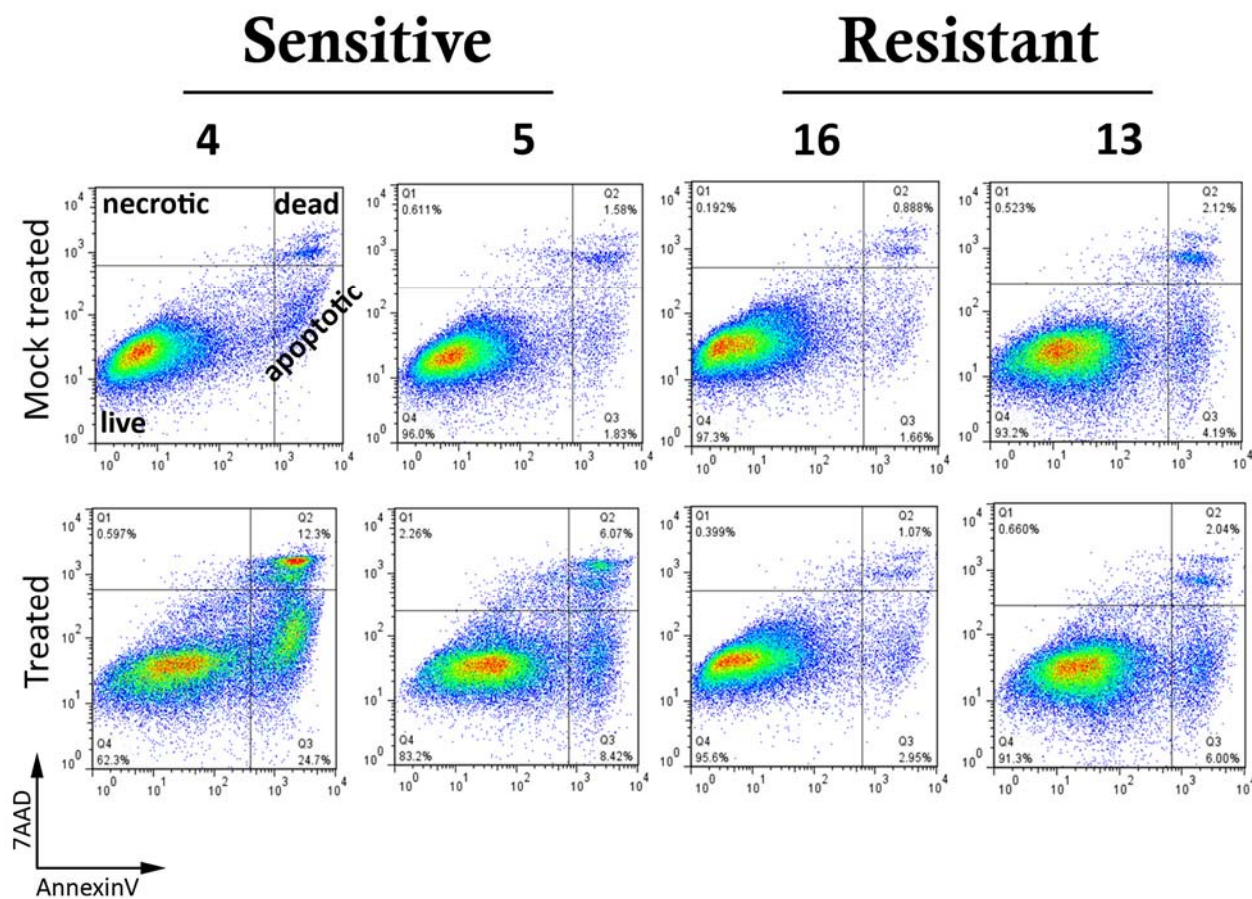


Figure 3.25: Annexin V/7AAD plots for the four cell lines at 48h post BCNU treatment

Flow cytometry scatter plots of Annexin V vs. 7AAD for cell lines 4, 5, 16 and 13 at 48h post BCNU treatment. The top panel shows mock-treated samples and the bottom panel shows BCNU treated samples. The four quadrants are marked as live, apoptotic, dead and necrotic as described in the text.

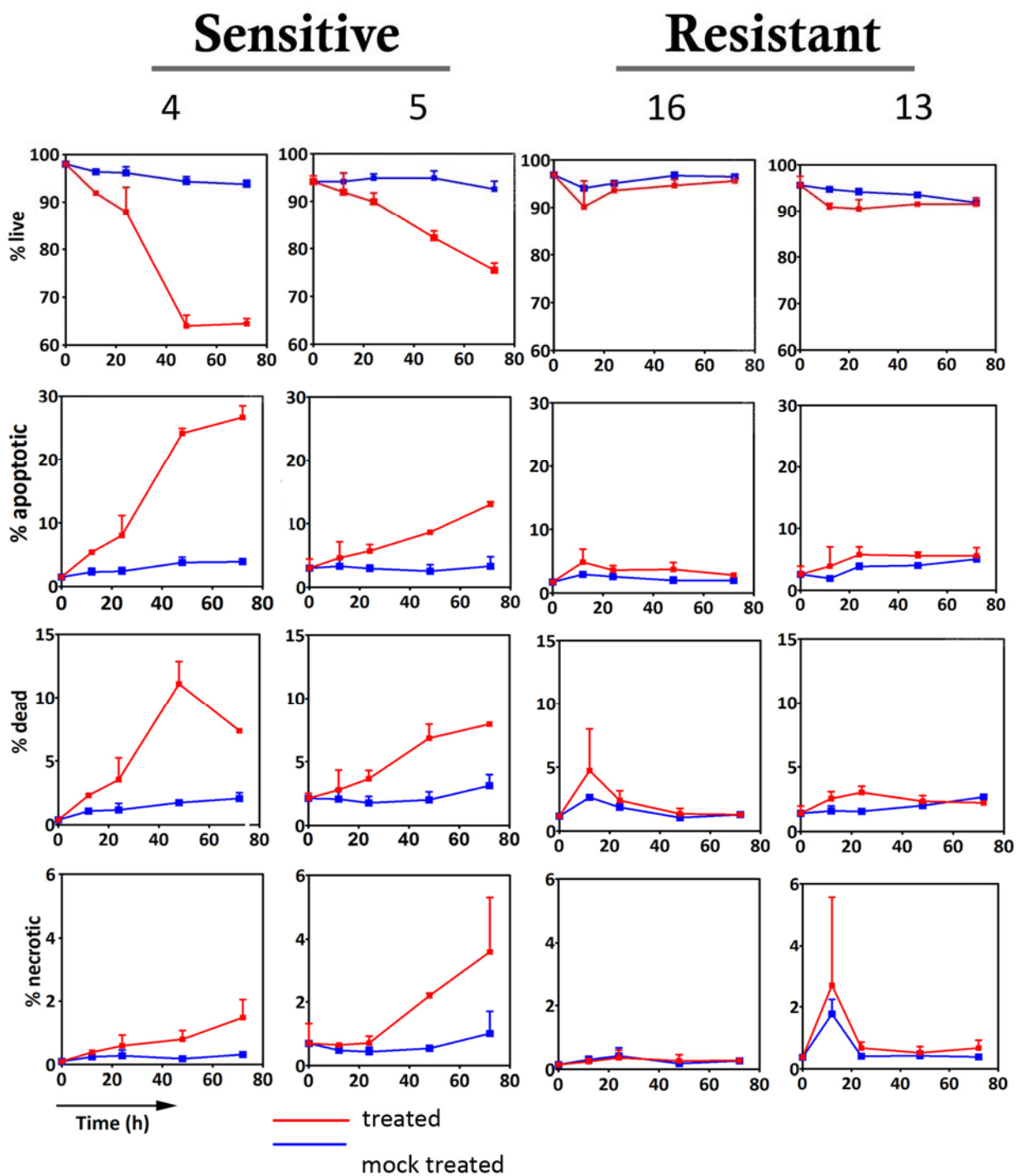


Figure 3.26: Quantification from AnnexinV/7AAD plots at multiple times post BCNU treatment
Plots of live, apoptotic, dead and necrotic populations as quantified from flow cytometry scatter plots at 0, 12, 24, 48 and 72 hours post BCNU treatment.

3.4.13 BCNU sensitive cell lines stall mitotic entry

Cell cycle profiles obtained from initial experiments (Figure 3.11 and Figure 3.11) had shown sensitive cells accumulating in late S more so than in the G2 phase of the cell cycle after BCNU treatment. Genes involved in mitosis are known to slowly accumulate during S phase of the cell cycle and peak in the G2 phase of the cell cycle immediately prior to mitosis. Therefore, although treated cells had a higher fraction of S-phase cells, and at least an equal fraction of G2 cells as compared to mock-treated samples, we observed a decrease in the mitotic genes. This suggested that BCNU treated cells failed to enter mitosis.

To confirm that BCNU treated cells failed to enter mitosis after BCNU treatment, we used flow cytometry based detection of histone H3 phosphorylation as a marker for mitotic cells (37). Histone H3 is phosphorylated during early mitosis (38). Therefore using an Alexafluor conjugated antibody against phosphorylated histone H3, we can detect the fraction of cells undergoing mitosis after BCNU treatment by flow cytometry. For this assay, cells were treated with 40 μ M BCNU, collected at 12, 24, 48 and 72 hours post BCNU treatment and fixed in paraformaldehyde and stored at -20⁰ C. Before analysis by flow cytometry, cells were incubated with anti-phosphorylated Histone H3 antibody and propidium iodide (see Materials and Methods) and analyzed by flow cytometry.

Figure 3.27 shows a snap shot of mitotic cells at 48hours post BCNU treatment where the mitotic population is almost completely absent in the sensitive cell lines but not in the resistant cell lines compared to mock-treated samples. Figure 3.28 shows the quantified mitotic fraction over time for the four cell lines. The drop in the mitotic population seen in sensitive cell lines occurs prior

to the 12 hour time point and remained low up to the 72 hour time-point. However no significant changes were seen in the resistant cell lines. These results show that the transcriptional down-regulation of mitotic genes correlates with a concomitant decrease in the mitotic population in treated, sensitive cell lines.

3.4.14 p53 is activated to a greater extent in sensitive cell lines as compared to resistant cell lines

Canonical pathway analysis showed that the gene set that was induced in BCNU treated, sensitive cell lines was enriched for p53 signaling. p53 is known to be activated after BCNU treatment in glioblastoma cell lines and p53 status affects the sensitivity of cells to BCNU (24,39,40). The standard mode of p53 activation involves its phosphorylation and accumulation which we expected to see happening to a greater extent in the two sensitive cell lines than in the resistant cell lines. To test this, we used immunoblot analysis with antibodies against total and phosphorylated p53 at serine 20 and serine15. Cells were treated with 40 μ M BCNU as described in the Materials and Methods section and cells were collected at 12, 24, 48 and 72 hours post BCNU treatment, for immunoblot analysis. Whole cell lysate was probed with antibodies against total p53 and p53 phosphorylated at serine 20 or serine 15 (see Materials and Methods). Figure 3.29 and Figure 3.30 shows the immunoblots and their quantified values respectively, for total and phosphorylated p53 (ser20) levels in sensitive and resistant cell lines in both the mock-treated and BCNU treated samples. Figure 3.31 and Figure 3.32 show the same for phosphorylated p53 (ser15). There is a significant increase in total and phosphorylated p53 for the most sensitive cell line 4. The other cell lines also show a similar trend of p53 activation, but not with statistical significance.

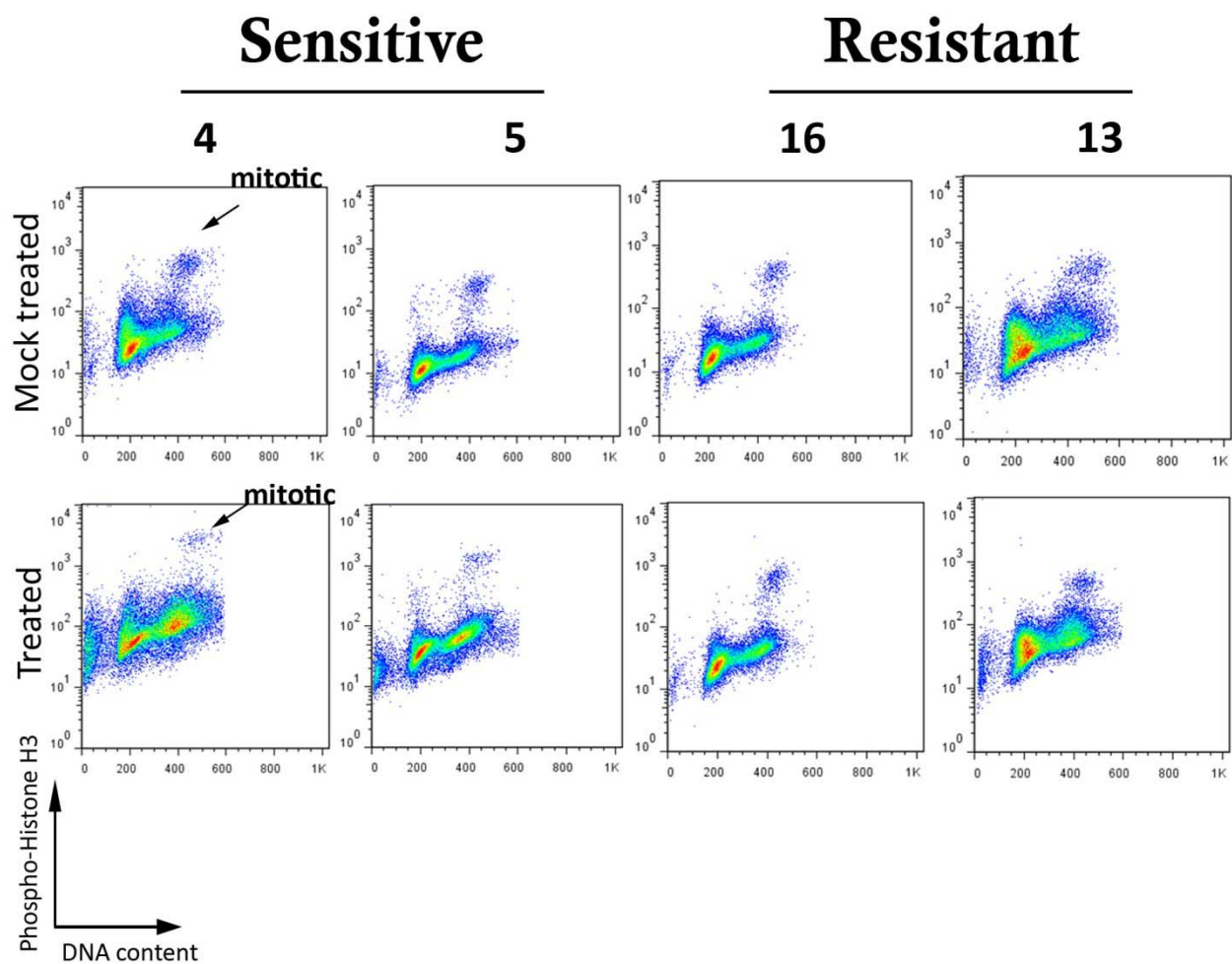


Figure 3.27: Phospho-histone H3 at 48h post BCNU treatment

Flow cytometry plots showing DNA content vs. Phospho-histone H3 for cell lines 4, 5, 13 and 16 at 48 hours post BCNU treatment. The population representing mitotic cells is highlighted.

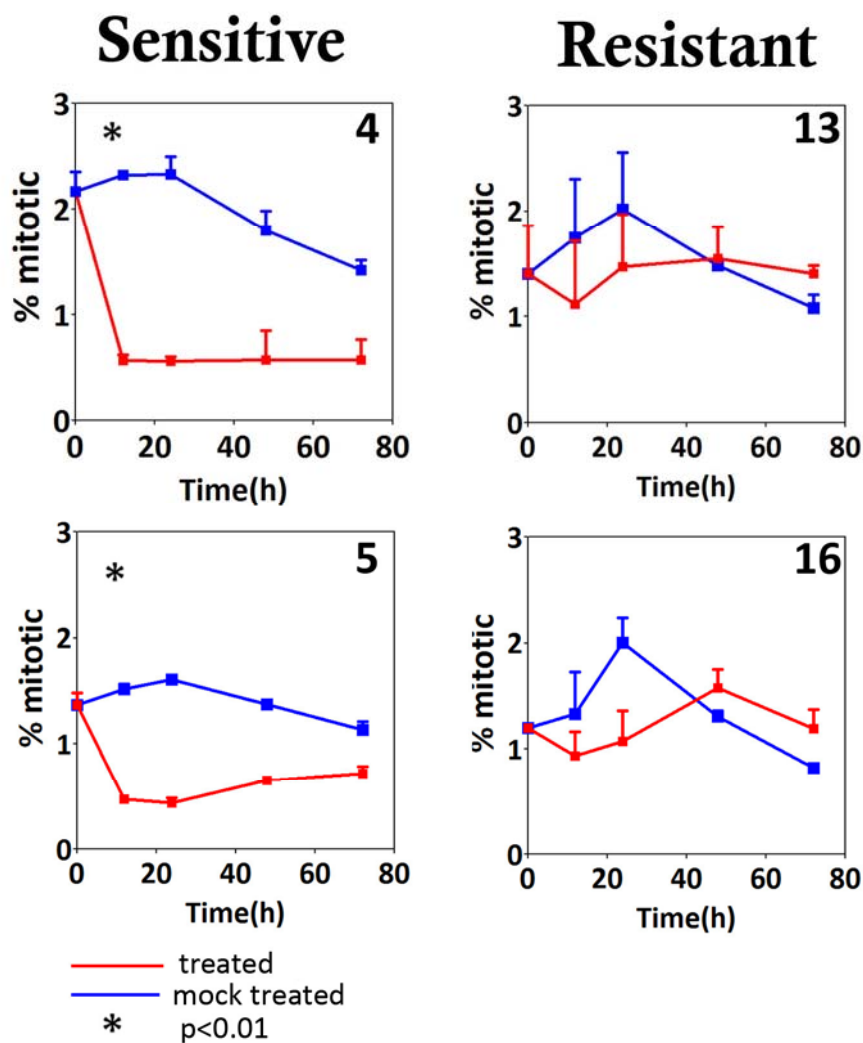


Figure 3.28: Quantified mitotic fraction at multiple time-points post BCNU treatment

Sensitive cell lines show a significant drop in the mitotic population after BCNU treatment as compared to the resistant cell lines.

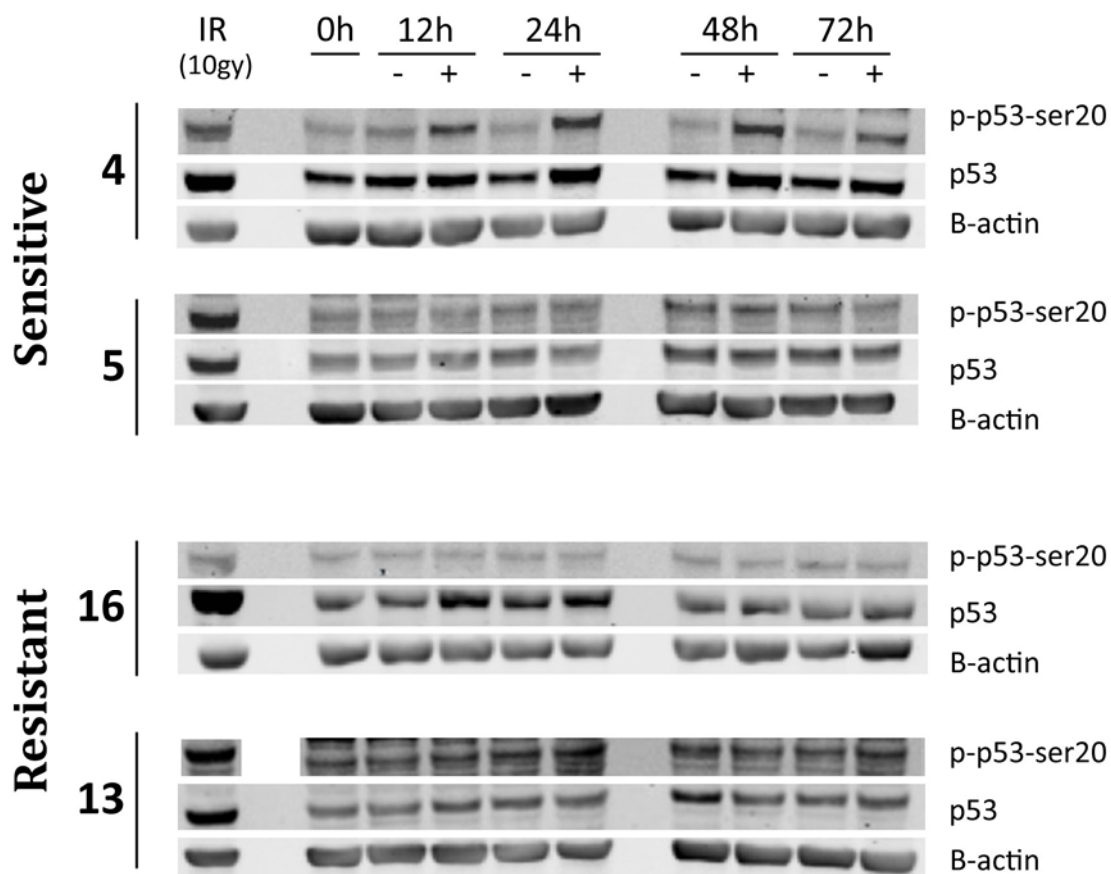


Figure 3.29: Immunoblots probed for phosphor-p53 (ser20)
 Immunoblots for BCNU and mock treated cells from cell lines 4, 5, 16 and 13 probed with antibodies against phosphorylated p53 (ser 20), p53 and B-actin (loading control).

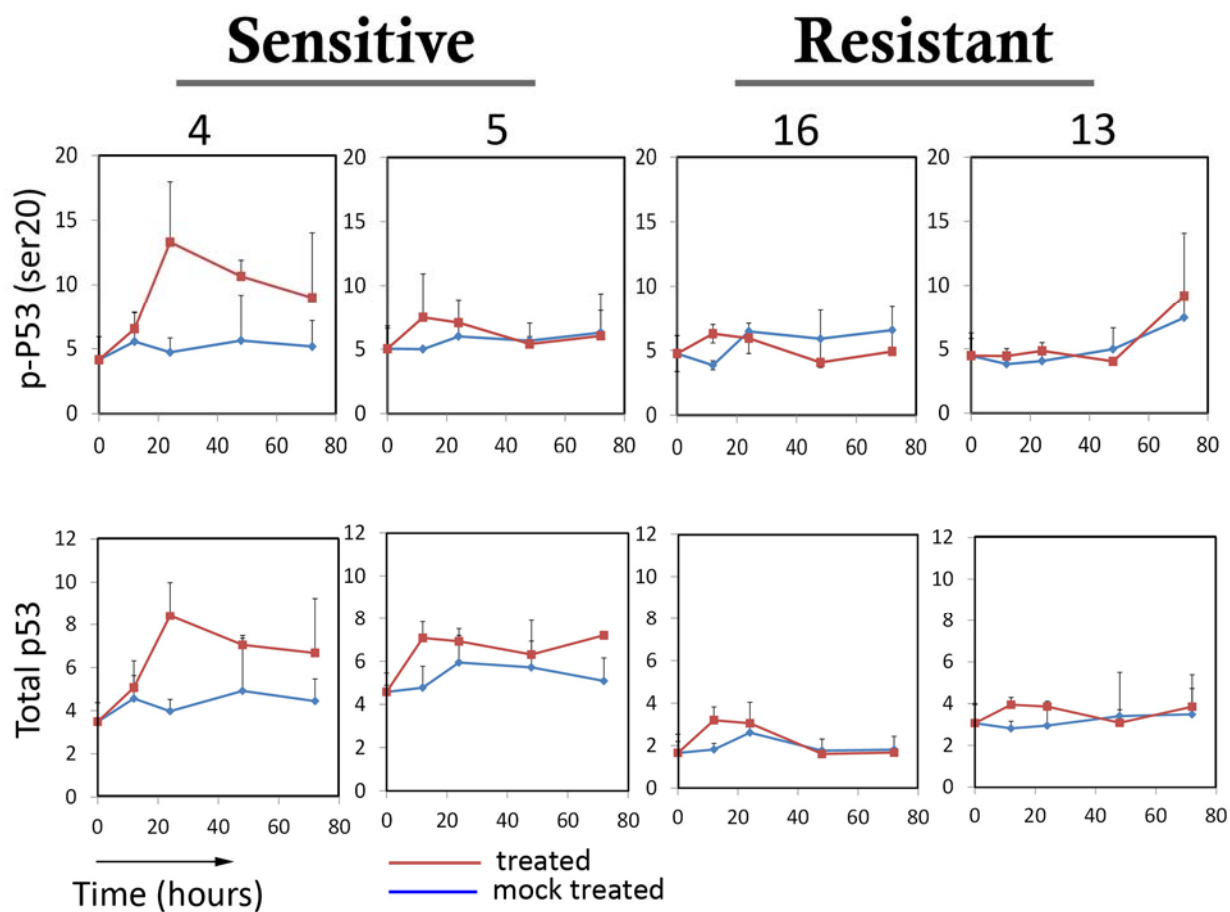


Figure 3.30: Total and phosphorylated p53 (Serine 20) as measured by immunoblot for BCNU treated samples

Quantitated data from immunoblot analysis of mock-treated and treated samples at 0, 12, 24, 48 and 72 hours post BCNU treatment for cell lines 4, 5, 16 and 13 using antibodies against total p53 and p53 phosphorylated at Serine 20. All values were normalized to b-actin as a loading control and scaled to a positive IR control.

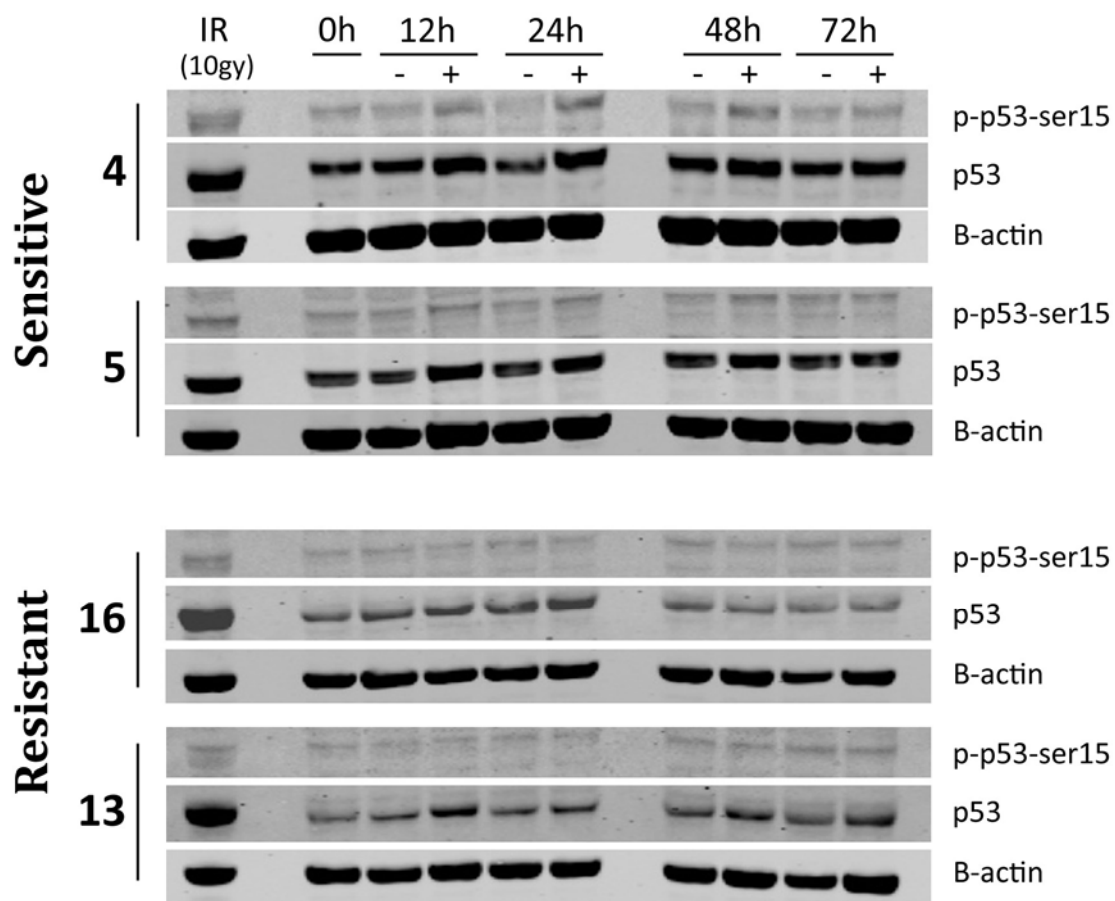


Figure 3.31: Immunoblots probed for phospho-p53 (ser15)

Immunoblots for BCNU and mock treated cells from cell lines 4, 5, 16 and 13 probed with antibodies against phosphorylated p53 (ser 15), p53 and B-actin (loading control).

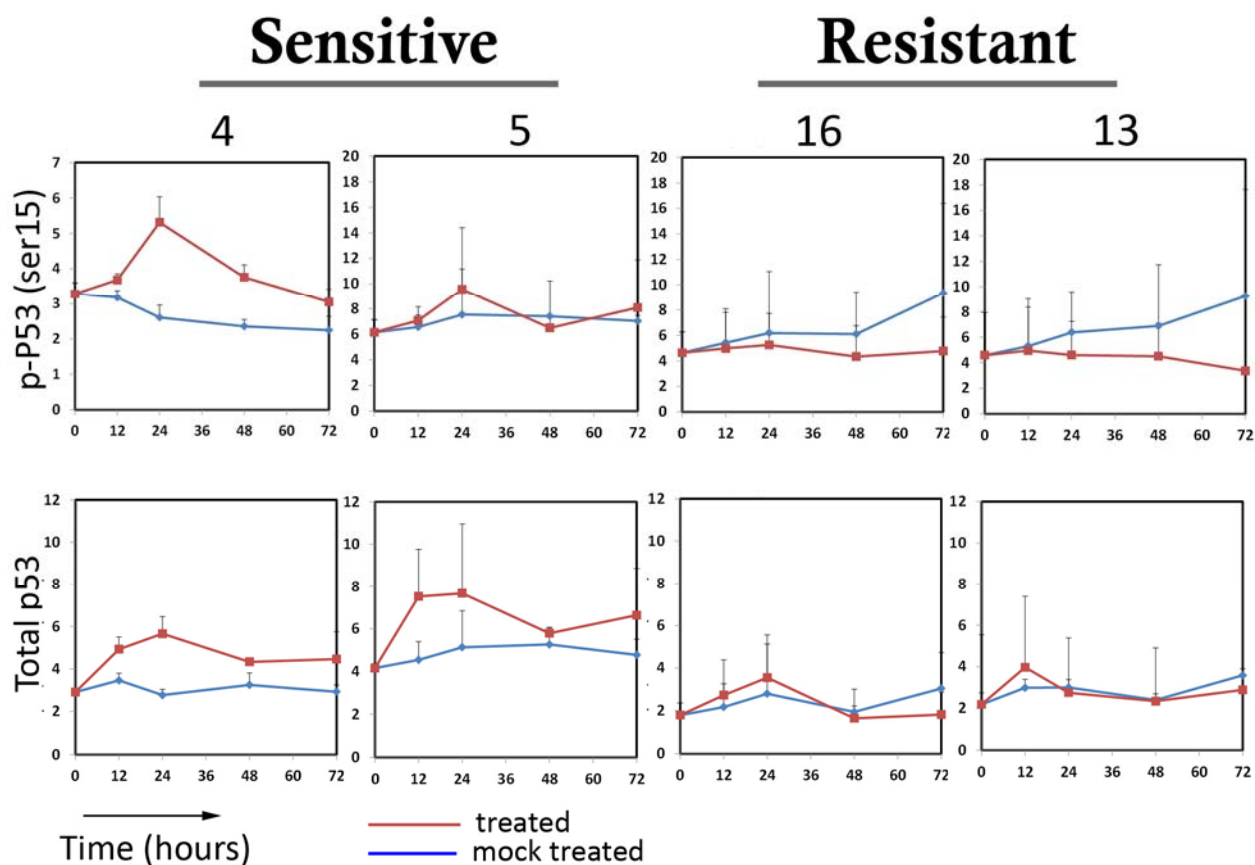


Figure 3.32: Total and phosphorylated p53 (Serine 15) as measured by immunoblot for BCNU treated samples

Quantitated data from immunoblot analysis of mock-treated and treated samples at 0, 12, 24, 48 and 72 hours post BCNU treatment for cell lines 4, 5, 16 and 13 using antibodies against total p53 and p53 phosphorylated at Serine 15. All values were normalized to b-actin as a loading control and scaled to a positive IR control.

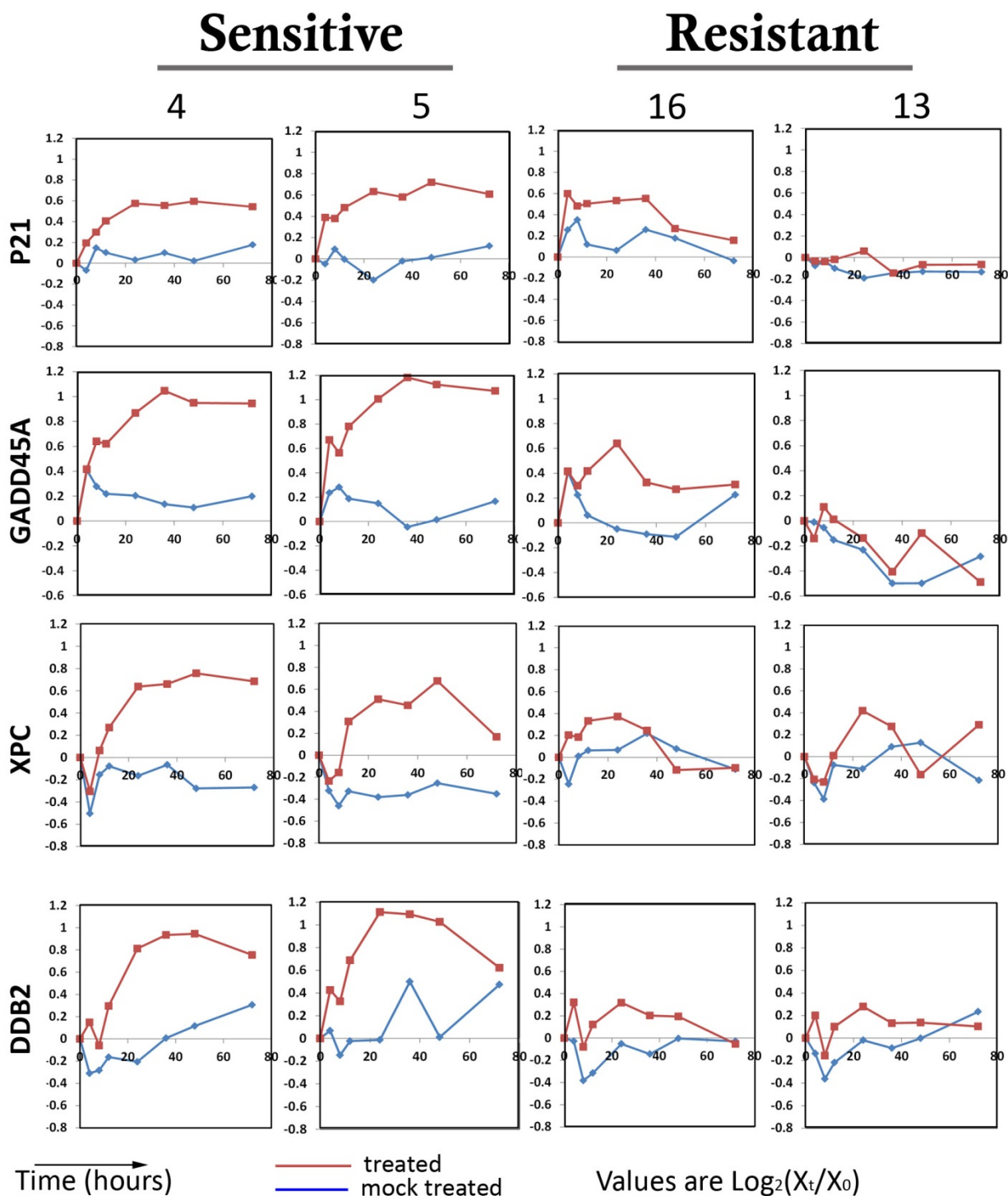


Figure 3.33: Expression of canonical p53 targets from transcriptional profiles

Expression values of canonical p53 targets in cell lines 4, 5, 16 and 13 in mock-treated (blue) and treated (red) samples at 0, 4, 8, 12, 24, 36, 48 and 72hours post BCNU treatment from microarray measurements. Values are \log_2 ratios of the expression at each time point compared to the zero hour time point.

To confirm p53 activation in the cell lines, we also looked at the gene expression of canonical p53 targets in the four cell lines from the microarray data set. Figure 3.33 shows induction of *P21*, *GADD45A*, *XPC* and *DDB2*, all known to be canonical targets of p53. Therefore, although only moderate p53 accumulation and phosphorylation was observed after BCNU treatment in the sensitive cell lines, this level of activation seems to be sufficient to activate expression of p53 targets.

3.4.15 Computational promoter sequence analysis shows that a subset of down-regulated genes are enriched for NF-Y binding motifs and predicted to be novel NF-Y targets

We used the EXPANDER (9-12) package and its built-in transcription factor enrichment platform PRIMA to identify transcription factors with motif matches that are enriched within the promoter regions (-1000 to 200bp region around the transcription start site) of either the up-regulated or down-regulated genes (see Materials and Methods section). As shown in Figure 3.34, the down-regulated genes showed significant enrichment for binding sites for four transcription factors- Nuclear Factor Y (NF-Y), Aryl Hydrocarbon Receptor (AhR), GA binding protein (GABP) and Nuclear respiratory factor 1 (Nrf-1). The genes whose promoter regions have binding motifs for these transcription factors are shown in Table 3.5, Table 3.6, Table 3.7 and Table 3.8. On the other hand, within the -1000 to 200bp region around the transcription start site, there were no significant transcription factors identified for the up-regulated genes.

To check whether the up-regulated genes contained transcription factor binding motifs outside of the -1000 to 200bp region, a larger range spanning the -3000 to 200bp region around the transcription start site (the maximum range allowed in EXPANDER) was searched for

transcription factor binding site enrichment. Table 3.9 shows the motifs enriched for various ranges of the promoter region analyzed. For the down-regulated genes, as the promoter range analyzed is increased, significance for enrichment is lost for Nrf-1 followed by GABP and finally AhR. The only motif enriched for all ranges tested is that of NF-Y, further increasing our confidence in the list of predicted NF-Y targets. For the up-regulated genes, performing the enrichment analyses for the -2000 to 200bp region as well as the -2500 to 200bp region yields enrichment in the p53 binding motif. This enrichment is no longer significant when the region is increased to the -3000 to 200bp region, or decreased to the -1500 to 200bp region. Searching the -2500 to -1500 bp region of the promoters of the up-regulated genes alone does not yield significant enrichment for p53 binding motifs, therefore, the region between -1500 to 200bp around the transcription start site is required for the enrichment of p53 motifs. The list of genes containing the p53 binding motif is shown in Table 3.10.

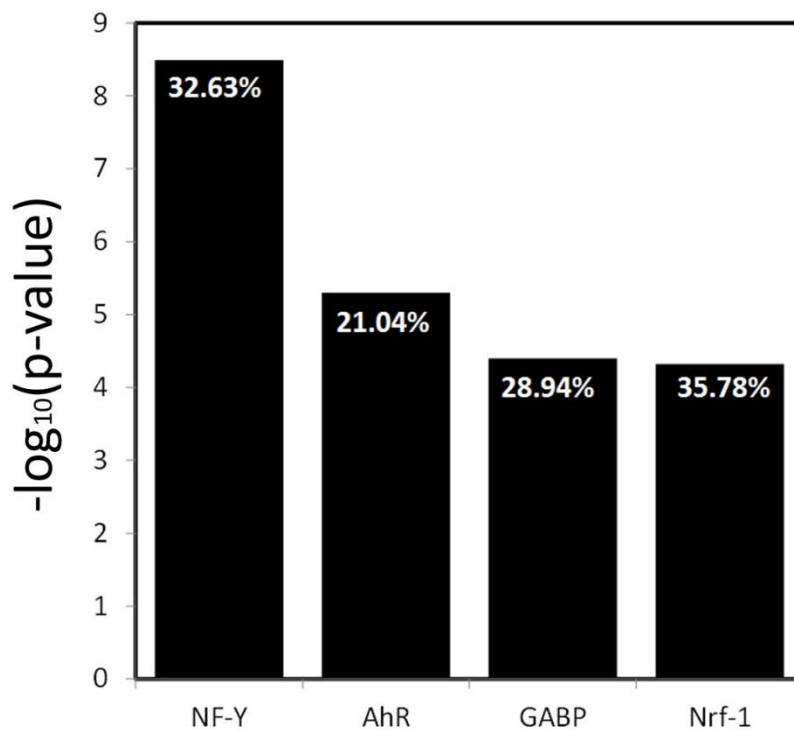


Figure 3.34: Transcription Factor Binding Site enrichment in the down-regulated cluster using EXPANDER

The figure shows significance of enrichment for the binding sites of four transcription factors within the promoter regions (-1000 to 200bp region around the transcription start site) of the down-regulated genes (190). The percentage of genes in the down-regulated gene-set also enriched for the binding site of each transcription factor is shown on each bar. NF-Y (Nuclear Factor Y), AhR (Aryl Hydrocarbon Receptor), GABP (GA binding protein) and Nrf-1 (Nuclear respiratory factor 1).

Table 3.5: Genes from the down-regulated gene set that are predicted to be NF-Y motif targets

Gene Symbol	Probe-set	Gene Name
AIMP2	202138_x_at	aminoacyl tRNA synthetase complex-interacting multifunctional protein 2
ARHGAP11A	204492_at	Rho GTPase activating protein 11A
ASPM	239002_at	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)
AURKA	204092_s_at	aurora kinase A; aurora kinase A pseudogene 1
BUB1	233445_at	budding uninhibited by benzimidazoles 1 homolog (yeast)
BUB3	201456_s_at	budding uninhibited by benzimidazoles 3 homolog (yeast)
C13orf34	219544_at	chromosome 13 open reading frame 34
C2orf67	1554791_a_at	chromosome 2 open reading frame 67
CCNA2	203418_at	cyclin A2
CCNB1	214710_s_at	cyclin B1
CCNB2	1560161_at	cyclin B2
CCNF	204826_at	cyclin F
CDC20	202870_s_at	cell division cycle 20 homolog (S. cerevisiae)
CDCA3	221436_s_at	cell division cycle associated 3
CENPA	204962_s_at	centromere protein A
CENPF	207828_s_at	centromere protein F, 350/400ka (mitosin)
DEPDC1	220295_x_at	DEP domain containing 1
DLGAP5	203764_at	discs, large (Drosophila) homolog-associated protein 5
ECT2	219787_s_at	epithelial cell transforming sequence 2 oncogene
EEF1E1	204905_s_at	eukaryotic translation elongation factor 1 epsilon 1
EIF4A1	201530_x_at	eukaryotic translation initiation factor 4A, isoform 1
EXOSC2	209527_at	exosome component 2
GK	207387_s_at	glycerol kinase
GTPBP4	218238_at	GTP binding protein 4
GTSE1	204315_s_at	G-2 and S-phase expressed 1
H2AFV	202487_s_at	H2A histone family, member V
HMMR	1562677_at	hyaluronan-mediated motility receptor (RHAMM)
HNRNPAB	201277_s_at	heterogeneous nuclear ribonucleoprotein A/E
HNRPDL	1554678_s_at	heterogeneous nuclear ribonucleoprotein D-like
HSP90B1	200598_s_at	heat shock protein 90kDa beta (Grp94), member 1
KIF14	206364_at	kinesin family member 14
KIF20A	218755_at	kinesin family member 20A
LTV1	225748_at	LTV1 homolog (S. cerevisiae)
MCM4	212141_at	minichromosome maintenance complex component 4
MCM6	201930_at	minichromosome maintenance complex component 6
METAP2	209861_s_at	methionyl aminopeptidase 2
MSH6	211449_at	mutS homolog 6 (E. coli)
NEK2	204641_at	NIMA (never in mitosis gene a)-related kinase 2
NF2	204991_s_at	neurofibromin 2 (merlin)
NUDCD2	1558124_at	NudC domain containing 2
ODF2	210415_s_at	outer dense fiber of sperm tails 2
PIP4K2B	1553047_at	phosphatidylinositol-5-phosphate 4-kinase, type II, beta
PKP4	201927_s_at	plakophilin 4
PLK1	202240_at	polo-like kinase 1 (Drosophila)
PRMT5	1564520_s_at	protein arginine methyltransferase 5
PSME3	200987_x_at	proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki)
PSRC1	201896_s_at	proline/serine-rich coiled-coil 1
SFRS2	200753_x_at	splicing factor, arginine/serine-rich 2
SGOL2	230165_at	shugoshin-like 2 (S. pombe)
SMC3	1556925_at	structural maintenance of chromosomes 3
STAG3L3	221191_at	stromal antigen 3-like 3
TAF15	202840_at	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa
TAF1B	214690_at	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa
TCEB1	202823_at	similar to elongin C; transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C
TCP1	208778_s_at	t-complex 1
THADA	1554492_at	thyroid adenoma associated
TMPO	203432_at	thymopoietin
TPX2	210052_s_at	TPX2, microtubule-associated, homolog (Xenopus laevis)
TUBB2C	208977_x_at	tubulin, beta 2C
UBE2C	202954_at	ubiquitin-conjugating enzyme E2C
WDR37	1554869_at	WD repeat domain 37
ZMYM4	202049_s_at	zinc finger, MYM-type 4

Genes highlighted in red are transcriptionally regulated by NF-Y and genes in blue have been shown to have NF-Y promoter occupancy (56,89-91).

Table 3.6: Genes from the down-regulated gene set that are enriched for the AhR motif

Gene ID	Gene Name
BUB1B	budding uninhibited by benzimidazoles 1 homolog beta (yeast)
C13ORF34	chromosome 13 open reading frame 34
CCT4	chaperonin containing TCP1, subunit 4 (delta)
CCT5	chaperonin containing TCP1, subunit 5 (epsilon)
CCT8P1	similar to chaperonin containing TCP1, subunit 8 (theta)
CDC20	cell division cycle 20 homolog (S. cerevisiae)
CDCA3	cell division cycle associated 3
DCUN1D5	DCN1, defective in cullin neddylation 1, domain containing 5 (S. cerevisiae)
DDAH1	dimethylarginine dimethylaminohydrolase 1
DKC1	dyskeratosis congenita 1, dyskerin
DLAT	dihydrolipoamide S-acetyltransferase
DLGAP5	discs, large (Drosophila) homolog-associated protein 5
EIF4A1P4	similar to eukaryotic translation initiation factor 4a
ESF1	ESF1, nucleolar pre-rRNA processing protein, homolog (S. cerevisiae)
EXOSC2	exosome component 2
GTPBP4	GTP binding protein 4
HEATR1	HEAT repeat containing 1
HMMR	hyaluronan-mediated motility receptor (RHAMM)
HNRPDL	heterogeneous nuclear ribonucleoprotein D-like
KIF18A	kinesin family member 18A
KIF20A	kinesin family member 20A
MAP1D	methionine aminopeptidase 1D
MCFD2	multiple coagulation factor deficiency 2
MCM4	minichromosome maintenance complex component 4
MKI67	antigen identified by monoclonal antibody Ki-67
MLLT10	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
MSH6	mutS homolog 6 (E. coli)
NUDCD2	NudC domain containing 2
NUP35	nucleoporin 35kDa
PAPOLA	poly(A) polymerase alpha
PDS5A	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)
PSRC1	proline/serine-rich coiled-coil 1
PWP1	PWP1 homolog (S. cerevisiae)
SEPHS1	selenophosphate synthetase 1; similar to selenophosphate synthetase 1
SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein
TCEB1	transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C)
TUBB2C	tubulin, beta 2C
TUBGCP3	tubulin, gamma complex associated protein 3
UTP11L	UTP11-like, U3 small nucleolar ribonucleoprotein, (yeast)
XPO5	exportin 5

Table 3.7: Genes from the down-regulated gene set that are enriched for the GABP motif

Gene ID	Gene Name
ADSL	adenylosuccinate lyase
ANLN	anillin, actin binding protein
ARHGAP11A	Rho GTPase activating protein 11A
AURKA	aurora kinase A; aurora kinase A pseudogene 1
BUB1	budding uninhibited by benzimidazoles 1 homolog (yeast)
C13ORF34	chromosome 13 open reading frame 34
C21orf59	chromosome 21 open reading frame 59
C7orf44	chromosome 7 open reading frame 44
CAPRIN1	cell cycle associated protein 1
CCNF	cyclin F
CCT8P1	chaperonin containing TCP1, subunit 8 (theta)
CSNK2A1	casein kinase 2, alpha 1 polypeptide pseudogene; casein kinase 2, alpha 1 polypeptide
DLGAP5	discs, large (Drosophila) homolog-associated protein 5
EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa
EIF4E	eukaryotic translation initiation factor 4E
EIF4G1	eukaryotic translation initiation factor 4 gamma, 1
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
FBXW2	F-box and WD repeat domain containing 2
G2E3	G2/M-phase specific E3 ubiquitin ligase
GFM1	G elongation factor, mitochondrial 1
GMEB1	glucocorticoid modulatory element binding protein 1
GPRASP2	golgi reassembly stacking protein 2, 55kDa
GTSE1	G-2 and S-phase expressed 1
HNRPDL	heterogeneous nuclear ribonucleoprotein D-like
HSP90B1	heat shock protein 90kDa beta (Grp94), member 1
ING1	inhibitor of growth family, member 1
KIF14	kinesin family member 14
LIG3	ligase III, DNA, ATP-dependent
MLLT10	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
MRPL30	mitochondrial ribosomal protein L30
MSH6	mutS homolog 6 (E. coli)
NF2	neurofibromin 2 (merlin)
NOL10	nucleolar protein 10
ODF2	outer dense fiber of sperm tails 2
PAPOLA	poly(A) polymerase alpha
PIP4K2B	phosphatidylinositol-5-phosphate 4-kinase, type II, beta
PSRC1	proline/serine-rich coiled-coil 1
RABEPK	Rab9 effector protein with kelch motifs
RG9MTD1	RNA (guanine-9-) methyltransferase domain containing 1
RPF2P	brix domain containing 1 pseudogene; brix domain containing 1
SGOL2	shugoshin-like 2 (S. pombe)
SKP2	S-phase kinase-associated protein 2 (p45)
SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein
TAF15	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa
TCP1	t-complex 1
TFB2M	transcription factor B2, mitochondrial
TMPO	thymopoietin
TUBB2C	tubulin, beta 2C
TUBGCP3	tubulin, gamma complex associated protein 3
TUBGCP4	tubulin, gamma complex associated protein 4
UCHL5	ubiquitin carboxyl-terminal hydrolase L5
VPS13B	vacuolar protein sorting 13 homolog B (yeast)
WDR37	WD repeat domain 37
WDR5	WD repeat domain 5
XPO5	exportin 5

Table 3.8: Genes from the down-regulated gene set that are enriched for the Nrf-1 motif

Gene ID	Gene Name
ANLN	anillin, actin binding protein
AURKA	aurora kinase A; aurora kinase A pseudogene 1
CAPRN1	cell cycle associated protein 1
CCDC99	coiled-coil domain containing 99
CCNA2	cyclin A2
CCNB1	cyclin B1
CCNB2	cyclin B2
CCNF	cyclin F
CCT5	chaperonin containing TCP1, subunit 5 (epsilon)
CCT8	chaperonin containing TCP1, subunit 8 (theta)
CENPF	centromere protein F, 350/400ka (mitosin)
DKC1	dyskeratosis congenita 1, dyskerin
DLAT	dihydrolipoamide S-acetyltransferase
EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa
EIF4G1	eukaryotic translation initiation factor 4 gamma, 1
ESF1	ESF1, nucleolar pre-rRNA processing protein, homolog (S. cerevisiae)
EWSR1	Ewing sarcoma breakpoint region 1
EXOSC2	exosome component 2
FARSB	phenylalanyl-tRNA synthetase, beta subunit
GMEB1	glucocorticoid modulatory element binding protein 1
GTPBP4	GTP binding protein 4
GTSE1	G-2 and S-phase expressed 1
HMMR	hyaluronan-mediated motility receptor (RHAMM)
HNRNPAB	heterogeneous nuclear ribonucleoprotein A/B
HNRPDL	heterogeneous nuclear ribonucleoprotein D-like
ING1	inhibitor of growth family, member 1
KIF14	kinesin family member 14
LIG3	ligase III, DNA, ATP-dependent
LYAR	Ly1 antibody reactive homolog (mouse)
MAP1D	methionine aminopeptidase 1D
MCM4	minichromosome maintenance complex component 4
MCM6	minichromosome maintenance complex component 6
MLLT10	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
MRPL15	mitochondrial ribosomal protein L15
MTPAP	mitochondrial poly(A) polymerase
NAA15	NMDA receptor regulated 1
NANP	N-acetylneuraminic acid phosphatase
NF2	neurofibromin 2 (merlin)
NOLC1	nucleolar and coiled-body phosphoprotein 1
NUDCD1	NudC domain containing 1
NUDCD2	NudC domain containing 2
PAPOLA	poly(A) polymerase alpha
PDP2	pyruvate dehydrogenase phosphatase catalytic subunit 2
PNO1	partner of NOB1 homolog (S. cerevisiae)
PSRC1	proline/serine-rich coiled-coil 1
PWP1	PWP1 homolog (S. cerevisiae)
SEC24D	SEC24 family, member D (S. cerevisiae)
SEPHS1	selenophosphate synthetase 1; similar to selenophosphate synthetase 1
SFRS2	splicing factor, arginine/serine-rich 2
SKP2	S-phase kinase-associated protein 2 (p45)
SNRPA1	small nuclear ribonucleoprotein polypeptide A'
SRPK1	SFRS protein kinase 1
SSB	Sjogren syndrome antigen B (autoantigen La)
SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein
TAF1B	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa
TFB2M	transcription factor B2, mitochondrial
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)
TLL4	tubulin tyrosine ligase-like family, member 4
TUBB2C	tubulin, beta 2C
TUBGCP3	tubulin, gamma complex associated protein 3
TUBGCP4	tubulin, gamma complex associated protein 4
UCK2	uridine-cytidine kinase 2
UTP11L	UTP11-like, U3 small nucleolar ribonucleoprotein, (yeast)
WDR12	WD repeat domain 12
WDR5	WD repeat domain 5
XPO5	exportin 5
ZFYVE20	zinc finger, FYVE domain containing 20
ZMYM4	zinc finger, MYM-type 4

Table 3.9: Transcription factor binding motif enrichment for different promoter ranges for the up- and down-regulated genes

	RANGES					
	-600 to 200	-1000 to 200	-1500 to 200	-2000 to 200	-2500 to 200	-3000 to 200
Upregulated genes						
p53	N/A	N/A	N/A	8.59E-05	4.40E-05	N/A
Downregulated genes						
NF-Y	1.53E-09	3.22E-09	6.42E-09	7.24E-08	5.82E-07	1.13E-06
AhR	4.30E-06	5.02E-06	4.35E-05	8.54E-05	N/A	N/A
GABP	6.72E-06	4.00E-05	6.41E-05	N/A	N/A	N/A
Nrf-1	1.03E-05	4.79E-05	N/A	N/A	N/A	N/A

Table 3.10: Up-regulated genes containing p53 motifs in their promoter region

Gene ID	Gene name
ABHD4	abhydrolase domain containing 4
ADRB2	adrenergic, beta-2-, receptor, surface
APAF1	apoptotic peptidase activating factor 1
BBC3	BCL2 binding component 3
BTG2	BTG family, member 2
C10orf10	chromosome 10 open reading frame 10
C10orf99	chromosome 10 open reading frame 99
C12orf54	chromosome 12 open reading frame 54
C1orf183	chromosome 1 open reading frame 183
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CEL	carboxyl ester lipase (bile salt-stimulated lipase)
DCP1B	DCP1 decapping enzyme homolog B (<i>S. cerevisiae</i>)
DQX1	DEAQ box RNA-dependent ATPase 1
DSE	dermatan sulfate epimerase
E2F7	E2F transcription factor 7
EI24	etoposide induced 2.4 mRNA
FAM149A	family with sequence similarity 149, member A
FDXR	ferredoxin reductase
FRMD8	FERM domain containing 8
FXYD2	FXYD domain containing ion transport regulator 2
GRM1	glutamate receptor, metabotropic 1
HPX	hemopexin
ICAM4	intercellular adhesion molecule 4 (Landsteiner-Wiener blood group)
IGFBP3	insulin-like growth factor binding protein 3
LAMP3	lysosomal-associated membrane protein 3
LCE1B	late cornified envelope 1B
LCE1E	late cornified envelope 1E
LCN15	lipocalin 15
LRDD	leucine-rich repeats and death domain containing
MRPL49	mitochondrial ribosomal protein L49
NTN4	netrin 4
NUP62CL	nucleoporin 62kDa C-terminal like
PANK1	pantothenate kinase 1
PHLDA3	pleckstrin homology-like domain, family A, member 3
PLEKHA8	pleckstrin homology domain containing, family A member 8
PLXNB2	plexin B2
PON2	paraoxonase 2
PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform
PPP1R14C	protein phosphatase 1, regulatory (inhibitor) subunit 14C
PPY	pancreatic polypeptide
RAP2B	RAP2B, member of RAS oncogene family
RB1	retinoblastoma 1
SCN4B	sodium channel, voltage-gated, type IV, beta
TCEA3	transcription elongation factor A (SII), 3
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)
TLR1	toll-like receptor 1
TP53I3	tumor protein p53 inducible protein 3
TRIM32	tripartite motif-containing 32
UPB1	ureidopropionase, beta
ZBTB4	zinc finger and BTB domain containing 4

The role of NF-Y (most significantly enriched binding motif in the down-regulated genes) in the transcriptional DNA damage response was explored further. The set of predicted NF-Y targets consisted of 62 genes shown in Table 3.5, including three known NF-Y transcriptional targets *CCNB1*, *CCNB2* and *CCNA2*. Although some of the 62 genes have been shown to have NF-Y promoter occupancy, a majority of the genes identified from our data set as having binding sites for NF-Y in their promoter regions are novel predicted NF-Y targets.

The presence of *bona-fide* NF-Y targets *CCNB1*, *CCNB2* and *CCNA1* among the set of 62 predicted targets improved our confidence in the set of predicted NF-Y targets. The known NF-Y targets contain an additional motif called the cell cycle homology region (CHR) within their promoter regions. Knowing that true NF-Y targets have a CHR motif within their promoter regions, we checked to see whether the novel predicted NF-Y targets had such a CHR motif in their promoter region as well. The promoter regions of the predicted NF-Y target genes (1000 bp upstream of the transcription start site) were scanned for the presence of the CHR motif. 53/62 genes had both CHR and NF-Y motifs (p-value < 1×10^{-9} , calculated using a hypergeometric distribution).

3.4.16 Chromatin immunoprecipitation followed by sequencing (ChIP-seq) confirm NF-Y occupancy for novel predicted NF-Y targets

To experimentally validate promoter occupancy of NF-Y for the computationally predicted, novel NF-Y targets, ChIP-seq was performed on log-growing cells from the most sensitive cell line 4. Briefly, cells were crosslinked in formaldehyde, and the chromatin was isolated, sonicated and immunoprecipitated with antibodies against NF-YA. The crosslinks were reversed and DNA was isolated and sequenced (see Materials and Methods). The sequencing results were analyzed

using Model-based Analysis of ChIP-Seq (MACS) (15) as described in the methods section. Together, the two separate sequencing runs yielded 17 million unique reads for chromatin from the NF-YA ChIP and 16 million unique reads from the control IgG ChIP. For a p-value cut-off of 1×10^{-15} , there were 3100 peaks called for the NF-YA ChIP as compared to the IgG control. At the same p-value cut-off, there were only 39 peaks called in the IgG control as compared to the NF-YA ChIP.

Remarkably, of the 62 genes predicted as having NF-Y promoter occupancy, 54 were positive by ChIP-seq. 48 genes pass the stringent p-value cut off of $p < 10^{-15}$. Three more genes pass a p-value cut-off of $p < 10^{-10}$ and an additional three pass a p-value cut-off of $p < 10^{-5}$. The remaining 8 genes are negative for peaks (see Table 3.11). The four genes with the lowest p-value for peaks from the NF-Y ChIP are shown in the Figure 3.35. The four genes with the highest p-value for peaks from the NF-Y ChIP are shown in the Figure 3.36 and four genes with no significant peaks are shown in Figure 3.37. The top 200 peak sequences were analyzed using the THEME algorithm (16) as a check to confirm that the peaks obtained are indeed from an NF-Y ChIP. As expected top motif enriched in this analysis was the NF-Y motif.

Table 3.11: Peaks called using MACS near the genes predicted to be NF-Y targets

	GeneSymbol	Chromosome	Start	End	pvalue	Fold enrichment	FDR(%)	Distance from feature	Map type	Map subtype
1	CDCA3	chr12	6830277	6832274	10 ⁻³¹⁰	133.46	0	-52	promoter	
2	E1F4A1	chr17	7416194	7417066	10 ⁻³¹⁰	98.47	0	374	gene	exon
3	CENPA	chr2	26861725	26862519	10 ⁻³¹⁰	357.17	0	-119	promoter	
4	UBE2C	chr20	43873963	43875157	10 ⁻³¹⁰	189.8	0	-106	promoter	
5	NF2	chr22	28328924	28329626	10 ⁻³¹⁰	207.48	0	-169	promoter	
6	HNRNPAB	chr5	177563834	177564331	10 ⁻³¹⁰	157.05	0	-55	promoter	
7	TCP1	chr6	160130386	160131899	10 ⁻³¹⁰	165.58	0	-680	promoter	
8	H2AFV	chr7	44853892	44854815	10 ⁻³¹⁰	279.53	0	-236	promoter	
9	TUBB2C	chr9	139254881	139255771	10 ⁻³¹⁰	147.98	0	-143	promoter	
10	HSP90B1	chr12	102847729	102849232	10 ⁻²⁸⁹	84.88	0	-127	promoter	
11	TCEB1	chr8	75046603	75047587	10 ⁻²⁷²	103.87	0	-280	promoter	
12	DLGAP5	chr14	54727553	54728442	10 ⁻²⁴¹	89.51	0	60	gene	exon
13	CENPF	chr1	212842798	212843654	10 ⁻²³⁶	135.48	0	-62	promoter	
14	BUB1	chr2	111152001	111152385	10 ⁻²⁰⁶	96.99	0	-67	promoter	
15	TMPD	chr12	97432860	97433928	10 ⁻²⁰⁰	92.84	0	-63	promoter	
16	WDR37	chr10	1084560	1085801	10 ⁻²⁰⁰	52.78	0	-6975	promoter	
17	CL3orf34	chr13	72199423	72200309	10 ⁻¹⁷⁷	96.68	0	-63	promoter	
18	ASPM	chr1	195382123	195383026	10 ⁻¹⁵⁴	76.61	0	-199	promoter	
19	BUB3	chr10	124903292	124904119	10 ⁻¹⁴³	66.29	0	-57	promoter	
20	CCNB2	chr15	57184020	57184859	10 ⁻¹²²	58.88	0	-41	promoter	
21	EEF1E1	chr6	8047134	8048233	10 ⁻¹¹²	83.25	0	3	gene	exon
22	ECT2	chr3	173950943	173951675	10 ⁻¹⁰⁹	51.94	0	-34	promoter	
23	HMMR	chr5	162819148	162820880	10 ⁻⁹⁸	69.25	0	-115	promoter	
24	NUDCD2	chr5	162819148	162820880	10 ⁻⁹⁸	69.25	0	-402	promoter	
25	CCNF	chr16	2418701	2419591	10 ⁻⁹²	66.41	0	-185	promoter	
26	TPX2	chr20	29790249	29791400	10 ⁻⁸⁸	59.23	0	135	gene	exon
27	CCNA2	chr4	122964064	122964809	10 ⁻⁸⁵	51.94	0	-12	promoter	
28	KIF14	chr1	198856063	198856881	10 ⁻⁸⁴	49.04	0	-26	promoter	
29	PRMT5	chr14	22468027	22469036	10 ⁻⁸⁰	35.23	0	-48	promoter	
30	AURKA	chr20	54400496	54401368	10 ⁻⁷⁹	35.32	0	37	gene	exon
31	CCNB1	chr5	68498277	68499142	10 ⁻⁷⁹	35.62	0	20	gene	exon
32	MCM6	chr2	136350264	136351023	10 ⁻⁷⁸	59.69	0	-84	promoter	
33	CDC20	chr1	43596723	43597632	10 ⁻⁷³	46.45	0	-30	promoter	
34	GTSE1	chr22	45070875	45071614	10 ⁻⁶⁶	56.65	0	-57	promoter	
35	JTV1	chr7	6014891	6015656	10 ⁻⁶²	28.76	0	-56	promoter	
36	THADA	chr2	43675898	43677160	10 ⁻⁵⁴	48.71	0	-58	promoter	
37	SGOL2	chr2	201098742	201099527	10 ⁻⁴⁸	27.5	0	-73	promoter	
38	NEK2	chr1	209915283	209915904	10 ⁻⁴⁴	37.59	0	-37	promoter	
39	GTPBP4	chr10	1024053	1024592	10 ⁻⁴²	32.7	0	-37	promoter	
40	MCM4	chr8	49035555	49036113	10 ⁻³⁷	33.4	0	-87	promoter	
41	KIF20A	chr5	137542165	137542913	10 ⁻³⁶	14.28	0	-133	promoter	
42	TAF15	chr17	31159774	31160773	10 ⁻²⁴	19.27	0	-99	promoter	
43	PKP4	chr2	159021410	159021947	10 ⁻²³	26.84	0.04	-171	promoter	
44	PLK1	chr16	23597207	23598099	10 ⁻²²	19.68	0.04	-43	promoter	
45	DEPDC1	chr1	68734941	68735804	10 ⁻²²	20.84	0.04	-31	promoter	
46	EXOSC2	chr9	132558483	132559321	10 ⁻¹⁹	15.23	0.07	-118	promoter	
47	ODF2	chr9	130257948	130258450	10 ⁻¹⁸	19.28	0.18	-92	promoter	
48	LTV1	chr6	144205514	144206524	10 ⁻¹⁸	14.87	0.28	-12	promoter	
49	SFRS2	chr17	72244755	72245501	10 ⁻¹³	9.05	3.85	-140	promoter	
50	C2orf67	chr2	210744295	210744689	10 ⁻¹²	13.68	5.76	-139	promoter	
51	SMC3	chr10	112316678	112317404	10 ⁻¹¹	10.46	13.4	-408	promoter	
52	ARHGAP11A	chr15	30694567	30695284	10 ⁻⁹	15.16	25.02	15	gene	exon
53	TAF1B	chr2	9900532	9901360	10 ⁻⁹	7.05	31.67	-132	promoter	
54	PSME3	chr17	38238437	38239313	10 ⁻⁵	6.01	98.53	-57	promoter	
55	ZMYM4									
56	PSRC1									
57	METAP2									
58	PIP4K2B									
59	MSH6									
60	HNRPDL									
61	GK									
62	STAG3L3									

Genes predicted as NF-Y targets were examined for highly significant peaks. Genes are ordered based on their p-value, with white rows having high significance, and grey rows containing low-confidence or no peaks. Genes highlighted in red are transcriptionally regulated by NF-Y. Genes in blue have been shown to have NF-Y promoter occupancy. (56,89-91).

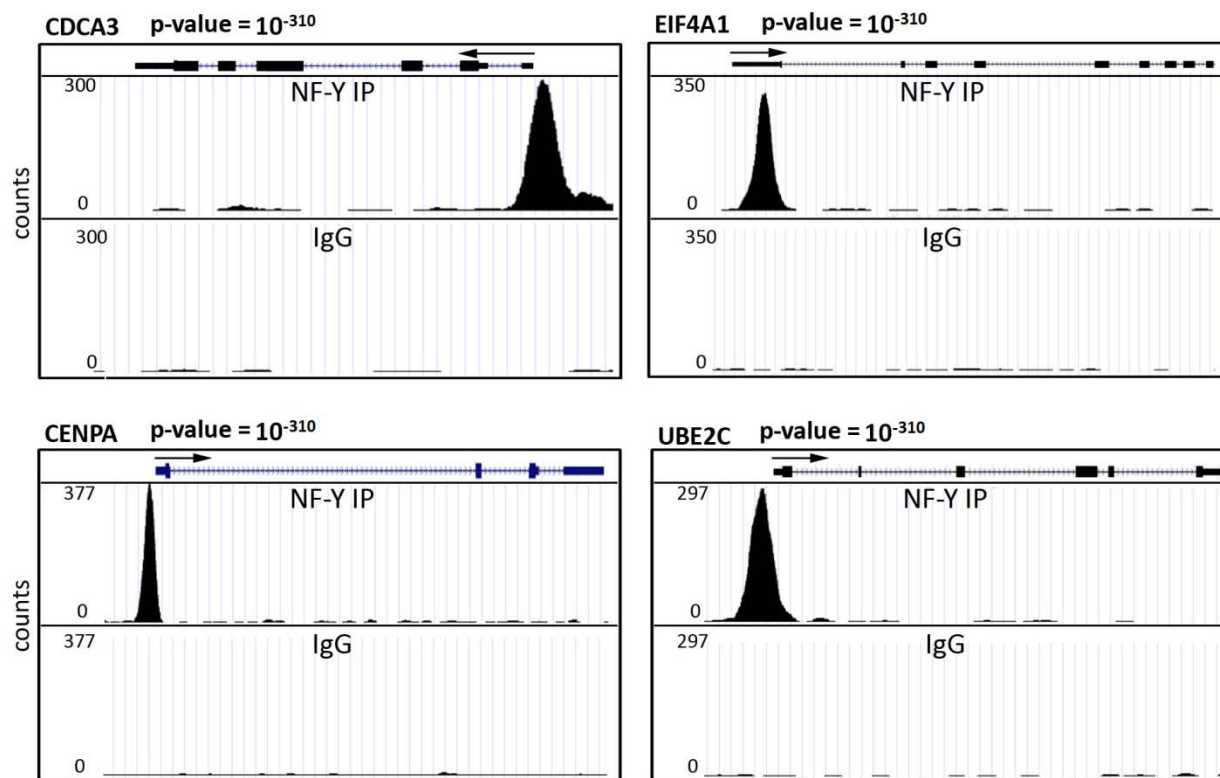


Figure 3.35: Highly significant peaks from MACS analysis of NF-Y ChIP-seq data

The figure shows peaks with the lowest p-value within the down-regulated genes obtained from the NF-Y ChIP-seq. For each gene, the top panel shows reads from the NF-Y IP and the bottom panel shows reads from the negative IgG control. The arrow indicates the direction of transcription

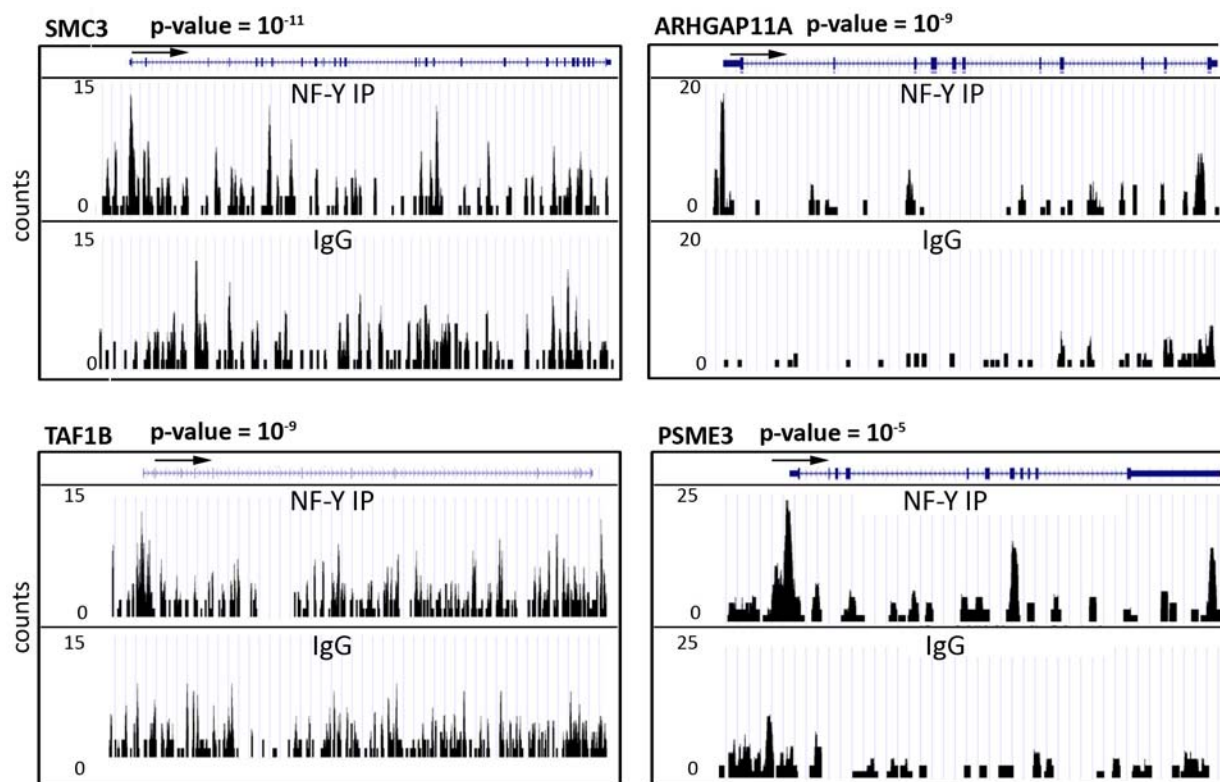


Figure 3.36: Weakly significant peaks from MACS analysis of NF-Y ChIP-seq data

The figure shows peaks with the highest p-value within the down-regulated genes obtained from the NF-Y ChIP-seq. For each gene, the top panel shows reads from the NF-Y IP and the bottom panel shows reads from the negative IgG control. The arrow indicates the direction of transcription

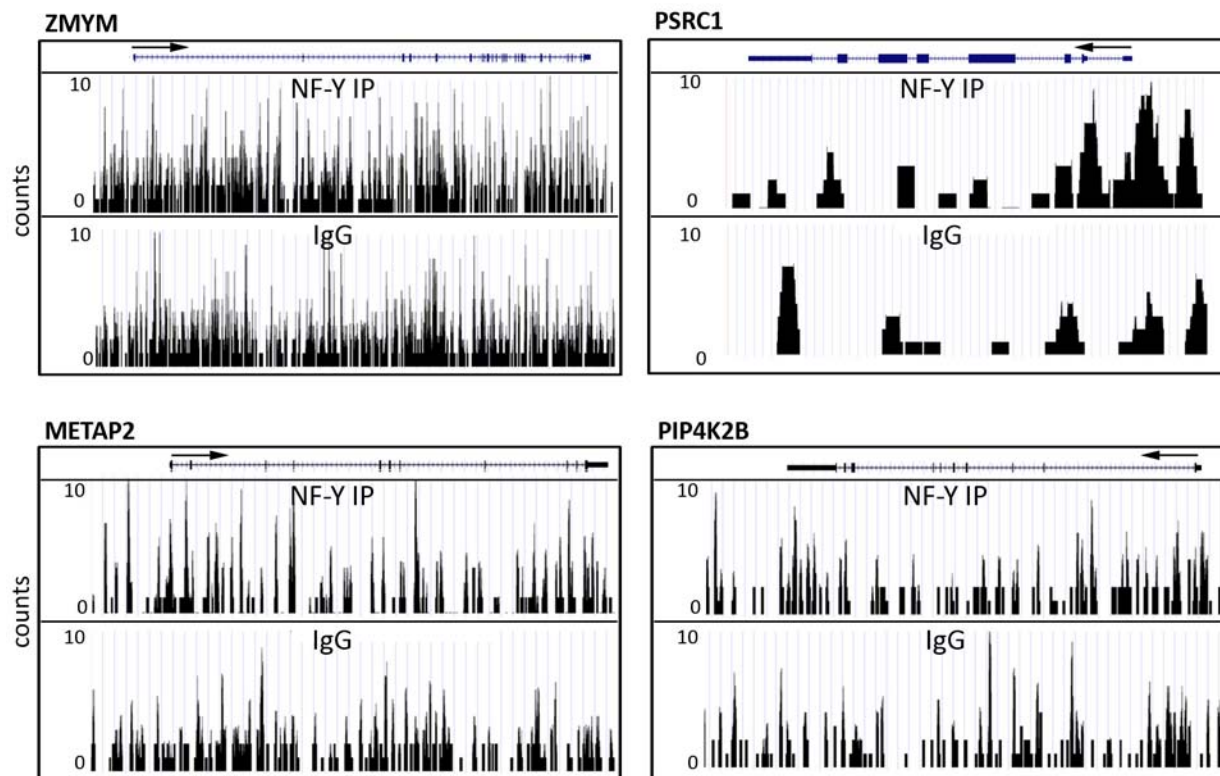


Figure 3.37: Genes with no significant peaks from MACS analysis NF-Y ChIP-seq data

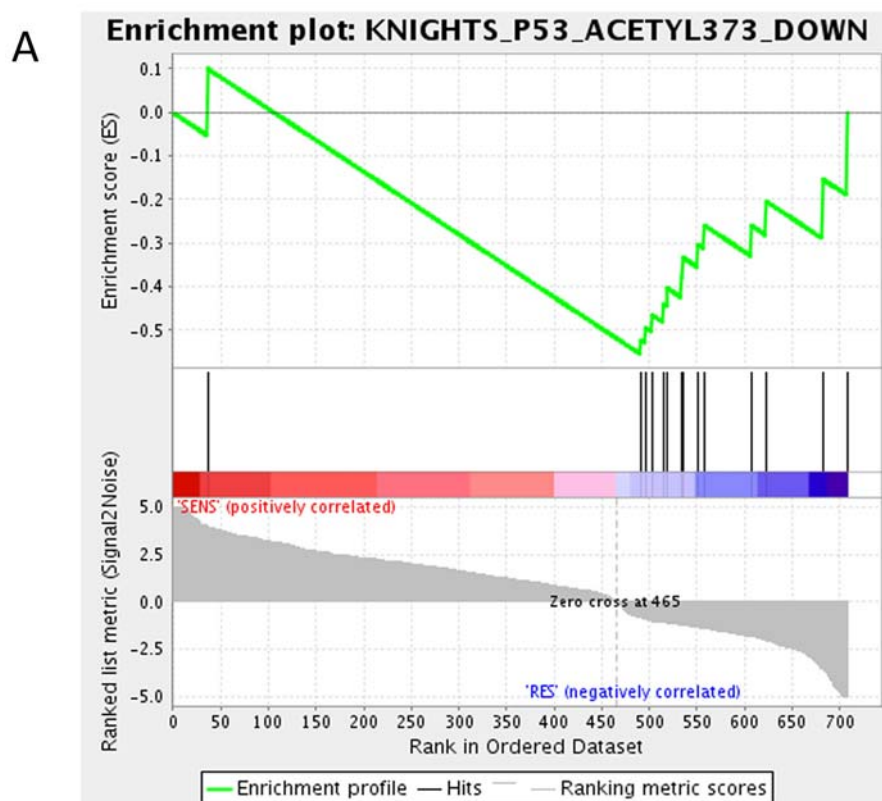
The figure shows peaks with the highest p-value within the down-regulated genes obtained from the NF-Y ChIP-seq. For each gene, the top panel shows reads from the NF-Y IP and the bottom panel shows reads from the negative IgG control. The arrow indicates the direction of transcription

3.4.17 Gene Set Enrichment Analysis (GSEA) shows enrichment of acetyl-p53 repressed genes within the transcription signature

p53 acetylation has been shown to be important for inhibition of NF- κ B transcriptional activity. 68 genes were identified as being repressed upon p53 acetylation at lysine 373 in (41). The 984 probesets identified from our transcription data were analyzed for enrichment in p53 acetylation mediated gene repression using GSEA (see Materials and Methods). This analysis shows that the transcriptional signature is indeed enriched for genes repressed by p53 acetylation, with 14/68 genes present within the transcriptional signature (FDR 0.3%). The enrichment plot is shown in Figure 3.38A with the enrichment score shown in green and the black lines showing the location of the 14 genes along the ranked list of genes within the transcriptional signature. 13/14 genes repressed after p53-acetylation were also repressed in sensitive cell lines after BCNU treatment as seen by the skewing of the black lines towards the blue region of the enrichment plot. Figure 3.38B shows the 14 genes that are repressed after p53 acetylation and present within the transcriptional signature set. 11 out of these 14 genes have strong NF- κ B promoter occupancy and are marked in Figure 3.38B.

3.4.18 Immunoblot analysis shows no significant p53 acetylation in BCNU treated cells

To check whether BCNU induced p53 acetylation at lysine 373 in the BCNU sensitive or resistant cell lines, cells were treated with 40 μ M BCNU and collected at 0, 12, 24, 48 and 72 hours for immunoblot analysis with antibodies against total p53 and acetyl p53 (lysine 373) (see Materials and Methods). Figure 3.39 shows that the antibody used detects p53 acetylation induced by etoposide treatment in cell line 16. The immunoblots for cell lines 4, 5, 13 and 16 treated with BCNU are shown in Figure 3.40 and the quantified values are shown in Figure 3.41. We see no significant differences in p53 acetylation between mock-treated and treated cell lines.



B

Gene ID	Gene Name	NFY occupancy
KITLG	KIT ligand	
CCNA2	cyclin A2	y
NEK2	NIMA (never in mitosis gene A)-related kinase	y
CCNF	cyclin F	y
NF2	neurofibromin 2	y
CCNB1	cyclin B1	y
CDC20	cell division cycle 20 homolog (<i>S.cerevisiae</i>)	y
CCNB2	cyclin B2	y
MKI67	antigen identified by monoclonal antibody Ki-67	y
ANLN	anillin, actin binding protein	
BUB1	BUB1 budding inhibited by benzimidazoles 1 homolog (yeast)	y
MCM6	MCM6 minichromosomes maintenance deficient 6 (<i>S.cerevisiae</i>)	y
API5	apoptosis inhibitor 5	
SMC3	structural maintenance of chromosomes 3	y

Figure 3.38: GSEA of the transcriptional signature shows enrichment of genes repressed upon p53 acetylation

Genes repressed upon p53-acetylation (list of 68 obtained from (41)) were enriched in the 984 probe-sets (FDR 0.3%) in the transcriptional signature. A) Enrichment plot showing the position of the 14 of the 68 genes –black lines– within the 984 probe-sets, ranked based on expression (red – high in sensitive cell lines, blue- low in sensitive cell lines). B) List showing these 14 genes and the promoter occupancy from NF-Y ChIP-seq (y – yes)

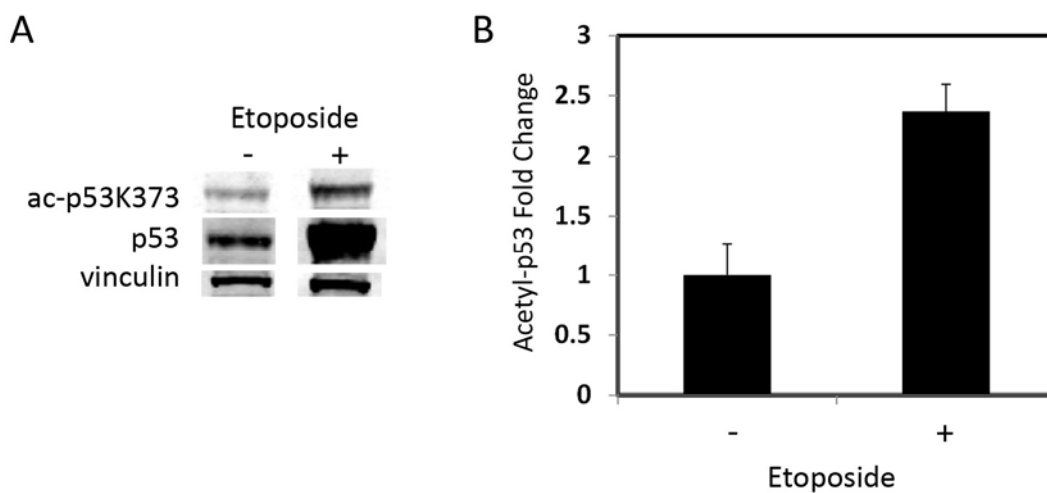


Figure 3.39: Etoposide treated samples probed with the antibody against acetyl-p53 (lys373)
(A) Immunoblots for cell lysate from cell line 16, treated with 50 μ M etoposide (see Materials and Methods) probed with antibodies against acetyl-p53 (lys 373), p53 and vinculin; (B) Bar chart showing the average fold-change in acetyl-p53 (lys373) for etoposide treated samples over untreated samples.

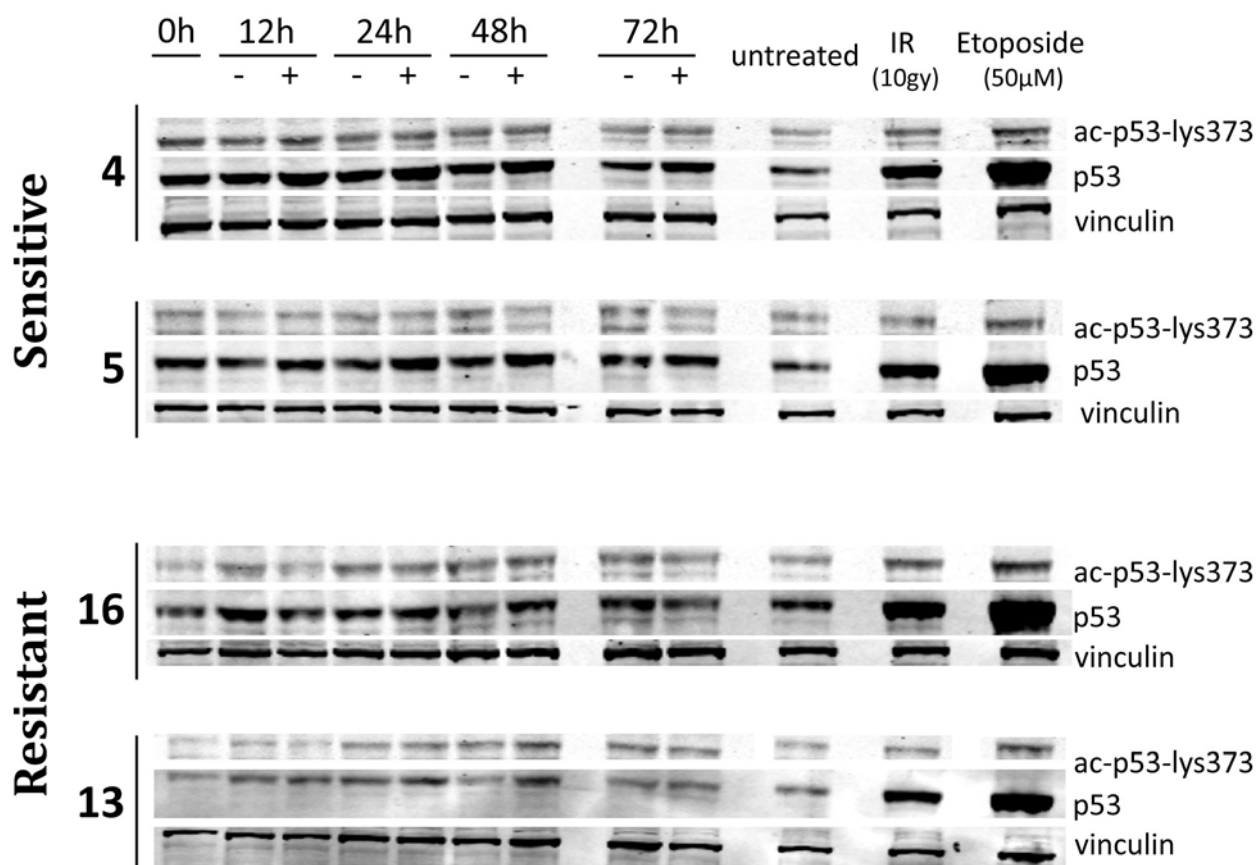


Figure 3.40: Immunoblot probed for acetyl-p53 (lysine 373)

Immunoblots for cell lines 4, 5, 16 and 13 at various time points after 40μM BCNU treatment probed with antibodies against acetyl-p53 (lys373), p53 and vinculin (loading control). Cell lysate from untreated, IR (10gy) treated and etoposide (50μM) treated samples from cell line 16 were used as controls for each blot and for comparison across blots.

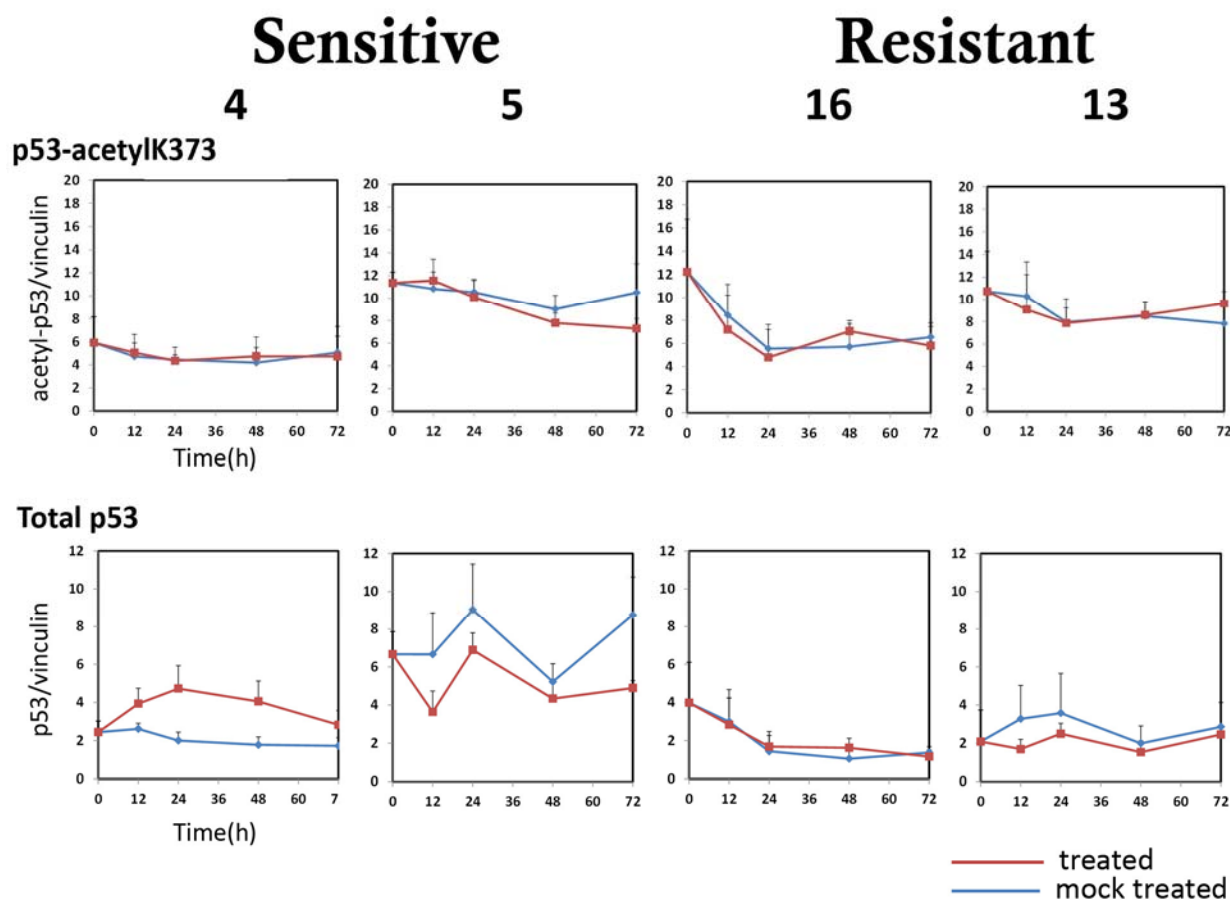


Figure 3.41: p53 acetylation as measured by immunoblot after BCNU treatment

Acetylated p53 (top) and total p53 (bottom) levels in cell lines 4, 5, 16 and 13 in mock-treated and BCNU treated samples at multiple time-points. All values are normalized to the loading control vinculin and scaled to a positive IR control..

3.5 Discussion

The goals of this study were: i) to identify a transcriptional signature that cell lines that are resistant or sensitive to BCNU treatment; and ii) to find putative transcriptional regulators of this transcription signature.

The wide range of BCNU sensitivity observed in the panel of twenty four genetically varied cell lines allowed the identification of cell lines with extreme BCNU sensitivity or resistance. The 706 genes identified as being differentially expressed between sensitive and resistant cell lines after BCNU treatment reveal interesting properties about the BCNU induced transcriptional differences between these cell lines. Genes within the transcriptional signature have visibly stronger gene induction or repression upon BCNU treatment in the sensitive cell lines as compared to the resistant cell lines. The muted transcriptional response in the resistant cell lines might suggest that the resistant cell lines induce DNA damage response pathways through mechanisms other than transcription. Additionally, since the ANOVA analysis enforced a differential gene expression signature between sensitive and resistant cell lines, it might also be the case that the transcripts induced or repressed in resistant cell lines are also induced and repressed to similar extents in the sensitive cell lines and therefore not included in the transcriptional signature.

The transcriptional signature was partitioned into two sub-groups with different biological and functional annotation that correlated with cellular phenotype. We infer that two major and distinct processes are initiated in the BCNU treated sensitive cell lines. The first is the induction of apoptotic cell death as suggested by gene ontology enrichment of apoptosis within the up-

regulated set, networks enriched in cell death related genes and the experimental observation of apoptotic cell death. The second process initiated in BCNU treated sensitive cell lines is the lack of entry into mitosis as suggested by gene ontology enrichment of the mitotic phase of the cell cycle within the down-regulated set, networks enriched in cell cycle related genes and experimental observation of a decreased mitotic population. Taken together, these results suggest either that the sensitive cell lines don't reach mitosis or that they initiate a protective program that prevents the entry of cells into mitosis or even both. All of these stall proliferation of cells containing BCNU induced DNA damage. This also suggests that sensitive cell lines simultaneously initiate cell death mechanisms most likely to eliminate cells that are unable to repair the BCNU induced damage.

The other observation from the transcriptional data set was the enrichment of the p53 signaling canonical pathway within the up-regulated set of genes using network building algorithms, which was also seen experimentally. Note that although the p53 signaling pathway was enriched only in the up-regulated gene set, p53 appears as a hub node in networks generated for both the up-regulated and down-regulated gene sets (Figure 3.23B and Figure 3.24B). This discrepancy arises from the definition of the canonical p53 signaling pathway that mostly include only those p53 targets that are transactivated as seen in Figure 3.23, but not those targets that are repressed or inhibited by p53 as seen in Figure 3.24. Therefore, although p53-mediated repression interactions are included in the Ingenuity database and seen in Figure 3.24B, they have not been incorporated into the p53 canonical pathway.

In addition to this, there are other assumptions made in the use of such network analysis techniques. Using curated knowledge to understand the gene set of interest assumes that transcript levels correctly depict protein levels. This, however, is not necessarily true, and genes that seem to be induced at the transcript level might still be lacking at the protein level due to rapid degradation of the protein or poor translation of the transcript (42-44). The other assumption with these analysis techniques is the generality of the curated data. Certain reactions, interactions, induction and repression are often specific to the particular cell type or condition being studied while being irrelevant for the system under consideration. Moreover, the quality of interactions in the database might vary greatly depending on what experimental techniques were used to identify the interactions. Nevertheless, using known biological interactions can help us identify pathways that are possibly activated or repressed within our data set, and produce a better understanding of the meaning of our transcriptional signature in the context of DNA damage. For example, the network and functional analysis not only suggested a role for p53 in the BCNU induced transcriptional response but also suggested the induction of apoptosis and the absence of entry into mitosis, both of which were also seen experimentally.

Significant binding site enrichment of the transcription factors NF-Y, AhR, GABP and Nrf-1 within promoter regions of the down-regulated genes suggested that these transcription factors might control the expression of genes involved in mitosis. The fact that the NF-Y motif was enriched for all ranges of the promoter region tested improved our confidence in the predicted targets of NF-Y. The only transcription factor binding site enrichment identified for the up-regulated set of genes was that of p53. This is in concurrence with the induction of canonical p53 targets (Figure 3.33) suggesting p53 transcriptional activation of a sub-set of up-regulated genes.

However, the lack of other significantly enriched transcription factor binding site enrichment within promoter regions of the up-regulated genes could be due to several reasons. One could be that the binding motifs might occur outside the search range possible in EXPANDER. Another reason could be that our knowledge of binding motifs for transcription factors is incomplete and so the motif enriched within the up-regulated set might still be unknown and absent in the TRANFAC data base that the motif-searching algorithm relies. Moreover, the set of up-regulated genes with the same expression patterns might be activated by different transcription factors, thus diluting out the enrichment for any one particular transcription factor within the set of up-regulated genes.

Further exploration of BCNU induced transcriptional repression of mitotic genes was pursued with the most significantly enriched transcription factor, NF-Y. The second most significant transcription factor, AHR was not explored because examination of its list of predicted targets did not include canonical AHR targets such as the *CYP1* family of genes. Moreover, the transcription factor motif for AHR is often identified as a false positive in computational promoter sequence analyses (personal communication, Shao-shan Huang). Furthermore, literature review of the transcription factors suggested an interesting role for NF-Y in the control of mitotic genes after DNA damage.

Nuclear Factor Y (NF-Y) is a trimeric complex composed of NF-YA, NF-YB and NF-YC (45, 46). NF-Y has been shown to bind to promoter regions of, and transactivate multiple genes in a CCAAT motif dependent manner. The CCAAT motif is over-represented in promoter regions of the human genome, with ~30% of promoters containing this pentanucleotide sequence (47). In

line with this, NF-Y has been shown to bind and transactivate genes showing a wide range of expression patterns including ubiquitously expressed genes, cell cycle dependent genes and stress responsive genes (48-50). Moreover, NF-Y has been shown to recruit and increase binding affinity of other transcription factors at promoter regions (50-53). These facts suggest a complex and combinatorial regulation of NF-Y targets. Over the past decade, NF-Y has emerged as an important regulator of cell cycle progression, cell cycle checkpoints and chromatin remodeling under both non-stressed and DNA damage conditions (54-56).

Bioinformatic analyses of promoter sequences have implicated NF-Y as a regulator of a large set of mitotic cell cycle genes due to the presence of the CCAAT motif in these promoter sequences (12,57). A few of these genes (*CCNB1*, *CCNB2*, *CCNA2* and *CDC25C*) have been shown experimentally to be transcriptionally regulated by NF-Y (56). These known NF-Y mitotic targets have varied expression levels for different phases of the cell cycle, with low or no expression during G1, expression beginning during S-phase, with transcript levels accumulating all the way up to the end of G2. Such gene expression regulation is dependent on a motif present within the promoter regions of the mitotic genes, referred to as a cell cycle homology region (CHR) (58-61). In the absence of CHR elements in the promoter region, constitutive expression of mitotic genes has been observed (62,63). The exact mechanism of transcriptional repression of mitotic genes via the CHR motifs is not well understood, but thought to involve multi-protein complexes (64,65). The significant enrichment of the CHR motif within the promoter regions of the predicted NF-Y targets within our data set further improved our confidence in the predicted set of novel NF-Y targets, and suggested similar expression regulation for these new putative NF-Y targets. Moreover, significant promoter occupancy of NF-Y for 54 ($p < 10^{-5}$) of the 62

predicted target genes implies NF-Y regulation of these mitotic genes and a possible role for NF-Y in a G2 arrest after BCNU treatment.

Previous studies showed that human cells treated with BCNU induce a late S or G2/M arrest (28,40,66). However, the transcriptional control in the initiation of such an arrest after BCNU treatment has not been studied yet. G2/M arrest after DNA damage is thought to be initiated mainly through the regulation of CDK1 activity through multiple processes including an increase in its inhibitory phosphorylation by WEE1 (67-70), inhibition of its activating phosphatase CDC25C (71-73), exclusion from the nucleus and sequestration of its binding partner CCNB1 (74,75). Additional transcriptional control of the G2/M arrest has been suggested by the p53 (76-80) and E2F family of transcription factors (81-85). Few studies have also implicated NF-Y in DNA damage induced G2/M arrest.

Protein levels of mitotic proteins such as CCNB1 and CDK1 drop during DNA damage induced G2/M arrest. This drop in protein level is partially attributed to protein degradation and partially attributed to transcriptional repression of the CCNB1 and CDK1 genes (76). These genes have CCAAT motifs in their promoter regions (84,86) and NF-Y has been shown to regulate expression from the promoters of these genes. NF-Y activity at the promoter regions of these mitotic genes is reduced upon DNA damage (54) thus explaining the decreased expression of CCNB1 and CDK1 during G2/M arrest after DNA damage. Moreover, cells lacking NF-YA or NF-YB show cell cycle defects, in particular the induction of a G2/M arrest with concomitant decrease in the expression of mitotic genes (55,87). This shows that NF-Y transcriptional control of cell cycle genes is important for proper control of G2/M both in the absence and presence of

DNA damage. In the context of these findings, our study adds to the list of mitotic genes with NF-Y promoter occupancy, that are most likely repressed during a G2/M arrest with the same mechanisms of repression seen for the key mitotic gene *CCNB1*. This also suggests a role for similar transcriptional control of a G2/M arrest in the sensitive cell lines after BCNU treatment.

One mechanism proposed for transcriptional inactivation of the NF-Y after DNA damage is the inhibition of NF-Y by acetylated p53 (88). p53 has been shown to localize to promoter regions of mitotic genes especially after DNA damage (89). Biochemical analysis revealed that this occurs through the interactions of p53 with the NF-YC component of trimeric NF-Y. Mutational analysis identified that the C-terminal regions of p53 is required for the p53-NF-Y interaction (88). Moreover, p53 acetylation at the C-terminal region of p53 was shown to be important for the repression of NF-Y activity after DNA damage. In addition to this, p53 acetylation alone was shown to induce transcriptional changes and G2/M arrest (41). The enrichment within our transcriptional signature (especially in the down-regulated set) for genes that are repressed by p53 acetylation suggested a similar mechanism of transcriptional control for the set of down-regulated mitotic genes. However, immunoblot analysis showed a lack of p53 acetylation after BCNU treatment in cell lines 4 and 16. This outcome could result from one of many possible reasons. One possibility is that p53 is not acetylated in these cell lines after BCNU treatment, at least at the site probed (lysine 373). In this case, it could be that lymphoblastoid cell lines used in this study employ mechanisms other than p53 acetylation for the repression of NF-Y bound mitotic genes. Another possibility is that the amounts of acetylated p53 might be too low to be detected using the technique we used. It could also be that acetylation might be occurring at lysines other than the site probed that might also be able to inhibit NF-Y. Therefore, we were

unable to prove or disprove the role of p53 acetylation in mediating the repression of NF-Y mitotic targets after DNA damage.

Taken together, our results suggest transcriptional control of the observed G2 arrest in the BCNU treated sensitive cell lines by NF-Y possibly mediated by p53 acetylation or some as yet unknown mechanism as shown in Figure 3.42. The model suggests that under normal cell cycle progression, NF-Y transcriptionally regulates a large panel of mitotic genes. In the presence of DNA damage, the transcriptional activity of NF-Y on this panel of mitotic genes is inhibited, possibly by acetylated p53 or some other novel mechanism. This model also suggests a possible link between DNA damage signaling and control of cell cycle thus facilitating constant monitoring of the presence of damaged DNA.

To conclude, we identified a transcriptional signature of genes differentially expressed between sensitive and resistant cell lines after BCNU treatment that was informative of two major and distinct processes initiated by sensitive cell lines after BCNU treatment – the stalling of entry into mitosis and the initiation of cell death by apoptosis. Furthermore, we identified NF-Y as a possible regulator of BCNU induced G2/M arrest and also identified novel NF-Y targets that are part of the transcriptional response for a G2 arrest after BCNU treatment. Finally, the lack of p53 acetylation in the cell lines used here suggests that these cell lines might use mechanism other than the known mechanism of p53 acetylation for the repression of NF-Y mitotic target genes after BCNU treatment.

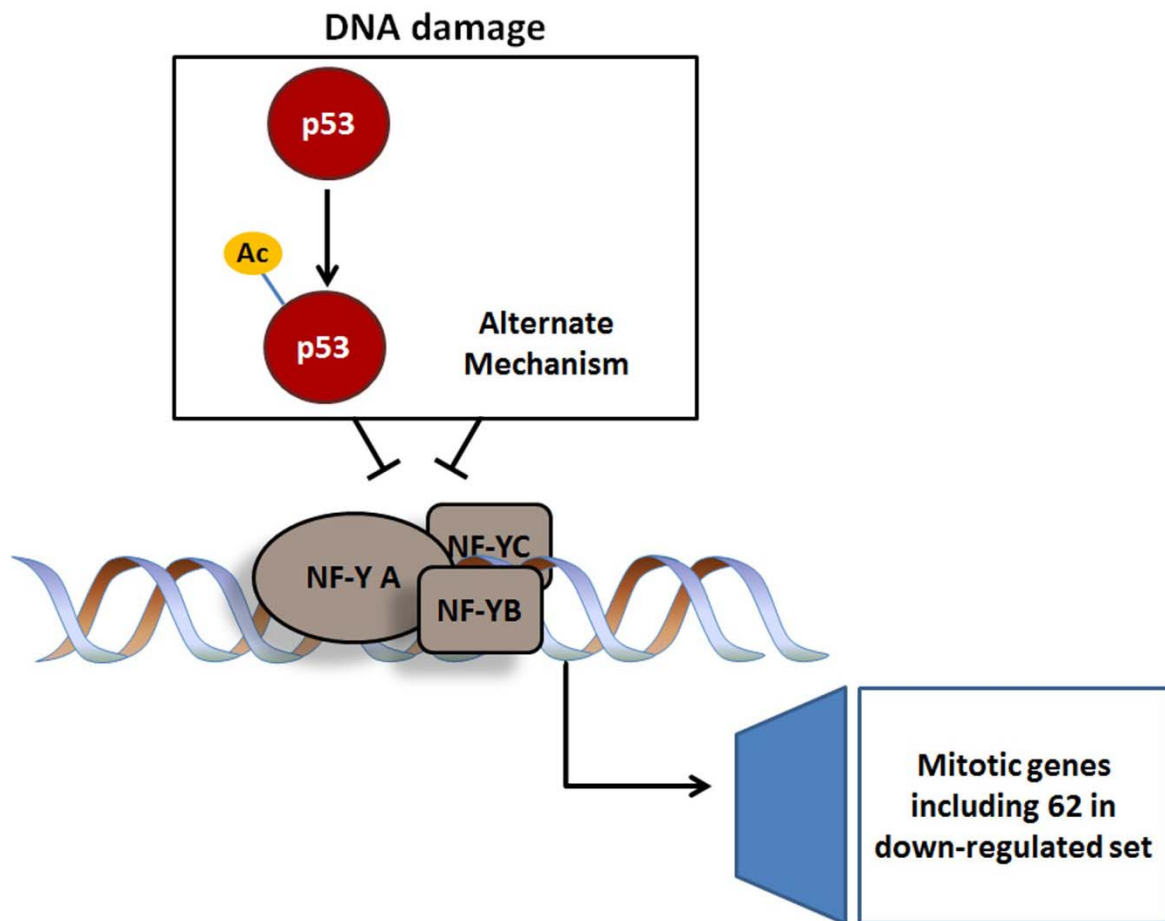


Figure 3.42: Model for repression of NF-Y mitotic targets mediated by p53 acetylation

3.6 References

1. Friedberg, E.C. (2006) *DNA repair and mutagenesis*. 2nd ed. ASM Press, Washington, D.C.
2. Jelinsky, S.A. and Samson, L.D. (1999) Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc Natl Acad Sci U S A*, **96**, 1486-1491.
3. Jen, K.Y. and Cheung, V.G. (2003) Transcriptional response of lymphoblastoid cells to ionizing radiation. *Genome Res*, **13**, 2092-2100.
4. Zhou, T., Chou, J., Mullen, T.E., Elkon, R., Zhou, Y., Simpson, D.A., Bushel, P.R., Paules, R.S., Lobenhofer, E.K., Hurban, P. *et al.* (2007) Identification of primary transcriptional regulation of cell cycle-regulated genes upon DNA damage. *Cell Cycle*, **6**, 972-981.
5. Mees, C., Nemunaitis, J. and Senzer, N. (2009) Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. *Cancer Gene Ther*, **16**, 103-112.
6. Begley, T.J., Rosenbach, A.S., Ideker, T. and Samson, L.D. (2002) Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res*, **1**, 103-112.
7. Begley, T.J., Rosenbach, A.S., Ideker, T. and Samson, L.D. (2004) Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. *Mol Cell*, **16**, 117-125.
8. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. and Mesirov, J.P. (2006) GenePattern 2.0. *Nat Genet*, **38**, 500-501.
9. Shamir, R., Maron-Katz, A., Tanay, A., Linhart, C., Steinfeld, I., Sharan, R., Shiloh, Y. and Elkon, R. (2005) EXPANDER--an integrative program suite for microarray data analysis. *BMC Bioinformatics*, **6**, 232.
10. Sharan, R., Maron-Katz, A. and Shamir, R. (2003) CLICK and EXPANDER: a system for clustering and visualizing gene expression data. *Bioinformatics*, **19**, 1787-1799.
11. Ulitsky, I., Maron-Katz, A., Shavit, S., Sagir, D., Linhart, C., Elkon, R., Tanay, A., Sharan, R., Shiloh, Y. and Shamir, R. (2010) Expander: from expression microarrays to networks and functions. *Nat Protoc*, **5**, 303-322.
12. Elkon, R., Linhart, C., Sharan, R., Shamir, R. and Shiloh, Y. (2003) Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res*, **13**, 773-780.
13. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*, **4**, P3.
14. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, **102**, 15545-15550.
15. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.
16. Macisaac, K.D., Gordon, D.B., Nekludova, L., Odom, D.T., Schreiber, J., Gifford, D.K., Young, R.A. and Fraenkel, E. (2006) A hypothesis-based approach for identifying the

- binding specificity of regulatory proteins from chromatin immunoprecipitation data. *Bioinformatics*, **22**, 423-429.
17. Poot, M., Silber, J.R. and Rabinovitch, P.S. (2002) A novel flow cytometric technique for drug cytotoxicity gives results comparable to colony-forming assays. *Cytometry*, **48**, 1-5.
 18. Ormerod, M.G. (2004) Cell-cycle analysis of asynchronous populations. *Methods Mol Biol*, **263**, 345-354.
 19. Ormerod, M.G. and Kubbies, M. (1992) Cell cycle analysis of asynchronous cell populations by flow cytometry using bromodeoxyuridine label and Hoechst-propidium iodide stain. *Cytometry*, **13**, 678-685.
 20. Poot, M., Schmitt, H., Seyschab, H., Koehler, J., Chen, U., Kaempf, U., Kubbies, M., Schindler, D., Rabinovitch, P.S. and Hoehn, H. (1989) Continuous bromodeoxyuridine labeling and bivariate ethidium bromide/Hoechst flow cytometry in cell kinetics. *Cytometry*, **10**, 222-226.
 21. Kubbies, M., Schindler, D., Hoehn, H. and Rabinovitch, P.S. (1985) BrdU-Hoechst flow cytometry reveals regulation of human lymphocyte growth by donor-age-related growth fraction and transition rate. *J Cell Physiol*, **125**, 229-234.
 22. Poot, M., Schindler, D., Kubbies, M., Hoehn, H. and Rabinovitch, P.S. (1988) Bromodeoxyuridine amplifies the inhibitory effect of oxygen on cell proliferation. *Cytometry*, **9**, 332-338.
 23. Fry, R.C., Svensson, J.P., Valiathan, C., Wang, E., Hogan, B.J., Bhattacharya, S., Bugni, J.M., Whittaker, C.A. and Samson, L.D. (2008) Genomic predictors of interindividual differences in response to DNA damaging agents. *Genes Dev*, **22**, 2621-2626.
 24. Xu, G.W., Mymryk, J.S. and Cairncross, J.G. (2005) Pharmaceutical-mediated inactivation of p53 sensitizes U87MG glioma cells to BCNU and temozolomide. *Int J Cancer*, **116**, 187-192.
 25. Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson, J., Jr., Boguski, M.S. *et al.* (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83-87.
 26. Wagner, V.E., Bushnell, D., Passador, L., Brooks, A.I. and Iglewski, B.H. (2003) Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *J Bacteriol*, **185**, 2080-2095.
 27. Murray, J.I., Whitfield, M.L., Trinklein, N.D., Myers, R.M., Brown, P.O. and Botstein, . (2004) Diverse and specific gene expression responses to stresses in cultured human cells. *Mol Biol Cell*, **15**, 2361-2374.
 28. Biroccio, A., Bufalo, D.D., Ricca, A., D'Angelo, C., D'Orazi, G., Sacchi, A., Soddu, S. and Zupi, G. (1999) Increase of BCNU sensitivity by wt-p53 gene therapy in glioblastoma lines depends on the administration schedule. *Gene Ther*, **6**, 1064-1072.
 29. Wu, M., Kelley, M.R., Hansen, W.K. and Martin, W.J., 2nd. (2001) Reduction of BCNU toxicity to lung cells by high-level expression of O(6)-methylguanine-DNA methyltransferase. *Am J Physiol Lung Cell Mol Physiol*, **280**, L755-761.
 30. Kim, J.H., Zheng, L.T., Lee, W.H. and Suk, K. (2011) Pro-apoptotic role of integrin beta3 in glioma cells. *J Neurochem*.
 31. Batista, L.F., Roos, W.P., Christmann, M., Menck, C.F. and Kaina, B. (2007) Differential sensitivity of malignant glioma cells to methylating and chloroethylating anticancer drugs: p53 determines the switch by regulating xpc, ddb2, and DNA double-strand breaks. *Cancer Res*, **67**, 11886-11895.

32. Zhou, Y., Li, W., Xu, Q. and Huang, Y. (2010) Elevated expression of Dickkopf-1 increases the sensitivity of human glioma cell line SHG44 to BCNU. *J Exp Clin Cancer Res*, **29**, 131.
33. Petak, I., Mihalik, R., Bauer, P.I., Suli-Vargha, H., Sebestyen, A. and Kopper, L. (1998) BCNU is a caspase-mediated inhibitor of drug-induced apoptosis. *Cancer Res*, **58**, 614-618.
34. Hickman, M.J. and Samson, L.D. (1999) Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proc Natl Acad Sci U S A*, **96**, 10764-10769.
35. van Engeland, M., Nieland, L.J., Ramaekers, F.C., Schutte, B. and Reutelingsperger, C.P. (1998) Annexin V-affinity assay: a review on an apoptosis detection system based on phosphatidylserine exposure. *Cytometry*, **31**, 1-9.
36. Andree, H.A., Stuart, M.C., Hermens, W.T., Reutelingsperger, C.P., Hemker, H.C., Frederik, P.M. and Willems, G.M. (1992) Clustering of lipid-bound annexin V may explain its anticoagulant effect. *J Biol Chem*, **267**, 17907-17912.
37. Taylor, W.R. (2004) FACS-based detection of phosphorylated histone H3 for the quantitation of mitotic cells. *Methods Mol Biol*, **281**, 293-299.
38. Juan, G., Traganos, F., James, W.M., Ray, J.M., Roberge, M., Sauve, D.M., Anderson, H. and Darzynkiewicz, Z. (1998) Histone H3 phosphorylation and expression of cyclins A and B1 measured in individual cells during their progression through G2 and mitosis. *Cytometry*, **32**, 71-77.
39. Jin, G., Cook, S., Cui, B., Chen, W.C., Keir, S.T., Killela, P., Di, C., Payne, C.A., Gregory, S.G., McLendon, R. *et al.* (2010) HDMX regulates p53 activity and confers chemoresistance to 3-bis(2-chloroethyl)-1-nitrosourea. *Neuro Oncol*, **12**, 956-966.
40. Xu, G.W., Mymryk, J.S. and Cairncross, J.G. (2005) Inactivation of p53 sensitizes astrocytic glioma cells to BCNU and temozolomide, but not cisplatin. *J Neurooncol*, **74**, 141-149.
41. Knights, C.D., Catania, J., Di Giovanni, S., Muratoglu, S., Perez, R., Swartzbeck, A., Quong, A.A., Zhang, X., Beerman, T., Pestell, R.G. *et al.* (2006) Distinct p53 acetylation cassettes differentially influence gene-expression patterns and cell fate. *J Cell Biol*, **173**, 533-544.
42. Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, **18**, 533-537.
43. Takechi, T., Okabe, H., Fujioka, A., Murakami, Y. and Fukushima, M. (1998) Relationship between Protein Levels and Gene Expression of Dihydropyrimidine Dehydrogenase in Human Tumor Cells during Growth in Culture and in Nude Mice. *Cancer Science*, **89**, 1144-1153.
44. Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X.S. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533-538.
45. Sinha, S., Kim, I.S., Sohn, K.Y., de Crombrughe, B. and Maity, S.N. (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol*, **16**, 328-337.
46. McNabb, D.S., Xing, Y. and Guarente, L. (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev*, **9**, 47-58.

47. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol*, **212**, 563-578.
48. Framson, P. and Bornstein, P. (1993) A serum response element and a binding site for NF-Y mediate the serum response of the human thrombospondin 1 gene. *J Biol Chem*, **268**, 4989-4996.
49. Wright, K.L., Moore, T.L., Vilen, B.J., Brown, A.M. and Ting, J.P. (1995) Major histocompatibility complex class II-associated invariant chain gene expression is up-regulated by cooperative interactions of Sp1 and NF-Y. *J Biol Chem*, **270**, 20978-20986.
50. Zwicker, J., Gross, C., Lucibello, F.C., Truss, M., Ehlert, F., Engeland, K. and Muller, R. (1995) Cell cycle regulation of cdc25C transcription is mediated by the periodic repression of the glutamine-rich activators NF-Y and Sp1. *Nucleic Acids Res*, **23**, 3822-3830.
51. Jackson, S.M., Ericsson, J., Mantovani, R. and Edwards, P.A. (1998) Synergistic activation of transcription by nuclear factor Y and sterol regulatory element binding protein. *J Lipid Res*, **39**, 767-776.
52. Ericsson, J., Jackson, S.M. and Edwards, P.A. (1996) Synergistic binding of sterol regulatory element-binding protein and NF-Y to the farnesyl diphosphate synthase promoter is critical for sterol-regulated expression of the gene. *J Biol Chem*, **271**, 24359-24364.
53. van Ginkel, P.R., Hsiao, K.M., Schjerven, H. and Farnham, P.J. (1997) E2F-mediated growth regulation requires transcription factor cooperation. *J Biol Chem*, **272**, 18367-18374.
54. Basile, V., Mantovani, R. and Imbriano, C. (2006) DNA damage promotes histone deacetylase 4 nuclear localization and repression of G2/M promoters, via p53 C-terminal lysines. *J Biol Chem*, **281**, 2347-2357.
55. Hu, Q., Lu, J.F., Luo, R., Sen, S. and Maity, S.N. (2006) Inhibition of CBF/NF-Y mediated transcription activation arrests cells at G2/M phase and suppresses expression of genes activated at G2/M phase of the cell cycle. *Nucleic Acids Res*, **34**, 6272-6285.
56. Manni, I., Mazzaro, G., Gurtner, A., Mantovani, R., Haugwitz, U., Krause, K., Engeland, K., Sacchi, A., Soddu, S. and Piaggio, G. (2001) NF-Y mediates the transcriptional inhibition of the cyclin B1, cyclin B2, and cdc25C promoters upon induced G2 arrest. *J Biol Chem*, **276**, 5570-5576.
57. Linhart, C., Elkon, R., Shiloh, Y. and Shamir, R. (2005) Deciphering transcriptional regulatory elements that encode specific cell cycle phasing by comparative genomics analysis. *Cell Cycle*, **4**, 1788-1797.
58. Katula, K.S., Wright, K.L., Paul, H., Surman, D.R., Nuckolls, F.J., Smith, J.W., Ting, J.P., Yates, J. and Cogswell, J.P. (1997) Cyclin-dependent kinase activation and S-phase induction of the cyclin B1 gene are linked through the CCAAT elements. *Cell Growth Differ*, **8**, 811-820.
59. Wasner, M., Tschop, K., Spiesbach, K., Haugwitz, U., Johne, C., Mossner, J., Mantovani, R. and Engeland, K. (2003) Cyclin B1 transcription is enhanced by the p300 coactivator and regulated during the cell cycle by a CHR-dependent repression mechanism. *FEBS Lett*, **536**, 66-70.

60. Lange-zu Dohna, C., Brandeis, M., Berr, F., Mossner, J. and Engeland, K. (2000) A CDE/CHR tandem element regulates cell cycle-dependent repression of cyclin B2 transcription. *FEBS Lett*, **484**, 77-81.
61. Wasner, M., Haugwitz, U., Reinhard, W., Tschop, K., Spiesbach, K., Lorenz, J., Mossner, J. and Engeland, K. (2003) Three CCAAT-boxes and a single cell cycle genes homology region (CHR) are the major regulating sites for transcription from the human cyclin B2 promoter. *Gene*, **312**, 225-237.
62. Lucibello, F.C., Truss, M., Zwicker, J., Ehlert, F., Beato, M. and Muller, R. (1995) Periodic cdc25C transcription is mediated by a novel cell cycle-regulated repressor element (CDE). *EMBO J*, **14**, 132-142.
63. Zwicker, J., Lucibello, F.C., Wolfrain, L.A., Gross, C., Truss, M., Engeland, K. and Muller, R. (1995) Cell cycle regulation of the cyclin A, cdc25C and cdc2 genes is based on a common mechanism of transcriptional repression. *EMBO J*, **14**, 4514-4522.
64. Tommasi, S. and Pfeifer, G.P. (1995) In vivo structure of the human cdc2 promoter: release of a p130-E2F-4 complex from sequences immediately upstream of the transcription initiation site coincides with induction of cdc2 expression. *Mol Cell Biol*, **15**, 6901-6913.
65. Muller, G.A. and Engeland, K. (2010) The central role of CDE/CHR promoter elements in the regulation of cell cycle-dependent gene transcription. *FEBS J*, **277**, 877-893.
66. Yan, L., Donze, J.R. and Liu, L. (2005) Inactivated MGMT by O6-benzylguanine is associated with prolonged G2/M arrest in cancer cells treated with BCNU. *Oncogene*, **24**, 2175-2183.
67. Fantes, P. (1979) Epistatic gene interactions in the control of division in fission yeast. *Nature*, **279**, 428-430.
68. Nurse, P. (1975) Genetic control of cell size at cell division in yeast. *Nature*, **256**, 547-551.
69. Russell, P. and Nurse, P. (1986) cdc25+ functions as an inducer in the mitotic control of fission yeast. *Cell*, **45**, 145-153.
70. Stark, G.R. and Taylor, W.R. (2006) Control of the G2/M transition. *Mol Biotechnol*, **32**, 227-248.
71. Furnari, B., Rhind, N. and Russell, P. (1997) Cdc25 mitotic inducer targeted by chk1 DNA damage checkpoint kinase. *Science*, **277**, 1495-1497.
72. Peng, C.Y., Graves, P.R., Thoma, R.S., Wu, Z., Shaw, A.S. and Piwnica-Worms, H. (1997) Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science*, **277**, 1501-1505.
73. Sanchez, Y., Wong, C., Thoma, R.S., Richman, R., Wu, Z., Piwnica-Worms, H. and Elledge, S.J. (1997) Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science*, **277**, 1497-1501.
74. Atherton-Fessler, S., Parker, L.L., Geahlen, R.L. and Piwnica-Worms, H. (1993) Mechanisms of p34cdc2 regulation. *Mol Cell Biol*, **13**, 1675-1685.
75. Hagting, A., Karlsson, C., Clute, P., Jackman, M. and Pines, J. (1998) MPF localization is controlled by nuclear export. *EMBO J*, **17**, 4127-4138.
76. Taylor, W.R., DePrimo, S.E., Agarwal, A., Agarwal, M.L., Schonthal, A.H., Katula, K.S. and Stark, G.R. (1999) Mechanisms of G2 arrest in response to overexpression of p53. *Mol Biol Cell*, **10**, 3607-3622.

77. Vogelstein, B., Lane, D. and Levine, A.J. (2000) Surfing the p53 network. *Nature*, **408**, 307-310.
78. Smits, V.A., Klompaker, R., Vallenius, T., Rijksen, G., Makela, T.P. and Medema, R.H. (2000) p21 inhibits Thr161 phosphorylation of Cdc2 to enforce the G2 DNA damage checkpoint. *J Biol Chem*, **275**, 30638-30643.
79. Jin, S., Antinore, M.J., Lung, F.D., Dong, X., Zhao, H., Fan, F., Colchagie, A.B., Blanck, P., Roller, P.P., Fornace, A.J., Jr. *et al.* (2000) The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression. *J Biol Chem*, **275**, 16602-16608.
80. Chan, T.A., Hermeking, H., Lengauer, C., Kinzler, K.W. and Vogelstein, B. (1999) 14-3-3Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature*, **401**, 616-620.
81. Boronat, S. and Campbell, J.L. (2007) Mitotic Cdc6 stabilizes anaphase-promoting complex substrates by a partially Cdc28-independent mechanism, and this stabilization is suppressed by deletion of Cdc55. *Mol Cell Biol*, **27**, 1158-1171.
82. Crosby, M.E., Oancea, M. and Almasan, A. (2004) p53 binding to target sites is dynamically regulated before and after ionizing radiation-mediated DNA damage. *J Environ Pathol Toxicol Oncol*, **23**, 67-79.
83. Cam, H. and Dynlacht, B.D. (2003) Emerging roles for E2F: beyond the G1/S transition and DNA replication. *Cancer Cell*, **3**, 311-316.
84. Taylor, W.R., Schonthal, A.H., Galante, J. and Stark, G.R. (2001) p130/E2F4 binds to and represses the cdc2 promoter in response to p53. *J Biol Chem*, **276**, 1998-2006.
85. Plesca, D., Crosby, M.E., Gupta, D. and Almasan, A. (2007) E2F4 function in G2: maintaining G2-arrest to prevent mitotic entry with damaged DNA. *Cell Cycle*, **6**, 1147-1152.
86. Sugarman, J.L., Schonthal, A.H. and Glass, C.K. (1995) Identification of a cell-type-specific and E2F-independent mechanism for repression of cdc2 transcription. *Mol Cell Biol*, **15**, 3282-3290.
87. Benatti, P., Dolfini, D., Vigano, A., Ravo, M., Weisz, A. and Imbriano, C. (2011) Specific inhibition of NF-Y subunits triggers different cell proliferation defects. *Nucleic Acids Res*.
88. Imbriano, C., Gurtner, A., Cocchiarella, F., Di Agostino, S., Basile, V., Gostissa, M., Dobbstein, M., Del Sal, G., Piaggio, G. and Mantovani, R. (2005) Direct p53 transcriptional repression: in vivo analysis of CCAAT-containing G2/M promoters. *Mol Cell Biol*, **25**, 3737-3751.
89. Ceribelli, M., Alcalay, M., Vigano, M.A. and Mantovani, R. (2006) Repression of new p53 targets revealed by ChIP on chip experiments. *Cell Cycle*, **5**, 1102-1110.
90. Testa, A., Donati, G., Yan, P., Romani, F., Huang, T.H., Vigano, M.A. and Mantovani, R. (2005) Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J Biol Chem*, **280**, 13606-13615.
91. Caretti, G., Salsi, V., Vecchi, C., Imbriano, C. and Mantovani, R. (2003) Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. *J Biol Chem*, **278**, 30435-30440.

Chapter 4: Conclusions and Future Directions

There are a few key ideas that provided the vision for, and formed the core of this thesis. The first idea was the use of computational and systems biology techniques to study the complex biological components of the transcriptional response of human cells to DNA damaging agents. The design of the experimental system and the choice of the measurement techniques were key decisions that allowed the generation of a multi-dimensional and information-rich data set amenable to computational analyses. Moreover, the use of computational tools enabled us to identify a biologically meaningful transcriptional signature and further to propose a mechanism for the transcriptional control of a subset of that signature, both of which would be difficult to identify otherwise. The interweaving of experimental and computational techniques that enhanced and complemented one another was an essential part of this project. The experimental and computational techniques chosen at each point of the project were motivated by the biologically relevant question at that step (see Figure 4.1) and yielded more insight into the transcriptional response to DNA damaging agents. For example, once the transcriptional signature was identified, bioinformatics analysis was used to determine the functions enriched within the gene set, which in turn led to the experimental measurement of cell death and mitosis in the cell lines. Similarly, computational identification of NF-Y as the transcription factor possibly regulating the mitotic genes led to the experimental measurement of NF-Y promoter occupancy for these genes.

Another important idea that was explored in this thesis was the identification of a common signature for cell death/survival after exposure to a DNA damaging agent. Identifying genes with similar expression patterns in two separate cell lines (although in this case they were both lymphoblastoid cell lines) allowed us to exclude transcriptional changes that were specific only

to one particular cell line and not essential for BCNU sensitivity; this would not have been possible if only one sensitive or one resistant cell line were compared. Therefore, this project suggests that such common signatures specific for cell death could be identified. A crucial next step is to use a similar approach to identify transcriptional signatures for cell death across different DNA damaging agents.

The third important point in this thesis was the power of using an exploratory approach. Many of the genes that fall within the transcriptional signature have not been previously identified as affecting sensitivity to DNA damaging agents. Therefore, if we had performed this same study with preselected sets of genes or transcription factors, we would have missed the identification of the signature. Moreover, without the signature gene set, we would not have identified the transcription factor NF-Y as playing a role in DNA damage response.

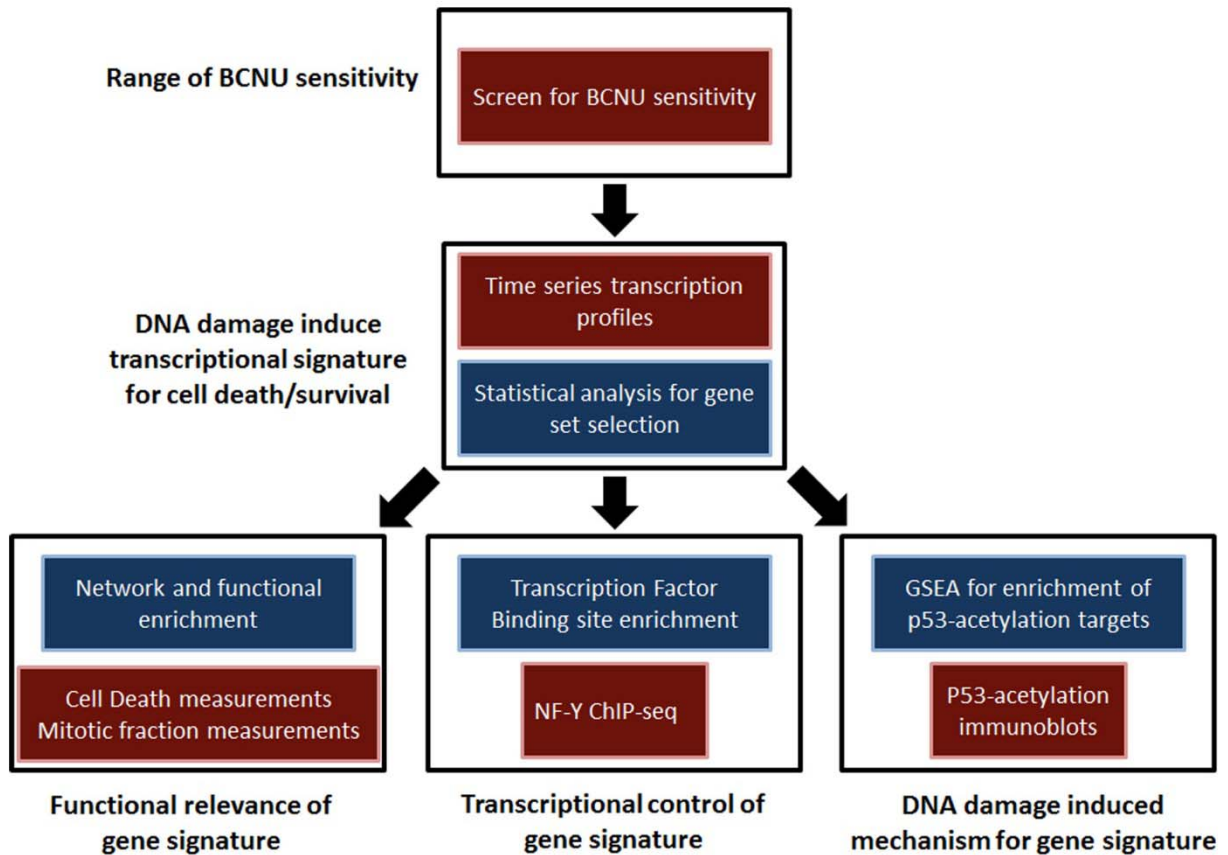


Figure 4.1: Experimental and computational techniques driven by biologically motivated questions
The experimental (red solid boxes) and computational (blue solid boxes) techniques used within this thesis are shown. The biologically relevant question that motivated these techniques are highlighted in black open boxes

The ideas, design and techniques implemented in this thesis yielded some important scientific contributions. The multi-well proliferation assay developed as part of this thesis is an efficient way to perform large scale screens to measure effects on proliferation of various types of drugs and treatments. Using this assay, high resolution measurements of survival can be obtained even for usually hard-to measure suspension cell lines. The assay enabled the otherwise impractical measurement of survival curves for the panel of 24 genetically diverse lymphoblastoid cell lines (grown in suspension) to BCNU sensitivity. More importantly, the technique has potential for use in a wide variety of projects, with a wide variety of cell lines and chemicals or toxic agents.

In the second part of the thesis, the transcriptional response after BCNU treatment in sensitive resistant cell lines was explored. This led to the identification of a transcriptional signature that suggested the activation of two important processes in the sensitive, BCNU treated cell lines; the induction of apoptosis and the stalling of entry into mitosis. Furthermore, the transcription signature correlated well with the observed cell phenotype which showed that these two processes were indeed occurring in the BCNU sensitive cell lines. This gene set was important in understanding the processes initiated in the sensitive cell lines after exposure to BCNU. Whether these genes behave similarly in response to other DNA damaging agents is not yet known and might be interesting to study.

From the transcriptional signature, NF-Y was identified as an important regulator for the transcriptional control of genes involved in the mitotic phase of the cell cycle after BCNU induced DNA damage. Although NF-Y has previously been implicated in the transcriptional repression of a few G2 genes, it had neither been observed in response to BCNU, nor to the extent observed here. Moreover, BCNU has been shown to induce a G2/M arrest in glioblastoma

cell lines, but the transcriptional control of mitotic entry and the role of NF-Y in such an arrest have not been previously identified. NF-Y mediated repression of a few mitotic genes has been observed in multiple cell lines treated with different types of DNA damaging agents. This suggests that the control of G2 arrest by NF-Y could potentially be a general mechanism used by cells after DNA damage.

The control of the G2/M checkpoint is important in the treatment of tumors since many of the DNA damaging agents used in the clinic induce a G2/M arrest. This arrest is thought to act as a protective mechanism, allowing time for the tumor cells to repair damage. Our results suggest that NF-Y could play an important role in the protective mechanisms employed by tumors against chemotherapy. However, our results also showed that the induction of apoptotic genes can overshadow the protective effects of a G2 arrest since the G2 arrest we observed in our sensitive cell lines did not ultimately protect them from BCNU induced cell death. Moreover, there was no G2 arrest observed in the resistant cell lines confirming that a G2 arrest is not required for resistance to DNA damaging agents and that other mechanisms such as induction of DNA repair proteins could also confer resistance, as has been seen in the clinic. These results suggest to some extent, an ordering of the effects of DNA repair, induction of apoptosis and cell cycle arrest on the cells' survival after exposure to DNA damaging agents. These ideas could be exploited for use in combination therapy in the clinic, where multiple effective strategies such as inhibition of repair and cell cycle arrest and acceleration of apoptosis might sensitize some resistant tumors.

To this end, understanding the mechanism of NF-Y mediated repression in the control of the G2/M arrest after DNA damage could provide possible avenues for enhanced or more efficient chemotherapy. p53 acetylation has been suggested as a possible mechanism that controls the repression of NF-Y targets after DNA damage. However, the lack of p53 acetylation observed in the sensitive cell line after BCNU exposure could suggest that alternate mechanisms are more important for NF-Y repression in these lymphoblastoid cell lines than what has been studied so far. Other mechanisms for NF-Y regulation have also been suggested, including post-translational modification of NF-Y and changes in chromatin structure. It would be interesting to explore which, if any, of these mechanisms are responsible for the observed NF-Y repression seen in BCNU treated cells. Interactions of NF-Y with varying partners and protein complexes have also been suggested as a method for the transcriptional control of NF-Y target genes. The identity of these interacting partners and protein complexes both in the absence and presence of DNA damage could also shed some light on other mechanisms of NF-Y mediated transcriptional control.

As a result of this thesis, experimental and computational techniques suitable for extracting general gene signatures for cell death/survival after DNA damage were identified. The next step would be to apply these techniques to other DNA damaging agents to determine whether a drug independent signature for cell death can be extracted using this technique. Such a study would, for example, reveal whether the G2 arrest controlled by NF-Y in BCNU treated cells is activated or important in the response to other chemotherapeutic agents. Preliminary results for the sensitivity of the BCNU sensitive/resistant cell lines to other DNA damaging agents used in the clinic are shown in Figure 2. The results show that the cell lines have different sensitivities to

Temozolomide, but not to cisplatin and ionizing radiation. However, it is yet remains to be seen whether the panel of twenty four cell lines will indeed show a range of sensitivities to IR and cisplatin, and whether a G2 arrest is induced in treated cells.

If NF-Y dependent transcription is observed in cell lines after treatment with the other DNA damaging agents, this would be an important step forward in understanding the common mechanisms for cell death/survival after DNA damage. The presence of such a common signature that is not dependent on the DNA damaging agent, and is unaffected by genetic variation across a population, would have the potential to improve current modes of chemotherapy by sensitizing treatment resistant tumors or by protecting sensitive non-tumor cells.

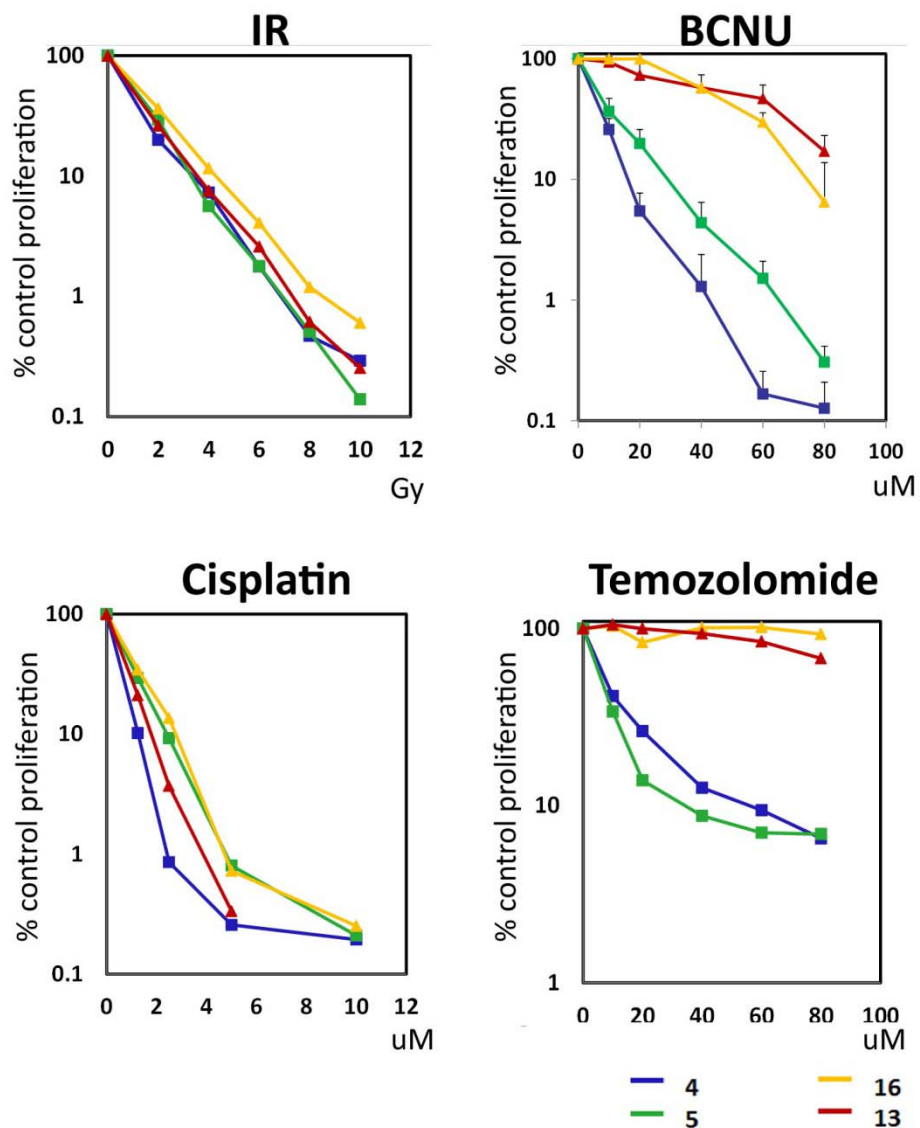


Figure 4.2: Survival curves for cell lines 4, 5, 13 and 16 for four clinically used cancer treatments
 Survival curves for cell lines 4 (blue), 5 (green), 13 (red) and 16 (yellow), when treated with ionizing radiation (IR), BCNU, Cisplatin and Temozolomide as measured using the proliferation assay.

Appendix A:

Genomic Predictors of Interindividual Differences in Response to DNA Damaging Agents

Contribution: Partial Least Squares Regression analysis and statistical analyses for significance of the identified gene set.

RESEARCH COMMUNICATION

Genomic predictors of interindividual differences in response to DNA damaging agents

Rebecca C. Fry,^{1,2,3,4,6} J. Peter Svensson,^{1,2,6}
Chandni Valiathan,^{1,2,3,6} Emma Wang,^{1,2}
Brad J. Hogan,^{1,2} Sanchita Bhattacharya,^{2,7}
James M. Bugni,^{1,2} Charles A. Whittaker,^{2,3}
and Leona D. Samson^{1,2,3,4,5,8}

¹Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ²Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ³Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ⁴Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Human lymphoblastoid cells derived from different healthy individuals display considerable variation in their transcription profiles. Here we show that such variation in gene expression underlies interindividual susceptibility to DNA damaging agents. The results demonstrate the massive differences in sensitivity across a diverse cell line panel exposed to an alkylating agent. Computational models identified 48 genes with basal expression that predicts susceptibility with 94% accuracy. Modulating transcript levels for two member genes, *MYH* and *C21ORF56*, confirmed that their expression does indeed influence alkylation sensitivity. Many proteins encoded by these genes are interconnected in cellular networks related to human cancer and tumorigenesis.

Supplemental material is available at <http://www.genesdev.org>.

Received April 24, 2008; revised version accepted August 7, 2008.

The interindividual differences in disease susceptibility, responsiveness to chemotherapeutics, and susceptibility to environmental exposures across human populations are influenced by a combination of gene-environment interactions. Over the past few years, studies aimed at dissecting the genetic basis underlying human phenotypic variation have built off of the dense genotyping

[*Keywords*: DNA damage susceptibility; alkylating agents; prediction]

⁶These authors contributed equally to this work.

⁷Present address: Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

⁸Corresponding author.

E-MAIL lsamson@mit.edu; FAX (617) 253-8099.

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.1688508>. Freely available online through the *Genes & Development* Open Access option.

established by the HapMap Consortium. A wealth of genome-wide association studies (GWAS) have described how human genetic variation, at the level of single nucleotide differences, is linked to such complex diseases as diabetes and breast cancer (Hunter et al. 2007; Zeggini and McCarthy 2007). In addition, GWAS have also linked DNA polymorphic variants to gene expression variation across populations (Cheung et al. 2005; Stranger et al. 2005, 2007; Dixon et al. 2007).

However, while it is known that human lymphoblastoid cells derived from different healthy individuals display considerable variation in their transcription profiles (Cheung et al. 2003, 2005; Stranger et al. 2005, 2007; Dixon et al. 2007), the influence this variation has on the response to environmental and chemotherapeutic agents is unknown. In this study, a panel of 24 cell lines previously derived from unrelated, healthy individuals with diverse ancestry (Collins et al. 1998) was tested for variation in sensitivity to the DNA damaging agent, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG). MNNG induces a variety of alkylated DNA bases, among which *O*⁶-methylguanine (*O*⁶MeG) is known to be particularly toxic as well as mutagenic because it pairs with thymine during replication. *O*⁶MeG can be repaired by the MGMT DNA repair methyltransferase (Pegg 1990, 2000), but left unrepaired, the ensuing *O*⁶MeG:T base pair can be processed by the DNA mismatch repair (MMR) pathway, and such processing actually triggers apoptotic cell death and cytotoxicity (Kaina et al. 1997; Hickman and Samson 1999, 2004). Therefore, cells deficient in MGMT but proficient for MMR are extremely sensitive to MNNG-induced killing, whereas cells deficient in both MGMT and MMR are extremely resistant or tolerant to MNNG, but at the cost of increased mutation (Karran 2001). While MGMT and MMR status are thus known to be associated with alkylation sensitivity, we questioned whether the expression level of these two repair pathways is sufficient to explain interindividual variation in alkylation sensitivity.

We show that there is extensive interindividual variation in the response of cell lines derived from a healthy, genetically diverse population upon exposure to the DNA alkylating agent MNNG. The differences in susceptibility to MNNG were associated with variation in gene expression to identify genomic predictors of cellular sensitivity. A set of 48 genes was identified that can predict, with a remarkable 94% accuracy, differences in cellular sensitivity to MNNG in a test population. The basal gene expression, rather than MNNG-treated expression level, was found to be the better predictor of cellular sensitivity. To validate the computational models, the expression level of two members of the predictive gene set, *C21ORF56* and *MYH*, were modulated and shown to alter cellular sensitivity to MNNG. These findings may have profound implications in the clinical setting, where the collective set of 48 genes may be used as predictors and modulators of cellular sensitivity to cancer chemotherapeutics.

Results and Discussion

To assess the range of interindividual differences in sensitivity to a DNA alkylating agent, we used a high-throughput growth inhibition assay (percentage of con-

Fry et al.

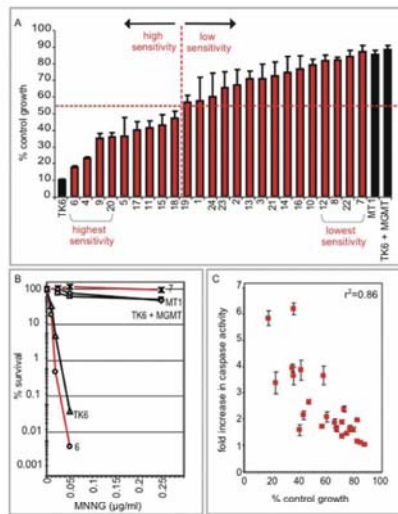


Figure 1. A considerable range of interindividual sensitivity to a DNA alkylating agent. (A) The percentage of control growth of the cell lines at 72 h after treatment with MNNG (0.5 µg/mL) using a growth inhibition assay. The division between high and low sensitivity among the cell lines is demarcated at 53% with a red dotted line. (B) The percentage of survival of cell lines 6 and 7 was determined 10 d after treatment with MNNG and compared with three control cell lines (TK6, TK6 + MGMT, and MT1) using a killing curve assay. (C) Fold increase in caspase-3 activity was determined 72 h post-treatment with MNNG across the cell line panel.

control growth) across the panel of cell lines. Control cell lines with very high MNNG sensitivity (TK6) or very low sensitivity (TK6 + MGMT, MT1) were included (Kat et al. 1993). The panel of cell lines displayed MNNG sensitivities spanning the entire range between the control cell lines (Fig. 1A). These large differences were even more apparent when measured by the lower throughput killing curve assay that has a greater dynamic range (Fig. 1B). Importantly, the MNNG sensitivities of the cell lines are not associated with individual differences in growth characteristics (e.g., cell doubling time) (Supplemental Fig. S1). We also monitored MNNG-induced apoptosis in the cell lines (by caspase-3 activation) and found a positive correlation with MNNG sensitivity (Fig. 1C). Thus, the growth inhibition, survival, and apoptosis assays each underscore the extensive range of interindividual responses to MNNG among genetically diverse cells.

To determine whether transcriptional profiles could predict cellular response to MNNG, a two-class prediction algorithm was applied to the gene expression profiles of the 24 cell lines under both untreated (basal) and MNNG-treated conditions. The cell lines were divided into two classes with either high sensitivity or low sensitivity, with a cut point of 53% control growth based on the midpoint between the most sensitive and least sensitive cell line (Fig. 1A, lines 6,7). For the two-class prediction model, a training population composed of the four most sensitive and the four least sensitive cell lines was selected and analyzed to identify genes that were not only differentially expressed but also showed significant positive or negative correlation with increasing MNNG sensitivity. Three alkylation sensitivity-as-

sociated (ASA) gene sets were identified as follows: (1) a set of 48 genes derived from basal gene expression (the BASA set), (2) a set of 39 genes derived from treatment-to-basal expression ratio (the TRASA set), and (3) a set of 121 genes derived from treatment-induced expression (the TASA set) (Supplemental Tables 1–3). The expression patterns of the ASA gene sets across the training population and the test population (i.e., the 16 cell lines not included in the training population) are visualized in Figure 2A.

The ability of the three ASA gene sets to predict

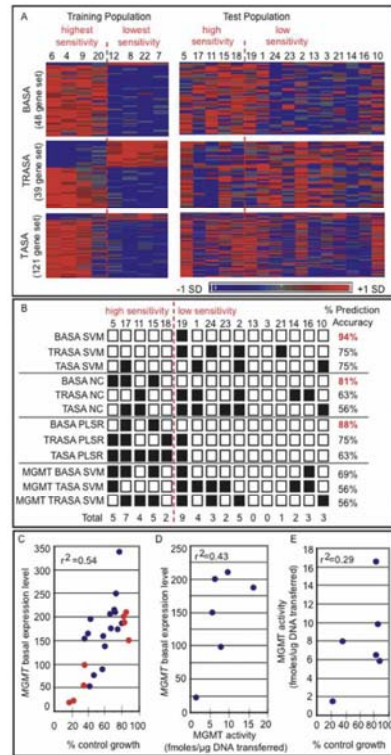


Figure 2. Identification of ASA genes that predict interindividual differences in alkylation sensitivity. (A) Three ASA gene sets were identified from a training population comprising the four most sensitive and the four least sensitive cell lines including (1) 48 genes derived from basal gene expression (the BASA set), (2) 39 genes derived from treatment-to-basal expression ratio (the TRASA set), and (3) 121 genes derived from treatment-induced expression (the TASA set). Expression patterns for the gene sets are shown for both the training and the test populations of cell lines. Expression values are mean centered with high relative expression indicated in red and low relative expression indicated in blue. (B) The sensitivity of the test population of cell lines to MNNG was predicted using three algorithms: SVM, NC, and PLSR. The two-class prediction algorithms were used with each of the three ASA gene sets as well as the MGMT transcript alone. Correct prediction is indicated with a white box; incorrect prediction, with a black box. (C) MGMT expression level is plotted versus the percentage of control growth of the cell lines treated with MNNG. Red circles indicate cell lines of the training population, and blue circles indicate cell lines of the test population. (D) *O*⁶-MeG DNA methyltransferase activity was determined in protein extracts derived from cell lines 4, 7, 20, 12, 22, and 8. Methyltransferase activity is plotted versus the baseline expression level of MGMT in each cell line. (E) Methyltransferase activity is plotted versus the percentage of control growth for the same cell lines as in (D).

MNNG sensitivity was assessed using the support vector machine (SVM) algorithm. The TRASA and TASA gene sets predicted the MNNG sensitivity of the test population with 75% accuracy (Fig. 2B). Remarkably, the BASA gene set accurately predicted sensitivity in 15 of 16 cell lines (94% accuracy), with only cell line 19 misclassified (Fig. 2B); note that cell line 19 falls on the boundary of the cut point between high and low sensitivity (Fig. 1A). To validate the SVM results, we applied two other prediction algorithms, namely, a nearest centroid (NC) and a partial least squares regression (PLSR) model (Fig. 2B). For all algorithms, the BASA gene set provided maximal prediction of MNNG sensitivity, with SVM providing the highest accuracy (Fig. 2B). The lack of prediction of the TRASA and TASA gene sets may be a result of time point selection, a feature that a more global assessment of temporal responses would potentially capture. That basal gene expression is the most accurate predictor of alkylation sensitivity bodes well for translating these findings to a clinical setting; for example, to predict whether a tumor will respond to alkylation chemotherapy.

The BASA gene set contained two genes that showed positive association of expression with lower MNNG sensitivity; namely, *MGMT* and the *C21ORF56* (Supplemental Table S1). *MGMT* efficiently repairs *O*⁶MeG (Pegg 1990, 2000), and its activity is known to vary among individuals (Vahakangas et al. 1991; Margison et al. 2003). Likewise, we identified considerable variation in the expression level of *MGMT* across the 24 cell lines, and we demonstrate a positive association of *MGMT* expression with lower MNNG sensitivity (Fig. 2C). Although *MGMT* expression level was positively associated with *MGMT* activity (Fig. 2D), and activity was positively associated with MNNG resistance (Fig. 2E), the correlations were relatively weak. This likely explains why *MGMT* expression alone is not as strong a predictor of alkylation sensitivity as the set of 48 transcripts together (Fig. 2B).

MGMT silencing is currently being used as a prognostic indicator of successful alkylation chemotherapy for glioblastoma (Hegi et al. 2005); our results suggest that expression levels for the 48 genes described here may prove a more accurate indicator. It should be mentioned that *MGMT* is the only member of the BASA gene set previously known to influence alkylation sensitivity. Interestingly, upon testing each member of the BASA gene set independently for predictive capacity, genes with equal or greater accuracy than *MGMT* were identified (Supplemental Table S1). However, while some genes showed higher prediction accuracy than *MGMT*, it was the collective set of 48 transcripts that provided the maximal prediction of 94% accuracy. The probability of 48 transcripts chosen at random from the pool of expressed genes predicting MNNG sensitivity with such accuracy is <0.002.

In addition to *MGMT*, the only other member of the BASA gene set with positive association of expression with lower MNNG sensitivity was *C21ORF56* (Supplemental Table S1). The *C21ORF56* protein exists in two isoforms that are highly conserved across mammals and show homology with SPATC1 (spermatogenesis and centriole-associated protein 1) (Supplemental Fig. S2A,B). Supporting our finding of the variation in the expression level of *C21ORF56* in these cell lines, its expression has been documented as highly variable

across four separate populations of cell lines. Specifically, *C21ORF56* showed variation in expression level across three of the four HapMap populations, namely, the populations of European (CEU), Chinese (CHB), and Japanese (JPT) origin (Stranger et al. 2005, 2007) and in a separate population of British descent (Dixon et al. 2007).

Given the significant association of *C21ORF56* gene expression with MNNG sensitivity across the cell line panel, our analysis suggested that *C21ORF56*, like *MGMT*, might play a role in protecting cells against MNNG-induced killing. To confirm the role of *C21ORF56* in modulating alkylation sensitivity, we show that TK6 cells with 80% knockdown of the *C21ORF56* transcript show greatly increased MNNG sensitivity relative to control cells expressing a nontargeting control shRNA (Fig. 3A). In addition, using an alternate *C21ORF56*-targeting shRNA, an empty vector control, and individually established clones with differential expression of *C21ORF56*, we show that decreased expression of *C21ORF56* is associated with increased sensitivity to MNNG (Supplemental Fig. S3). These results validate the computational prediction that *C21ORF56* expression modulates the response of human cells to alkylating agents.

The gene with the most significant positive association of expression with alkylation sensitivity in the BASA gene set was *MYH* (Supplemental Table S1), a DNA glycosylase that initiates base excision repair by removing adenines mispaired opposite 8-oxoguanine lesions (Slupska et al. 1996; Parker and Eshleman 2003) and not previously known to modulate alkylation sensitivity. In support of our finding of high variance of *MYH* expression across a genetically diverse population, its expression levels were found to vary among individuals of a separate population (Dixon et al. 2007). Importantly, however, to date no link between the variation in expression level of *MYH* and interindividual differences in sensitivity upon exposure to a DNA alkylating agent has been made.

As *MYH* expression was higher in the more sensitive cells, we hypothesized that *MYH* deficiency might con-

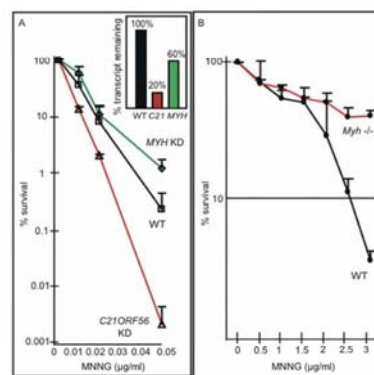


Figure 3. Modulation of *C21ORF56*, *MYH*, and *Myh* influences MNNG sensitivity. (A) TK6 cells expressing a control shRNA (WT) or shRNA specifically targeting the *C21ORF56* transcript (v1), or the *MYH* transcript (v1), were assessed for the percentage of survival after exposure to MNNG. The inset shows the percentage of transcript remaining *C21ORF56* and *MYH* in knockdown cells. (B) Percentage of survival of *Myh*^{-/-} or wild-type MEFs determined after treatment with MNNG.

Fry et al.

fer resistance, in much the same way that MMR deficiency confers MNNG resistance (Kaina et al. 1997; Karran 2001). This was tested by decreasing the expression of *MYH* in TK6 cells. The *MYH* knockdown cells were indeed significantly less sensitive than the parent TK6 cells expressing a control shRNA to cell killing (Fig. 3A). In addition, using an alternate *MYH*-targeting shRNA, an empty vector control, and individually established clones with differential expression of *MYH*, we show that increased expression of *MYH* is associated with MNNG sensitivity (Supplemental Fig. S3). Importantly, the influence of *Myh* in modulating MNNG sensitivity was further established in mouse cells by showing that *Myh*^{-/-} mouse embryonic fibroblasts (MEFs) were much less sensitive than wild type to MNNG-induced cell killing (Fig. 3B). These results indicate that *MYH* expression indeed correlates with increased alkylation sensitivity. Likewise, *MYH* expression was also significantly higher in the alkylation-sensitive TK6 cell line relative to the much less sensitive MT1 cell line (Fig. 1A,B; data not shown) compatible with the trend seen across the panel of genetically diverse cell line (Fig. 1A). Interestingly, although not known to repair damage induced by MNNG, *MYH* is known to interact with the MutS α heterodimer (Gu et al. 2002) that binds O⁶MeG mispairs in DNA to initiate the triggering of apoptotic cell death (Ceccotti et al. 1996; Hickman and Samson 2004). It remains to be determined whether *MYH* influences alkylation sensitivity via its interaction with the MMR machinery. Finally, it was surprising that no MMR transcripts were represented in the BASA gene set. It turned out that the differences in expression for *MSH2*, *MSH6*, and *MLH1* did not exceed 1.5-fold, eliminating them from the gene set; however, all three transcripts were in fact higher in the most sensitive versus the least sensitive cell line (data not shown).

To gain a more comprehensive view of the various pathways that influence alkylation sensitivity, all genes that were differentially expressed under basal conditions between cell lines with the highest and lowest MNNG sensitivity (this time not imposing a requirement for trend significance) were analyzed for network properties (Fig. 4A; Supplemental Table S4). For the 240 genes identified as differentially expressed, 148 gene products are found in the Ingenuity database, and a remarkable ~85% of these are contained in a single significant interacting network ($P < 10^{-10}$) (Fig. 4B; Supplemental Tables S5, S6). It thus seems that proteins likely to play a role in determining interindividual differences in alkylation sensitivity are highly connected, a phenomenon previously observed in *Saccharomyces cerevisiae* (Begley

et al. 2004; Said et al. 2004). Contained within the large interacting network are subnetworks that integrate 18 BASA transcripts (including *MGMT* and *MYH*), and these subnetworks are enriched for proteins that are associated with tumorigenesis and cancer predisposition (Fig. 4C,D; Supplemental Tables S1, S6). In general, the expression level of the tumorigenesis-associated transcripts showed elevated basal expression in the cell lines with high MNNG sensitivity. Finally, by analyzing the promoter regions of the ASA genes for enriched transcription factor-binding sites, we find evidence for a common regulatory factor, namely, the octamer-binding transcription factor, Oct-1 (Supplemental Table S7). Oct-1 is known to respond to DNA alkylation damage (Zhao et al. 2000) and is a known regulator of stress responses (Tantin et al. 2005). Here we find Oct-1-binding sites significantly enriched in transcripts with elevated basal expression in cell lines with high MNNG sensitivity as well as in tumorigenesis-associated proteins that show higher basal expression in cells with high sensitivity to MNNG ($P = 2.05 \times 10^{-5}$) (Fig. 4D).

To conclude, our findings may have profound implications in the clinical setting, where the expression of the 48 transcripts encompassing *MGMT*, *C21ORF56*, *MYH*, and many others may not only predict interindividual responses to alkylating agents but could be modulated to affect cancer treatment response. Furthermore, as cell lines derived from different individuals indeed

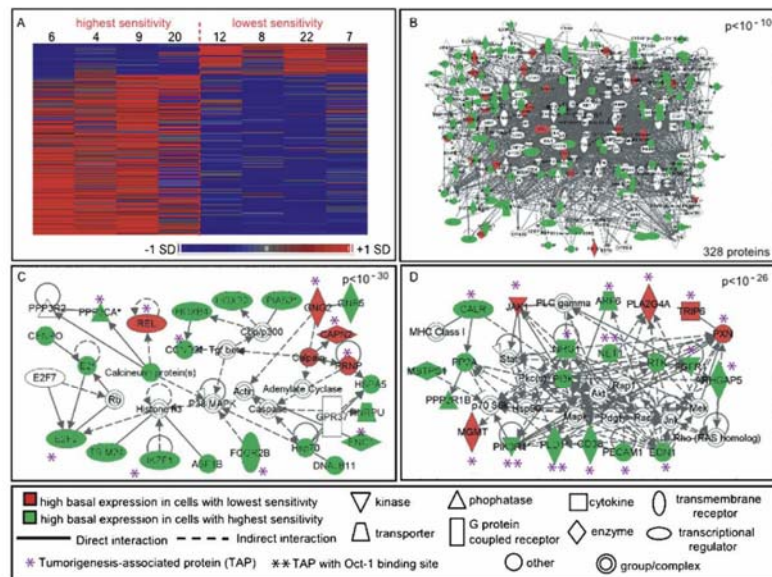


Figure 4. Basal expression networks associated with interindividual differences in sensitivity to MNNG. (A) A heat map of the 240 basally differentially expressed genes identified between two classes of the training population, those with highest and lowest MNNG sensitivity. Expression values are mean centered with high relative expression indicated in red and low relative expression indicated in blue. (B) Of the 240 genes from A, 148 were present in the Ingenuity database. These 148 proteins were analyzed for significant enrichment of molecular interactions. A significant ($P < 10^{-10}$) interactome of 328 total proteins containing 125 of the 148 proteins was identified. (C) The most significant subnetwork ($P < 10^{-30}$) of ASA proteins. (D) The second most significant subnetwork ($P < 10^{-26}$) of ASA proteins. Proteins in red are encoded by transcripts with high basal expression in cells with low MNNG sensitivity, and proteins in green are encoded by transcripts with high basal expression in cells with high MNNG sensitivity; proteins in white are associated with these ASA proteins. Tumorigenesis-associated proteins (TAPs) are indicated with an asterisk and TAP proteins containing Oct-1-binding sites are indicated with two asterisks.

preserve genetic diversity at the level of gene expression (Cheung et al. 2003; Correa and Cheung 2004; Morley et al. 2004), they serve as an ideal tool for establishing interindividual differences in DNA damage responses. We propose that upon exposure of these cell lines to other environmental toxicants and cancer chemotherapeutics we will discover more genes of hitherto unknown function responsible for interindividual differences in sensitivity to DNA damaging agents.

Materials and methods

Cell line panel, drug treatment, and RNA extraction

The 24 lymphoblastoid cell lines, established using EBV transformation, were obtained from the Coriell Institute (New Jersey) and were designated 1–24 as follows: #1 (GM15029), #2 (GM15036), #3 (GM15215), #4 (GM15223), #5 (GM15245), #6 (15,224), #7 (GM15236), #8 (GM15510), #9 (GM15213), #10 (GM15221), #11 (GM15227), #12 (GM15385), #13 (GM15590), #14 (GM15038), #15 (GM15056), #16 (GM15072), #17 (GM15144), #18 (GM15216), #19 (GM15226), #20 (GM15242), #21 (GM15268), #22 (GM15324), #23 (GM15386), and #24 (GM15061). Percentage of control growth was measured using logarithmically growing cells. Cells were treated with 0.5 $\mu\text{g}/\mu\text{L}$ MNNG or untreated, and viable cells were counted 72 h after treatment using a coulter counter coupled with trypan blue staining (total number of viable treated cells/total number of viable untreated cells). Percentage of survival was determined using a killing curve assay counted 10 d after treatment with MNNG as described (Furth et al. 1981). Total RNA was isolated from log phase cells according to the mammalian cell protocol (Qiagen) and labeled according to the Affymetrix protocol. RNA was hybridized to HGU133 Plus 2.0 full genome human arrays in technical duplicate totaling 96 arrays.

Microarray data analysis

Data were normalized using a PLIER algorithm and filtered for nonexpressed transcripts across all arrays as described in Fry et al. (2007), resulting in a reduction of the probesets from the original 54,675 to 19,290. ASA gene sets were determined as follows. The four cell lines with highest MNNG sensitivity and four cell lines with lowest MNNG sensitivity were used as a training population. For the ASA sets, genes with differential expression between the two groups were identified with (1) significant fold change (≥ 1.5 or ≤ -1.5 , $P < 0.05$ *t*-test), and (2) a significant positive or negative trend for association of gene expression with increasing growth inhibition (percentage of control growth) using a correlation measurement ($r \geq 0.7$ or ≤ -0.7 , $P < 0.01$ Trend) calculated using a linear regression model in S-PLUS 7.0 (<http://www.insightful.com>). For the fourth gene set (Fig. 4), a *P* for trend was not imposed. Three methods for two-class prediction were used, including SVM algorithm carried out using Gene Pattern Software (version 2.0.1) (<http://www.broad.mit.edu>), the NC algorithm based in R, and PLSR analysis programmed in MATLAB (Mathworks, Inc.) and adapted from Geladi et al. (1996). Network and gene ontology analysis was performed using Ingenuity software (<http://www.ingenuity.com>). Cancer and tumorigenesis-associated proteins were identified using the Ingenuity database as well as the Genomica module analysis (<http://genie.weizmann.ac.il>). Transcription factor-binding site analysis was performed as described (Fry et al. 2007) using the EXPANDER program. Microarray data have been deposited in the Gene Expression Omnibus repository (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE10313.

Sequence analysis

Orthologs of human C21ORF56 [NM_032261] were identified by comparing *Homo sapiens* protein sequence to those existing sequences currently available in the University of California at Santa Cruz Genome Browser (<http://genome.ucsc.edu>). Sequences were aligned using ClustalX.

Caspase 3 activation

After treatment with MNNG, caspase-3 activity was measured using a caspase activity assay kit (Promega). Cells were resuspended in PBS and incubated for 45 min with the proluminescent caspase-3 substrate con-

taining the DEVD sequence. Luminescence was measured with a luminometer.

MGMT activity assay

Lymphoblastoid cells in log phase growth were resuspended in MTase buffer [50 mM HEPES at pH 7.8, 10 mM DTT, 1 mM EDTA, 5% glycerol], sonicated, and lysate cleared by centrifugation. Protein concentration was determined using the Pierce Better Bradford assay. MTase activity for each cell line was determined using calf thymus DNA methylated in vitro with [³H] MNU as described (Glassner et al. 1999).

MEF colony forming assay

MEFs were cultured in DMEM media (10% fetal bovine serum, 1% penicillin, and streptomycin). Cells were seeded at 200 cells per 10 mL of culture in 100-mm dishes. Media was replaced 24 h after seeding, and cells were preincubated with *O*⁶-benzylguanine (10 mM) for 2 h. Cells were then treated with MNNG (0–3 $\mu\text{g}/\mu\text{L}$) and colonies counted 5 d after treatment.

shRNA knockdown cell line generation and treatment

shRNAs expressed in a lentiviral plasmid (pGIPZ) were purchased from Open Biosystems to target the *C21ORF56* transcript (v1:#RHS4430-98844079 and v2:#RHS4430-98477469) or the *MYH* (v1:#RHS4430-98904053 and v2:#RHS4430-99140608) transcript. Knockdown cells were compared with TK6 cells expressing a nontargeting shRNA (#RHS4346) or an empty vector control (#RHS4349). Virus was generated in 293T cells using packaging plasmids (psPAX2, pMD2.G Addgene). The TK6 cell line was infected with virus and stable clones selected using Puromycin. The percentage of survival was measured using a killing curve assay 14 d after treatment with MNNG as described (Furth et al. 1981).

Acknowledgments

We are grateful to Margherita Bignami for the generous gift of the *Myh*^{-/-} MEFs. This work was supported by the National Institute of Environmental Health Sciences (ES11399 and ES002109) and the National Cancer Institute (CA055042 and U54-CA112967). J.P.S. was supported by a Swedish Research Council Fellowship. C.V. received support from NIH training grant 1-R90-DK071503-03. L.D.S. is an American Cancer Society Research Professor.

References

- Begley, T.J., Rosenbach, A.S., Ideker, T., and Samson, L.D. 2004. Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. *Mol. Cell* **16**: 117–125.
- Ceccotti, S., Aquilina, G., Macpherson, P., Yamada, M., Karran, P., and Bignami, M. 1996. Processing of *O*⁶-methylguanine by mismatch correction in human cell extracts. *Curr. Biol.* **6**: 1528–1531.
- Cheung, V.G., Conlin, L.K., Weber, T.M., Arcaro, M., Jen, K.Y., Morley, M., and Spielman, R.S. 2003. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* **33**: 422–425.
- Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., and Burdick, J.T. 2005. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**: 1365–1369.
- Collins, F.S., Brooks, L.D., and Chakravarti, A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**: 1229–1231.
- Correa, C.R. and Cheung, V.G. 2004. Genetic variation in radiation-induced expression phenotypes. *Am. J. Hum. Genet.* **75**: 885–890.
- Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., et al. 2007. A genome-wide association study of global gene expression. *Nat. Genet.* **39**: 1202–1207.
- Fry, R.C., Navasumrit, P., Valiathan, C., Svensson, J.P., Hogan, B.J., Luo, M., Bhattacharya, S., Kandjanapa, K., Soontararuks, S., Nookabkaew, S., et al. 2007. Activation of inflammation/NF- κ B signaling in infants born to arsenic-exposed mothers. *PLoS Genet.* **3**: e207. doi: 10.1371/journal.pgen.0030207.
- Furth, E.E., Thilly, W.G., Penman, B.W., Liber, H.L., and Rand, W.M. 1981. Quantitative assay for mutation in diploid human lympho-

Fry et al.







- blasts using microtiter plates. *Anal. Biochem.* **110**: 1–8.
- Geladi, P., Martens, H., Hadjiski, L., and Hopke, P.A. 1996. A calibration tutorial for spectral data. Part 2. Partial least squares regression using Matlab and some neural network results. *J. Near Infrared Spectrosc.* **4**: 243–255.
- Glassner, B.J., Weeda, G., Allan, J.M., Broekhof, J.L., Carls, N.H., Donker, I., Engelward, B.P., Hampson, R.J., Hersmus, R., Hickman, M.J., et al. 1999. DNA repair methyltransferase (Mgmt) knockout mice are sensitive to the lethal effects of chemotherapeutic alkylating agents. *Mutagenesis* **14**: 339–347.
- Gu, Y., Parker, A., Wilson, T.M., Bai, H., Chang, D.Y., and Lu, A.L. 2002. Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. *J. Biol. Chem.* **277**: 11135–11142.
- Hegi, M.E., Diserens, A.C., Gorlia, T., Hamou, M.F., de Tribolet, N., Weller, M., Kros, J.M., Hainfellner, J.A., Mason, W., Mariani, L., et al. 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352**: 997–1003.
- Hickman, M.J. and Samson, L.D. 1999. Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proc. Natl. Acad. Sci.* **96**: 10764–10769.
- Hickman, M.J. and Samson, L.D. 2004. Apoptotic signaling in response to a single type of DNA lesion, O⁶-methylguanine. *Mol. Cell* **14**: 105–116.
- Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., et al. 2007. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**: 870–874.
- Kaina, B., Ziouta, A., Ochs, K., and Coquerelle, T. 1997. Chromosomal instability, reproductive cell death and apoptosis induced by O⁶-methylguanine in Mex⁻, Mex⁺ and methylation-tolerant mismatch repair compromised cells: Facts and models. *Mutat. Res.* **381**: 227–241.
- Karran, P. 2001. Mechanisms of tolerance to DNA damaging therapeutic drugs. *Carcinogenesis* **22**: 1931–1937.
- Kat, A., Thilly, W.G., Fang, W.H., Longley, M.J., Li, G.M., and Modrich, P. 1993. An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair. *Proc. Natl. Acad. Sci.* **90**: 6424–6428.
- Margison, G.P., Povey, A.C., Kaina, B., and Santibanez Koref, M.F. 2003. Variability and regulation of O⁶-alkylguanine-DNA alkyltransferase. *Carcinogenesis* **24**: 625–635.
- Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**: 743–747.
- Parker, A.R. and Eshleman, J.R. 2003. Human MutY: Gene structure, protein functions and interactions, and role in carcinogenesis. *Cell. Mol. Life Sci.* **60**: 2064–2083.
- Pegg, A.E. 1990. Mammalian O⁶-alkylguanine-DNA alkyltransferase: Regulation and importance in response to alkylating carcinogenic and therapeutic agents. *Cancer Res.* **50**: 6119–6129.
- Pegg, A.E. 2000. Repair of O⁶-alkylguanine by alkyltransferases. *Mutat. Res.* **462**: 83–100.
- Said, M.R., Begley, T.J., Oppenheim, A.V., Lauffenburger, D.A., and Samson, L.D. 2004. Global network analysis of phenotypic effects: Protein networks and toxicity modulation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **101**: 18006–18011.
- Slupska, M.M., Baikalov, C., Luther, W.M., Chiang, J.H., Wei, Y.F., and Miller, J.H. 1996. Cloning and sequencing a human homolog (hMYH) of the *Escherichia coli* mutY gene whose function is required for the repair of oxidative DNA damage. *J. Bacteriol.* **178**: 3885–3892.
- Stranger, B.E., Forrest, M.S., Clark, A.G., Minichiello, M.J., Deutsch, S., Lyle, R., Hunt, S., Kahl, B., Antonarakis, S.E., Tavare, S., et al. 2005. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**: e78. doi: 10.1371/journal.pgen.0010078.
- Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., et al. 2007. Population genomics of human gene expression. *Nat. Genet.* **39**: 1217–1224.
- Tantin, D., Schild-Poulter, C., Wang, V., Hache, R.J., and Sharp, P.A. 2005. The octamer binding transcription factor Oct-1 is a stress sen-
- sor. *Cancer Res.* **65**: 10750–10758.
- Vahakangas, K., Trivers, G.E., Plummer, S., Hayes, R.B., Krokan, H., Rowe, M., Swartz, R.P., Yeager Jr., H., and Harris, C.C. 1991. O(6)-methylguanine-DNA methyltransferase and uracil DNA glycosylase in human broncho-alveolar lavage cells and peripheral blood mononuclear cells from tobacco smokers and non-smokers. *Carcinogenesis* **12**: 1389–1394.
- Zeggini, E. and McCarthy, M.I. 2007. Identifying susceptibility variants for type 2 diabetes. *Methods Mol. Biol.* **376**: 235–250.
- Zhao, H., Jin, S., Fan, F., Fan, W., Tong, T., and Zhan, Q. 2000. Activation of the transcription factor Oct-1 in response to DNA damage. *Cancer Res.* **60**: 6276–6280.

Appendix B:

Activation of Inflammation/NF- κ b Signaling in Infants Born to Arsenic-exposed Mothers

Contributions: Statistical prediction and significance analysis

Activation of Inflammation/NF- κ B Signaling in Infants Born to Arsenic-Exposed Mothers

Rebecca C. Fry^{1,2}, Panida Navasumrit³, Chandni Valiathan^{1,2}, J. Peter Svensson^{1,2}, Bradley J. Hogan^{1,2}, Manlin Luo^{1,2}, Sanchita Bhattacharya^{1,2}, Kritinee Kandjanapa³, Sumitra Soontararuks³, Sumontha Nookabkaew³, Chulabhorn Mahidol³, Mathuros Ruchirawat³, Leona D. Samson^{1,2}

1 Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **3** Chulabhorn Research Institute, Bangkok, Thailand

The long-term health outcome of prenatal exposure to arsenic has been associated with increased mortality in human populations. In this study, the extent to which maternal arsenic exposure impacts gene expression in the newborn was addressed. We monitored gene expression profiles in a population of newborns whose mothers experienced varying levels of arsenic exposure during pregnancy. Through the application of machine learning-based two-class prediction algorithms, we identified expression signatures from babies born to arsenic-unexposed and -exposed mothers that were highly predictive of prenatal arsenic exposure in a subsequent test population. Furthermore, 11 transcripts were identified that captured the maximal predictive capacity to classify prenatal arsenic exposure. Network analysis of the arsenic-modulated transcripts identified the activation of extensive molecular networks that are indicative of stress, inflammation, metal exposure, and apoptosis in the newborn. Exposure to arsenic is an important health hazard both in the United States and around the world, and is associated with increased risk for several types of cancer and other chronic diseases. These studies clearly demonstrate the robust impact of a mother's arsenic consumption on fetal gene expression as evidenced by transcript levels in newborn cord blood.

Citation: Fry RC, Navasumrit P, Valiathan C, Svensson JP, Hogan BJ, et al. (2007) Activation of inflammation/NF- κ B signaling in infants born to arsenic-exposed mothers. *PLoS Genet* 3(11): e207. doi:10.1371/journal.pgen.0030207

Introduction

Arsenic is a ubiquitous environmental pollutant and a known human carcinogen [1]. Chronic arsenic exposure is an important public health hazard around the world, with millions of people exposed to drinking water with levels far exceeding the guideline of 10 μ g/l established by the WHO. Exposure to arsenic-contaminated drinking water is alarmingly high in many countries, most notably Bangladesh, where >25 million people are chronically exposed to extreme arsenic levels. Arsenic contamination is also a significant health concern in the United States, with numerous public water supplies measuring above the WHO limit [2].

Epidemiological studies indicate that chronic arsenic exposure in drinking water is associated with increased risk of skin, bladder, lung, liver, and kidney cancer [1]; in 1987, arsenic was classified as a Group 1 carcinogen by the International Agency for Research on Cancer. Although the mechanism of arsenic-induced carcinogenesis is not clearly established, it has been attributed to genotoxicity associated with reactive oxygen species [3]. Arsenic is also implicated in other human diseases such as vascular disorders, peripheral neuropathy, bronchiectasis, and diabetes [1].

The long-term health consequences of prenatal arsenic exposure in human populations are pronounced, with increased mortality rates caused by prenatal and early childhood exposures [4]. The detrimental health impact of prenatal arsenic exposure has also been shown in rodent models where in utero arsenic exposure resulted in a striking carcinogenic response (5-fold increase in hepatocellular carcinomas) among offspring; in utero arsenic exposure also

changed the expression of genes involved in cell proliferation, stress, and cell-cell communication that are evident even when the offspring reach adulthood. These results have profound implications suggesting that in utero arsenic exposure may result in epigenetic changes that persist through the life of the organism, ultimately impacting health status. A landmark study in mouse models shows that, indeed, in utero exposures via the maternal diet can cause permanent gene expression changes in the offspring that affect susceptibility to disease in the adult [7].

Given the implications of prenatal exposure on human health and the known public health hazard of chronic arsenic exposure, we set out to establish the extent to which maternal arsenic exposure in a human population impacts newborn gene expression. Additionally, these studies were aimed at understanding exactly how arsenic affects biological systems

Editor: Vivian G. Cheung, University of Pennsylvania, United States of America

Received: July 4, 2007; **Accepted:** October 4, 2007; **Published:** November 23, 2007

Copyright: © 2007 Fry et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: FDR, false discovery rate; GSEA, Gene Set Enrichment Analysis; WHO, World Health Organization

* To whom correspondence should be addressed. E-mail: lsamson@mit.edu; mathuros@cri.or.th

© These authors contributed equally to this work.

‡ Current address: Lawrence Berkeley National Laboratory, Berkeley, California, United States of America

Author Summary

Arsenic is an environmental pollutant and known human carcinogen. Chronic exposure to arsenic-contaminated water is an important public health hazard around the world, including the United States, with millions exposed to drinking water with levels that far exceed World Health Organization (WHO) guidelines. Given the implications of prenatal exposure on human health and the known public health hazard of chronic arsenic exposure, this study was aimed at establishing the extent to which maternal arsenic exposure in a human population affects newborn gene expression. The authors show that prenatal arsenic exposure in a human population results in alarming gene expression changes in newborn babies. The gene expression changes monitored in babies born to mothers exposed to arsenic during pregnancy are highly predictive of prenatal arsenic exposure in a subsequent test population. The study establishes a subset of just 11 transcripts that captured maximal predictive capability that could prove promising as genetic biomarkers of prenatal arsenic exposure. Pathway analysis of the genome-wide response in the babies exposed to arsenic in utero indicates robust activation of an integrated network of pathways involving NF- κ B, inflammation, cell proliferation, stress, and apoptosis. This study contributes to our understanding of biological responses to arsenic exposure.

and identifying genes that could be used as predictors, and therefore potential biomarkers, of prenatal arsenic exposure.

Results

Our study was based in the Ron Pibul and Bangkok districts of Thailand (Figure S1). The first case of arsenicosis (arsenic poisoning) in Thailand was reported in 1987 from the Ron Pibul district [8]. Rather than natural leaching of arsenic from geologic sources, Ron Pibul arsenic contamination is attributed to tin mining that took place from the 1960s to the 1980s. Arsenic concentrations in groundwater and shallow wells have been classified at a mean level of 503.5 μ g/l, about 50 times higher than WHO guidelines [9].

Using a population of arsenic-exposed and -unexposed mothers (as defined by WHO standards of chronic exposure to \sim 10 μ g/l arsenic), we set out to identify gene expression changes in the cord blood of newborns significantly associated with the extent of prenatal arsenic exposure. Cord blood is derived almost exclusively from the fetus; therefore, gene expression changes assessed in cord blood are representative of the newborn [10]. For this study, exposure classification was based on arsenic concentration in the mother's toenails, as this is representative of long-term arsenic accumulation [11,12]. Toenail samples were taken from a population of 32 volunteer subjects to quantify arsenic exposure in the mothers. A level of 0.5 μ g/g toenail arsenic corresponds to chronic consumption of water with \sim 10 μ g/l (see Materials and Methods), which is the official WHO maximum recommended concentration of arsenic in drinking water [11,12]. For the purposes of this study, women with toenail arsenic levels of $<$ 0.5 μ g/g were considered unexposed, and women with toenail levels of \geq 0.5 μ g/g were considered exposed. The levels of toenail arsenic across the 32 pregnant women ranged from 0.1 to 68.63 μ g/g (Figure 1A). Given the paucity of available unexposed newborn cord blood from Ron Pibul, the experimental design required additional utilization of unexposed newborn cord blood samples from Bangkok.

Gene Expression Signatures Highly Predictive of Arsenic Exposure

We set out to determine whether gene expression changes in a set of infants born to arsenic-exposed women versus unexposed women (as judged by WHO guidelines) could be used to predict arsenic exposure in a test population. For these analyses, two-class prediction was employed, where a training population was used to derive gene sets that were then tested as predictors of exposure in a separate population. The analyses were carried out in two phases: (i) where the training population was selected at random and the analyst "blinded" to arsenic exposure level in the test population and (ii) where all arsenic exposure levels of the population were revealed and used to define new training populations.

The first training population comprised 13 newborn subjects selected at random from the 32 newborns (Figure 1A). Specifically, RNA was extracted from cord blood of newborns 1–13, and hybridized to whole human genome arrays (Materials and Methods). To identify genes whose expression was associated with prenatal arsenic exposure, we used an approach that combined differential expression testing between the populations, plus a positive or inverse correlation of expression with increasing arsenic exposure (Materials and Methods). From the 13 newborn subjects, we identified the first expression signature (first gene set, Figure 1B) composed of 170 genes (Table S1) that differentiated the unexposed newborns (subjects 1–6) from the arsenic-exposed newborns (subjects 7–13). This prenatal arsenic exposure expression signature of 170 genes was then used to predict prenatal exposure in the remaining population of 19 newborns (subjects 14–32). The percent accuracy of class prediction was determined post-analysis by revealing the arsenic exposure of the test population to the analyst. Expression of these 170 genes accurately predicted prenatal arsenic exposure in 15 of 19 (79%) of the newborns (Figure 1B).

When the arsenic levels of the entire population were revealed, it became apparent that the first training population was composed of newborns with a wide range of exposure levels distributed over almost the entire range (Figure 1B). We hypothesized that a training population based on extreme exposures might yield higher predictive capacity. To assess this, arsenic-associated genes were identified using newborns at the extremes of arsenic exposure (i.e., the lowest versus the highest exposures) as the second training population (Figure 1A, second training population). Six newborns comprised the low-exposure population (subjects 1, 14, 15, 2, 16, and 3), and six newborns comprised the high-exposure population (subjects 29, 30, 12, 13, 31, and 32) (Figure 1A). As with the first gene set, differential expression testing and correlation analysis identified an expression signature, this time composed of 38 genes (Table S2) that differentiated infants born to mothers with very low and very high arsenic exposure levels (Figure 1A). These 38 genes were used to predict arsenic exposure in the remaining test population of 20 newborns. Even though the gene set was smaller (38 versus 170), prediction was just as high as that of the first gene set, with prenatal arsenic exposure accurately predicted in 16 of 20 (80%) of the newborns (Figure 1B, second test population).

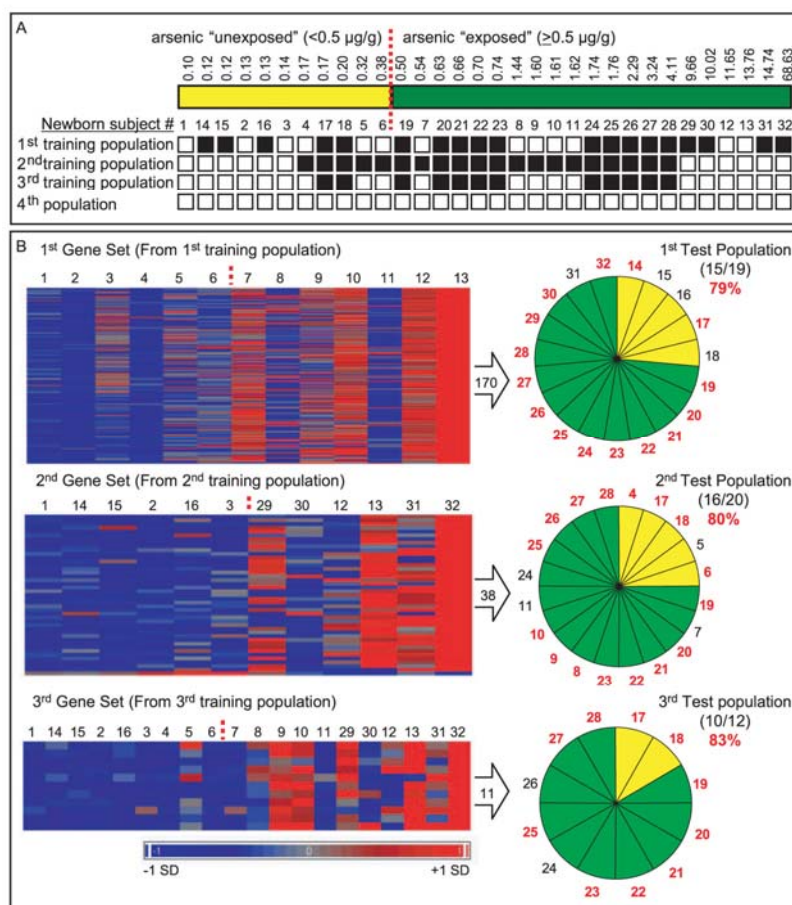


Figure 1. Gene Expression Signatures Predict Arsenic Exposure in Test Populations

(A) A population of newborns (subjects 1–32) born to mothers with varying levels of arsenic exposure was used to establish arsenic-associated gene expression signatures. Arsenic exposure levels were determined by maternal toenail arsenic concentration ($\mu\text{g/g}$). Babies born to unexposed (yellow) or arsenic-exposed mothers (green) were classified based on WHO guidelines with the cut point demarcated by the red dotted line. Subjects used in the populations to establish arsenic-associated gene sets are indicated with a white box. For two-class prediction, those subjects not included in the training population comprise the test population and are indicated with a black box.

(B) Three arsenic expression signatures (gene sets) were derived from populations spanning the range of arsenic exposure (first gene set), at the extremes of exposure (second gene set), or a combined population of the previous two (third gene set). To be included in the gene set, the transcript had to not only be differentially expressed (on average) between the exposed and unexposed groups, but also display a significant trend across increasing arsenic exposure levels. Expression values are mean centered with high relative expression indicated in red and low relative expression indicated in blue. The three derived gene sets (170 genes, 38 genes, or 11 genes) were used to predict prenatal arsenic exposure in test populations where correct classification is indicated by a red number. doi:10.1371/journal.pgen.0030207.g001

We next determined whether a training population derived from a combination of all of the training samples used to generate the first and second gene set would yield an expression signature with higher predictive capacity. This third training population was composed of nine unexposed newborns and 11 exposed newborns (Figure 1A). Differential expression testing and correlation analysis identified an expression signature of 11 genes (Figure 1B) that could predict prenatal arsenic exposure in 10 of 12 (83% accuracy) of the remaining newborn test population (Figure 1B). It is noteworthy that with only 11 genes, the power of prediction is as high as the first and second gene sets.

Many of the genes in the third gene set were represented in the gene sets derived from the first and second training

populations. Specifically, five of the 11 were identified in the first gene set and all 11 were present in the second gene set (Table 1). Given the high predictive capacity of these 11 genes, we hypothesize that these are key genes involved in the prenatal response of babies to arsenic and represent potential biomarkers of arsenic exposure. The potential arsenic biomarker set is composed of transcripts for the *CXLI*, *DUSP1*, *EGR-1*, *IER2*, *JUNB*, *MIRN21*, *OSM*, *PTGS2*, *RNF149*, *SFRS5*, and *SOC3* genes (Table 1). The dose response of expression level of each of the identified biomarkers is evident when plotted versus arsenic exposure across the population (Figure S2). Furthermore, to substantiate the association of the expression of the biomarkers with arsenic exposure, a multivariate model was employed (Materials and

Table 1. Potential Gene Biomarkers of Prenatal Arsenic Exposure—Third Gene Set

Gene	Description	Representative Public ID	Affymetrix Probeset ID	*Present in First Gene Set	*Present in Second Gene Set
<i>CXCL1</i>	chemokine (C-X-C motif) ligand 1	NM_001511	204470_at		*
<i>DUSP1</i>	dual specificity phosphatase 1	NM_004417	201041_s_at	*	*
<i>EGR1</i>	early growth response 1	NM_001964	201694_s_at	*	*
<i>IER2</i>	immediate early response 2	NM_004907	202081_at	*	*
<i>JUNB</i>	jun B proto-oncogene	NM_002229	201473_at		*
<i>MIRN21</i>	microRNA 21	BF674052	224917_at	*	*
<i>OSM</i>	oncostatin M	A1079327	230170_at	*	*
<i>PTGS2</i>	prostaglandin-endoperoxide synthase 2	NM_000963	204748_at		*
<i>RNF149</i>	ring finger protein 149	A1640483	235536_at		*
<i>SFRS5</i>	splicing factor	NM_006925	203380_x_at		*
<i>SOC3</i>	suppressor of cytokine signaling 3	BG035761	206359_at		*

Genes with expression significantly associated with prenatal arsenic exposure in a training population were identified and are highly predictive of arsenic exposure in a test population. doi:10.1371/journal.pgen.0030207.t001

Methods). The model was employed to determine significance of association of expression with two factors: (i) arsenic exposure and (ii) geographic source of samples (Materials and Methods). Geographic source was determined to be a nonsignificant factor for the expression level of the biomarkers ($p = 0.11$), whereas arsenic exposure was determined to be a highly significant factor ($p = 1.3 \times 10^{-9}$). Furthermore, for the set of biomarkers, the two factors of arsenic exposure and geographic source were not associated ($p = 0.77$).

Notably, associated molecular functions for the 11 gene products include stress response and cell cycle regulation. The zinc finger DNA binding transcription factor *EGR-1* (early growth response 1) is related to cell proliferation and is induced by mitogens such as EGF [13]. *EGR-1* regulates both proinflammatory cytokine activation and p53 transcription [14,15]. Not surprisingly, as *EGR-1* is known to activate cytokines, such signaling molecules are present in the arsenic biomarker gene set; namely, *OSM* (oncostatin M), a member of the interleukin-6 (IL-6) family of cytokines known to control cell cycle progression [16], *CXLI* (chemokine ligand 1), and *SOC* (suppressor of cytokine signaling 3). Additionally, *DUSP1* (dual specificity phosphatase 1) is involved in cell cycle regulation and is known to modulate cytokine expression [17,18]. An inflammation-activated acute phase response is indicated by the presence of the *JUNB* transcription factor, and *IER2* (immediate early response 2) transcripts in the biomarker set.

Genome-Wide Changes Associated with Prenatal Arsenic Exposure Are Robust

For a more global assessment of the impact of prenatal arsenic exposure on fetal gene expression, all biological pathways modulated in response to arsenic exposure were identified by studying the ontology of all the genes differentially expressed between the exposed and unexposed newborns across the entire population. For these analyses, the entire newborn population was used (the fourth population, Figure 1A) to define the fourth gene set that was differentially expressed between the two populations: the 21 newborns whose mothers were exposed to arsenic and the 11 newborns whose mothers were unexposed. It should be noted that for this analysis of global changes between the populations, the requirement for correlation with increasing

arsenic exposure was not imposed (Materials and Methods). This analysis identified 447 genes differentially expressed between the two populations of newborns, of which 404 (90%) were upregulated (Figure 2A; Table S3). Gene ontology enrichment analysis was performed to classify the genes modulated by prenatal arsenic exposure (Materials and Methods). This analysis identified ten gene ontology categories that were significantly enriched in the list of 447 genes (Table 2). Among the gene ontology categories that are significantly enriched are immune and inflammatory response ($p < 0.001$) (Table 2).

As an alternative approach to determine if groups of genes with common function are differentially expressed between the two newborn populations (arsenic exposed or unexposed), we have employed the knowledge-based Gene Set Enrichment Analysis (GSEA) (Materials and Methods). GSEA identified significant enrichment (false discovery rate [FDR] q -value < 0.01) of ten expression signatures with common biological function that are differentially expressed between the unexposed and exposed newborns. The groups of genes include three that represent stress-response signatures and three that represent tumor/cancer signatures (Table 3). The GSEA results also highlight that genes associated with estrogen receptor signaling are differentially expressed between the unexposed and exposed newborn populations (Table 3).

Arsenic-Modulated Networks Represent Numerous Biological Processes

We next determined whether known molecular interactions exist among the proteins encoded by the arsenic modulated transcripts. Of the 447 arsenic modulated transcripts, 285 gene products were identified in the Ingenuity knowledge base and overlaid with known human molecular interactions (Materials and Methods). Among these proteins, we identified the presence of a large arsenic-modulated interacting network of proteins (Figure 2B). Specifically, we identified a large interacting network comprised of 105 human proteins encoded by arsenic-modulated transcripts (indicated as red and green nodes) (Figure 2B; Table S4). The probability of finding 105 arsenic-modulated transcripts that encode for a protein network of this size by chance is $p <$

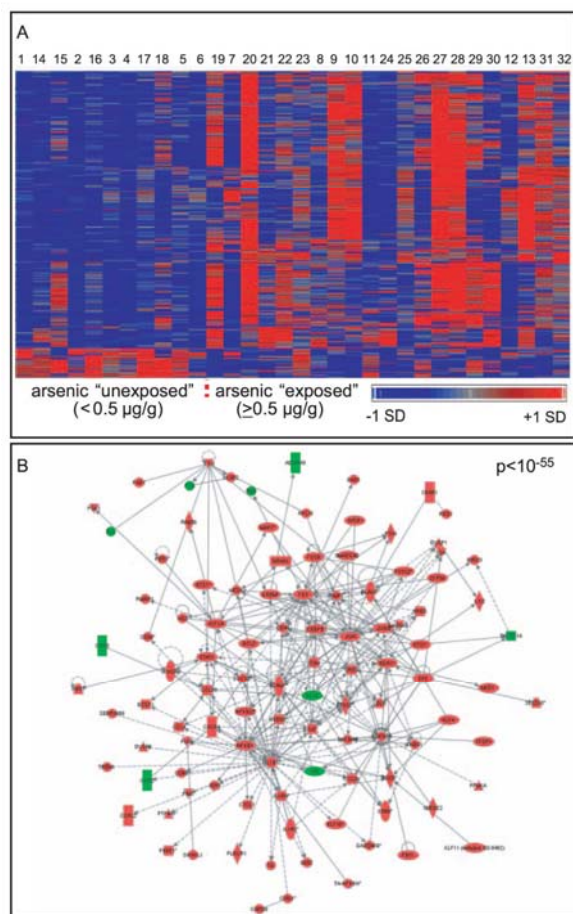


Figure 2. Prenatal Arsenic Exposure Results in Robust Genome-Wide Changes

(A) Heat map of the 447 differentially expressed genes identified between two newborn populations, those born to unexposed or arsenic exposed mothers. The cut point of exposure is indicated with a red dotted line. Unlike Figure 1, the differentially expressed transcripts did not have to display a significant trend with increasing arsenic exposure. Expression values are mean centered with high relative expression indicated in red and low relative expression indicated in blue.

(B) The 285 arsenic-modulated gene products existing in the Ingenuity database were analyzed for significant enrichment of molecular interactions. A significant ($p < 10^{-55}$) interactome containing 105 arsenic-modulated gene products was identified. Proteins in red represent arsenic-induced transcripts, proteins in green represent arsenic-repressed transcripts.

doi:10.1371/journal.pgen.0030207.g002

10^{-55} . Of the 105 proteins, 96 (91%) had transcripts that were upregulated in response to arsenic exposure.

Further analysis identified three highly significant ($p < 10^{-55}$) sub-networks embedded within the large interacting network (Figure 3A–3C). The first sub-network centers around the nuclear transcription factor NF-κB and the pro-inflammatory interleukin 1 family member IL1-β (Figure 3A). This network integrates two members of the potential biomarkers; namely, SOC3 and CXCL1 (Figure 3A). Note that transcripts for all proteins directly associated with NF-κB in this sub-network are upregulated in infants born to arsenic-exposed mothers (Figure 3A).

Table 2. Gene Ontology Enrichment

Gene Ontology Category	Gene Ontology Number	FDR q-Value
Immune response	GO:0006955	$p < 0.001$
Inflammatory response	GO:0006954	$p < 0.001$
Response to stress	GO:0006950	$p < 0.001$
Response to other organism	GO:0051707	$p < 0.001$
Response to pest, pathogen, or parasite	GO:0009613	$p < 0.001$
Response to wounding	GO:0009611	$p < 0.001$
Response to biotic stimulus	GO:0009607	$p < 0.001$
Response to external stimulus	GO:0009605	$p < 0.001$
Cytokine activity	GO:0005125	$p < 0.001$
Cell death	GO:0008219	0.0008

Genes differentially expressed between arsenic unexposed and exposed newborns were analyzed for significant enrichment of gene ontology categories. doi:10.1371/journal.pgen.0030207.t002

The second sub-network integrates biomarker member DUSP1 with two stress-activated transcription factors; namely, signal transducer and activator of transcription (STAT1) and hypoxia inducible factor-1 α (HIF-1α) (Figure 3B). Transcripts for both *STAT1* and *HIF-1α* were upregulated in infants with arsenic-exposed mothers (Figure 3B). *STAT1* is involved in cytokine signal transduction and is known to be activated by arsenic [19]. *HIF-1α* activation and resultant tumorigenesis has been linked to chronic arsenic exposure [20].

The third sub-network integrates four of the 11 potential arsenic biomarkers; namely, *EGR-1*, *OSM*, *PTGS2*, and *JUNB* (Figure 3C). These arsenic biomarker gene products are highly integrated with proteins known to be involved in cell cycle regulation, including *JUN* and *FOS*, as well as stress-response proteins such as interleukin-8 (*IL-8*) (Figure 3C). An overlay of molecular processes represented in this sub-network highlights the finding that prenatal arsenic exposure modulates numerous biological processes including stress response, signal transduction, cell adhesion, and transcription (Figure 3C).

Using network analyses, we also established that there are known molecular interactions among the 11 potential arsenic biomarker genes. Eight of the 11 biomarker gene products (exclusive of *SFRS5*, *MIRN21*, and *RNF149*) are highly integrated with tumor necrosis factor-α (*TNF-α*), another proinflammatory cytokine (Figure 3D). *TNF-α* is involved in the control of both cell proliferation and apoptosis [21]. Here, we identify *TNF-α* activation in newborn cord blood upon exposure to prenatal arsenic.

Evidence for Arsenic-Activated Transcriptional Control of Prenatal Responses

In an effort to uncover potential regulatory mechanisms underlying the transcription of the arsenic-modulated gene sets, we performed transcription factor binding site analysis within the promoters of the arsenic-modulated genes (Materials and Methods). Promoter region comparisons for the arsenic-modulated genes identified significant enrichment ($p < 0.05$) for two transcription factor binding sites across all four gene sets. Specifically, binding sites for NF-κB and serum response factor (SRF) are enriched in all four arsenic-modulated gene sets (Table 4). Moreover, metal

Table 3. GSEA

Gene Set	Gene Set Name (Molecular Signatures Database)	FDR q -value
Hypoxia/stress response signature	Hypoxia_Review	4.74E-03
Breast cancer/estrogen signature	Breast_Cancer_Estrogen_Signaling	3.00E-03
Transcription factor-induced signature	Gery_Cebp_Targets	2.33E-03
Hypoxia/stress response signature	Hypoxia_Reg_Up	3.71E-03
UV-induced signature	UVB_Nhek1_Up	2.97E-03
Trichostatin-A-induced signature	Tsa_Panc50_Up	2.67E-03
Tumor promoter signature	Tpa_Sens_Early_Up	2.48E-03
Tumor progression signature	Emt_Dn	2.47E-03
Cytokine-induced signature	Croonquist_IL6_Stroma_Up	5.10E-03
Butyrate-induced signature	Hdac1_Colon_But2hrs_Up	8.88E-03

GSEA was performed; gene sets from the Molecular Signatures Database were identified as significantly enriched.
doi:10.1371/journal.pgen.0030207.t003

response element binding sites (MREs) for the metal-responsive transcription factor-1 (MTF1) are enriched in three of the four gene sets (sets 1, 3, and 4) (Table 4). The MTF1 binding site enrichment was highest for the third gene set with five of the 11 genes containing the MRE element (Figure 3D). Notably, the enrichment for MTF1 in the second gene set only narrowly misses the enrichment $p < 0.05$ cutoff, at $p = 0.054$ (Table 4). MTF1 was shown to be activated upon arsenic exposure in animal models [23,24]. It is noteworthy that gene targets for a known arsenic-inducible transcription factor are found among the transcripts modulated in the cord blood of infants born to arsenic exposed mothers.

NF- κ B and Inflammation Signaling Identified in Arsenic-Exposed Newborns from Ron Pibul and Common Arsenic-Induced Stress Signaling across Species

As the unexposed samples utilized in this study were obtained from two different locations and could confound expression testing, we have used an alternative approach to substantiate the identified arsenic-induced pathways. Differential expression testing was performed between the cord blood of exposed and unexposed newborns from Ron Pibul (Materials and Methods). These analyses identified 321 genes that were differentially expressed between the arsenic-unexposed and -exposed newborns (Table S5). Notably, a direct comparison of gene expression changes identified considerable overlap between the transcripts differentially expressed between the newborns from Ron Pibul and transcripts differentially expressed across the whole population (fourth gene set) (Table S5).

To identify the biological pathways modulated by prenatal arsenic exposure, the proteins encoded by the 321 transcripts were analyzed for significant enrichment of molecular networks (Materials and Methods). Three highly significant protein sub-networks ($p < 10^{-30}$) were identified (Figure S3). As with the network findings from the entire population of newborns, the networks identified here integrate proteins known to be involved in cell cycle regulation including JUN, as well as stress-response proteins such as interleukin-8 (IL-8), the pro-inflammatory interleukin 1 family member IL-1 β , and hypoxia inducible factor-1 α (HIF-1 α) (Figure S3). Furthermore, the NF- κ B protein is integrated into the sub-networks and found to be activated in the cord blood of newborns

exposed to arsenic within the Ron Pibul population (Figure S3).

Finally, our analyses included comparisons of the gene expression changes identified in this study with arsenic-induced gene expression changes reported in the literature in mouse models as well as a separate arsenic-exposed human population. Our results were compared with (i) expression changes in livers of mice treated with arsenic [24], (ii) expression changes identified in arsenic-induced tumors resulting from in utero exposures to arsenic in mice [6], and (iii) expression changes in blood from a human population from Taiwan exposed to arsenic [25]. These comparisons identify overlap of similarly modulated transcripts in response to arsenic exposure that include: *BCL6* (B-cell CLL/lymphoma 6), *CD14* (CD14 antigen), *CXCL1* (chemokine ligand 1), *EGR1* (early growth response 1), *FOS* (v-fos FBJ murine osteosarcoma), *FOSB* (FBJ murine osteosarcoma viral oncogene homolog B), *GADD45B* (growth arrest and DNA damage inducible beta), *IFNGR1* (interferon gamma receptor 1), *IL1B* (interleukin 1 beta), *IL1R1* (interleukin 1 receptor 1), *JUN* (v-jun sarcoma virus oncogene), *MAPK6* (mitogen-activated protein kinase 6), *MTIX* (metallothionein IX), *RAD23B* (RAD23 homolog B), and *TOPI* (topoisomerase DNA 1) (Tables S3 and S5). These results highlight the modulation of stress related transcripts in both mice (acute and in utero exposures) and a separate adult human population in response to arsenic exposure.

Discussion

Globally, millions of people are at risk for the detrimental effects of chronic arsenic exposure with drinking water levels far exceeding the WHO guideline [1]. Prenatal arsenic exposure in human populations has been associated with pronounced long-term health consequences [4]. Here, we address the impact of maternal arsenic exposure on fetal gene expression in a human population. Our goals were 2-fold: first, to establish the extent to which chronic arsenic exposure in mothers impacts newborn gene expression, and second, to identify genes that could be used as potential biomarkers of prenatal arsenic exposure and targets for remedial therapy.

Differential expression testing of training populations of

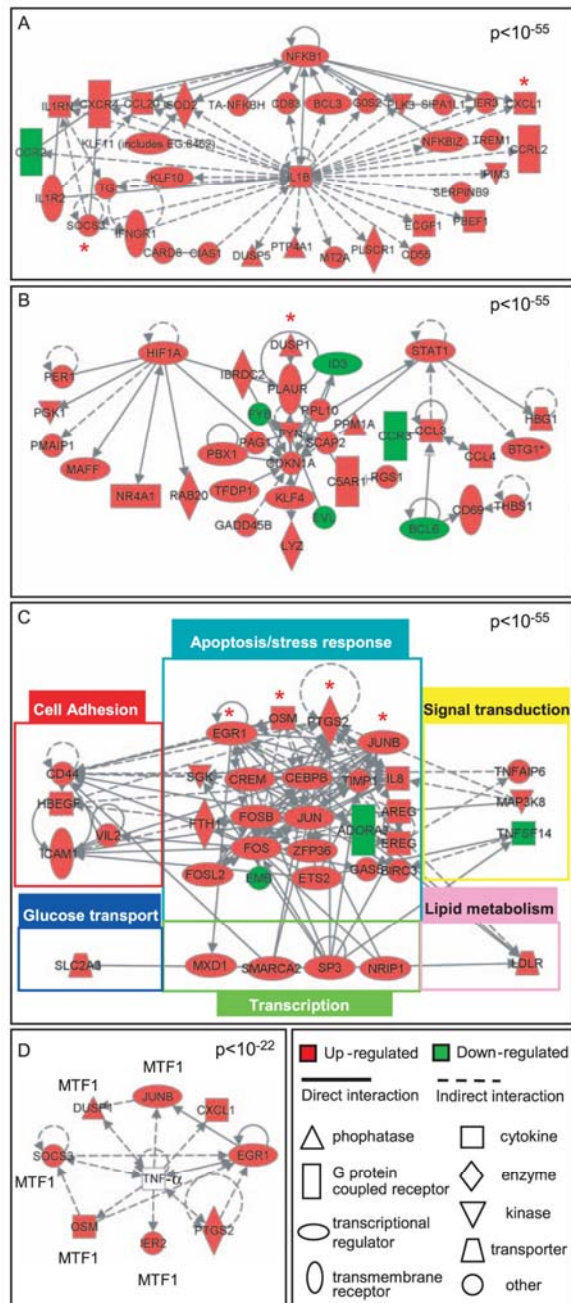


Figure 3. Sub-networks of Prenatal Arsenic-Modulated Gene Products (A) A sub-network that integrates NF- κ B and IL1- β with SOC3 and CXCL1 was identified. Note that SOC3 and CXCL1 are among the 11 potential gene biomarkers for arsenic exposure shown in Table 1. (B) A sub-network that integrates STAT1 and HIF1- α with DUSP1. (C) An EGR-1, OSM, JUNB focused sub-network highlights numerous biological processes modulated in response to arsenic. Proteins encoded by 11 potential gene biomarkers for arsenic exposure are indicated with a red asterisk. (D) TNF- α -associated network composed of eight core members of the potential gene biomarkers for arsenic exposure. Biomarker genes with binding sites for MTF transcription factor are indicated. Proteins in red

represent arsenic-induced transcripts, proteins in green represent arsenic-repressed transcripts. doi:10.1371/journal.pgen.0030207.g003

newborns whose mothers had varied exposures to arsenic identified three arsenic-associated gene expression signatures comprised of 170, 38, and 11 genes. Analysis of the predictive capacity of each of these gene sets using the Support Vector Machine two-class prediction algorithm showed that each of these gene sets is highly predictive of arsenic exposure in a test population. Notably, even the smallest gene set comprised of 11 genes was powerful, with 83% accuracy in predicting prenatal arsenic exposure in the test population. The 11 potential biomarkers of prenatal arsenic exposure include *CXCL1*, *DUSP1*, *EGR-1*, *IER2*, *JUNB*, *MIRN21*, *OSM*, *PTGS2*, *RNF149*, *SFRS5*, and *SOC3*. The set of 11 genes show a striking dose response to prenatal arsenic exposure. Stress response and cell cycle regulation are associated molecular functions of the potential biomarker set. Arsenic exposure is known to activate stress-related transcripts in yeast, animal models and human subjects [24–26]. Here, we find that stress-response genes are differentially expressed among a population of newborns whose mothers were exposed to varying levels of arsenic.

To assess the genome-wide impact of prenatal arsenic exposure on newborn gene expression, we identified all transcripts that showed differential expression between two populations; the 21 newborns whose mothers had been exposed to arsenic versus the 11 newborns whose mothers were unexposed. These analyses identified ~450 genes differentially expressed between the two populations, of which 90% had expression levels that were increased (rather than decreased) by arsenic exposure. Clearly, there is a robust genome-wide response to prenatal arsenic exposure with ~3% of the expressed genes significantly altered in the newborn. Gene ontology and GSEA highlight the activation of stress-related transcripts in the cord blood of infants exposed prenatally to arsenic.

Furthermore, integration of the gene products of the ~450 transcripts with known molecular interactions identified the existence of a large arsenic-modulated interacting network of 105 proteins. Embedded within this large interacting network are three sub-networks that highlight that prenatal arsenic exposure activates inflammation-related molecules. Specifically, the first of the sub-networks centers around NF- κ B and IL1- β . NF- κ B regulates a large number of genes critical for apoptosis, as well as inflammation-related molecules such as cytokines (interleukins). IL1- β belongs to the class of acute phase proteins known to be increased in response to inflammation. Links between prenatal arsenic exposure and the activation of a stress response are also evident in the second and third sub-networks. Prenatal arsenic exposure resulted in the induction of the stress-related transcription factors STAT1 and HIF-1 α , both of which are known to be activated by arsenic in model systems [19]. Here, we identify STAT1 and HIF-1 α activation in newborn cord blood upon prenatal arsenic exposure. The activation of stress-response proteins such as interleukin-8 (IL-8) in response to prenatal arsenic exposure is also evident in sub-network three. The gene expression signatures identified here as modulated by prenatal arsenic exposure were compared to arsenic-induced gene expression changes in the mouse model and also with a

Table 4. Transcription Factor Binding Site Enrichment

Transcription Factor	Description	Gene Set <i>p</i> -Values			
		First (170 Genes)	Second (38 Genes)	Third (11 Genes)	Fourth (447 Genes)
SRF	Serum response factor	8.46×10^{-4}	0.011	0.008	1.7×10^{-5}
NF- κ B	Nuclear factor- κ B	1.39×10^{-4}	0.011	0.011	8.52×10^{-6}
MTF-1	Metal transcription factor	0.0173	0.054	0.007	0.021

Four arsenic-associated gene sets were analyzed for significant enrichment of transcription factor binding sites. Binding sites for three transcription factors were identified as enriched (*p*-values for significance are shown).
doi:10.1371/journal.pgen.0030207.t004

separate human population. These comparisons highlight the common pattern of activation of stress-related transcripts in response to arsenic exposure.

Additionally, eight of the 11 biomarker gene products were found to have significant interactions with the proinflammatory cytokine TNF- α . Several studies in animal models have shown that arsenic exposure results in TNF- α stimulation [27–29]. In this study, TNF- α activation is identified in newborn cord blood upon prenatal arsenic exposure. Taken together, the network findings underscore that a mother's arsenic exposure results in a robust response in the fetus, indicative of a systemic inflammatory response along with the modulation of numerous other biological processes including apoptosis, signal transduction, cell adhesion, and transcription.

We further show that the extensive genome-wide newborn response to prenatal arsenic exposure may be regulated by at least three transcription factors. Analysis of the promoter regions of the arsenic-modulated genes showed enrichment for NF- κ B and SRF in all four arsenic-modulated gene sets. SRF transcriptionally activates the expression of immediate early response genes, including *C-FOS* and *EGR-1* [30], two members of the potential arsenic biomarker set. Moreover, binding sites for the metal-responsive transcription factor-1 (MTF1) are enriched in three of the four gene sets (sets 1, 3, and 4). MTF1 was shown to be activated upon arsenic exposure in animal models [23,24]. That gene targets for a known arsenic-inducible transcription factor are found among the transcripts modulated in the cord blood of infants born to arsenic exposed mothers supports our conclusions that the transcriptional changes reported here are likely due to prenatal arsenic exposure.

Our findings clearly demonstrate the robust impact of a mother's arsenic consumption on gene expression in utero as evidenced by transcript levels in the newborn's cord blood. More specifically, our data suggest that prenatal arsenic exposure acts as an inflammatory stimulus that activates the NF- κ B signaling cascade. NF- κ B activation plays a critical role in inflammation-driven tumor progression [31], and thus key players in tumor progression are modulated in the blood of newborns exposed to arsenic. To determine the extent to which these exposures and the resultant expression changes are associated with susceptibility to disease in later life, the health status of these children is currently being followed.

Conclusions

In summary, class prediction algorithms identified gene expression signatures that predict arsenic exposure in a test population with about 80% accuracy. Notably, by integrating

training populations with varied exposures, a highly predictive potential biomarker gene set composed of just 11 genes was identified. These genes are promising as genetic biomarkers for prenatal arsenic exposure. Currently, we cannot eliminate the possibility that the gene expression signatures identified here are not absolutely specific for arsenic; they may also be predictive of other environmental exposures, e.g., exposure to other heavy metals. Nevertheless, this study underscores that there is a robust prenatal response that correlates with arsenic-exposure levels that could modulate numerous biological pathways including apoptosis, cell signaling, the inflammatory response, and other stress responses, and ultimately affect health status. Arsenic contamination of the drinking water in the Ron Pibul area of Thailand is representative of that seen in many other areas of South East Asia, most notably Bangladesh [9], suggesting that prenatal exposures are likely to be endemic in these areas. Moreover, arsenic contamination of the Ron Pibul drinking water is roughly the same as that known to be present in many of the western United States [2,9], suggesting that prenatal arsenic exposure may also be a problem in the United States. These data contribute to our understanding of biological responses upon arsenic exposure, and show that prenatal exposure in humans results in measurable phenotypic responses in the newborn.

Materials and Methods

Study locations and subjects. The study was conducted in Bangkok and the Ron Pibul District of the Nakhon Sri Thammarat Province located in the southern peninsula of Thailand (Figure S1). Five villages in the Ron Pibul district were selected for the study location as they had been classified as high level arsenic contaminated areas, and arsenicosis had been reported there [8]. Arsenicosis has not been reported in Central Thailand, specifically Bangkok, where arsenic concentrations in water and soil have been determined to be very low [8]. The study subjects consisted of 32 pregnant women (20–40 y old). All subjects were healthy, pregnant volunteers undergoing vaginal childbirth without birth stimulation or anesthesia. Twenty-three pregnant women living in the Ron Pibul District and nine women living in Bangkok for at least 1 y were recruited for the study. Women from both sites were age, educational level, and socioeconomically matched. Questionnaires were administered to all participants to obtain personal information regarding residential history, health history and potential confounding factors, birth and pregnancy information (number of births, abortions or complications), use of community drinking water and well water, plus water and food consumption habits. Cord blood samples were collected from January 2004 to December 2005 in the Ron Pibul Hospital (Ron Pibul District) and the Rajvithi Hospital (Bangkok). This study was conducted according to the recommendations of the Declaration of Helsinki (World Medical Association 1989) for international health research. All subjects gave written informed consent to participate in this study.

Sample collection and arsenic measurement. Pregnant participants were asked to provide toenail samples during pregnancy for analysis of total arsenic concentration, which was determined by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) (Agilent 7500c). After delivery, 2.5 ml of newborn cord blood was collected into a PAXgene Blood RNA (Qiagen) tube for study of gene expression. All cord blood samples were kept at -70°C until analysis.

Microarray analysis. Total RNA was isolated from 32 cord blood samples according to the PAX gene protocol and Qiagen RNA extraction kit. RNA was labeled using a globin reduction protocol (Affymetrix) and hybridized to HGU133 Plus 2.0 full genome human arrays in technical duplicate for a total of 64 arrays. Data were first normalized using Robust Multi-Chip Average (RMA) [32] and filtered for expressed transcripts across all arrays ($+2$ standard deviations above mean background) resulting in reduction of the probesets from the original 54,675 to 15,265. A mean absolute expression value was calculated from technical duplicates of the arrays for all expressed transcripts. Differential gene expression and association with increasing arsenic concentration was calculated as follows. The samples comprising the training sets were separated into two groups based on arsenic exposure level. The two groups were unexposed (maternal toenail <0.5 $\mu\text{g/g}$) or exposed (maternal toenail ≥ 0.5 $\mu\text{g/g}$). The two-class exposure designation is based on the WHO standards for exposure to arsenic of 10 $\mu\text{g/l}$ arsenic. A mean toenail arsenic concentration of 0.5 $\mu\text{g/g}$ corresponding to chronic consumption of drinking water at 10 $\mu\text{g/l}$ arsenic was derived from two studies associating arsenic toenail concentration and drinking water in a population from Bangladesh [12] and the United States [11]. Differential expression was determined as a significant difference in the expression of a gene (exposed versus unexposed) where the average fold change was greater than ± 1.5 and $p < 0.05$ (t -test). Additionally, significant association of gene expression and increasing arsenic level was determined by correlation measurements ($r^2 \geq +0.6$, $r^2 \leq -0.6$; $p < 0.01$) calculated using the linear regression model in S-PLUS 7.0 (<http://www.insightful.com>). The two-class prediction model used for assessing arsenic exposure in test populations was Support Vector Machine, carried out in Gene Pattern Software (version 2.0.1) (<http://www.broad.mit.edu>). Multivariate analysis was performed as follows: the expression values (Y) for each gene were modeled using $Y = \beta_1 + \beta_2 \text{ars} + \beta_3 \text{loc}$ (arsenic) + $\beta_3 \text{loc}$ (geographic location), where toenail arsenic concentration is a continuous variable and location is binary. Statistical significance was determined by subjecting β_2 and β_3 to t -statistics. A χ^2 test for dependence (association) of the two factors (e.g., arsenic and geographic location) was performed for the set of arsenic biomarkers. A Fisher's exact test was employed to determine overrepresentation of the biomarkers within the genes significantly associated with either geographic source or arsenic exposure ($p < 0.01$). Network analyses were performed using the Ingenuity software (<http://www.ingenuity.com>). Gene ontology enrichment analysis was performed using GO Miner [33]. GSEA [34] was performed using the GSEA desktop software [35], with a false discovery rate correction (Benjamini-Hochberg) employed. Microarray data have been deposited to the Gene Expression Omnibus repository.

Transcription factor binding site analysis. Transcription factor binding site analysis was performed using Expander software [36] and Genomatix software (<http://www.genomatix.de>). For both analyses, Affymetrix probesets were linked to sequence data for regions 1,000 base pairs upstream and 200 base pairs downstream of the transcription start sites, and these were analyzed for significant enrichment of transcription factor binding sites. Significance ($p \leq 0.05$) was calculated where significance is the probability of obtaining an equal or greater number of sequences with a model match in a randomly drawn sample of the same size as the input sequence set.

Supporting Information

Figure S1. Map of Study Location

The study was conducted in Bangkok and the Ron Pibul District of the Nakhon Sri Thammarat Province located in the southern peninsula of Thailand. Study locations are indicated with red circles. Found at doi:10.1371/journal.pgen.0030207.sg001 (6.3 MB AI).

References

1. NRC (1999) National Research Council Report: Arsenic in the drinking water. Washington, DC: National Academy Press.

Figure S2. Expression Patterns of 11 Biomarkers of Prenatal Arsenic Exposure

Transcripts that are predictive of prenatal arsenic exposure are plotted with expression intensity versus arsenic exposure. Found at doi:10.1371/journal.pgen.0030207.sg002 (615 KB AI).

Figure S3. Arsenic-Modulated Sub-networks Identified from Differentially Expressed Genes from the Ron Pibul Population

Significant sub-networks of arsenic-modulated gene products were identified between the unexposed and exposed newborn populations of Ron Pibul. Found at doi:10.1371/journal.pgen.0030207.sg003 (10.9 MB AI).

Table S1. Genes Differentially Expressed between Newborns That Composed the First Training Population

A total of 170 genes were differentially expressed between newborns born to mothers unexposed to arsenic and newborns born to mothers exposed to arsenic that composed the first training population. Found at doi:10.1371/journal.pgen.0030207.st001 (54 KB DOC).

Table S2. Genes Differentially Expressed between Newborns That Composed the Second Training Population

A total of 38 genes were identified as differentially expressed between newborns born to mothers unexposed to arsenic and newborns born to mothers exposed to arsenic that composed the second training population. Found at doi:10.1371/journal.pgen.0030207.st002 (29 KB DOC).

Table S3. Genes Differentially Expressed between Two Newborn Populations; Those Born to Mothers Unexposed to Arsenic and Those Born to Mothers Exposed to Arsenic

A total of 447 genes were identified as differentially expressed between the two newborn populations; those newborns born to mothers unexposed to arsenic or to mothers exposed to arsenic. Found at doi:10.1371/journal.pgen.0030207.st003 (122 KB DOC).

Table S4. Arsenic-Modulated Gene Products

A total of 105 arsenic modulated gene products are contained in a significant large interactome ($p < 10^{-55}$). Found at doi:10.1371/journal.pgen.0030207.st004 (38 KB DOC).

Table S5. Genes Differentially Expressed between Arsenic-Unexposed and -Exposed Newborns in Ron Pibul

A list of the 321 genes differentially expressed between the two newborn populations from the Ron Pibul Province; those born to mothers unexposed to arsenic, and those born to mothers exposed to arsenic. Found at doi:10.1371/journal.pgen.0030207.st005 (85 KB DOC).

Accession Numbers

Microarray data have been deposited to the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus repository under Series Record GSE7967 (www.ncbi.nlm.nih.gov/geo/).

Acknowledgments

We would like to thank Jill Mesirov, Gabriela Alexe, Bevin Engelward, Gerald Wogan, John Essigmann, and David Hunter for their comments on the manuscript.

Author contributions. RCF, CM, MR, and LDS conceived and designed experiments. SS isolated cord blood samples. KK and SN analyzed arsenic concentration in biological samples. RCF, BJH, and ML extracted RNA and performed microarray hybridization. RCF, PN, CV, JPS, SB, and LDS analyzed the data. RCF, PN, and LDS wrote the manuscript.

Funding. This work was supported by grants ES11399 and ES002109.

Competing interests. The authors have declared that no competing interests exist.

2. Welch AH, Helsel DR, Focazio MJ, Watkins SA (1999) Arsenic in ground water supplies of the United States. In: Chappell WR, Abernathy CO, Calderon RL, editors. Arsenic exposure and health effects. New York: Elsevier Science, pp. 9–17.

3. Hei TK, Liu SX, Waldren C (1998) Mutagenicity of arsenic in mammalian cells: role of reactive oxygen species. *Proc Natl Acad Sci U S A* 95: 8103–8107.
4. Smith AH, Marshall G, Yuan Y, Ferreccio C, Liaw J, et al. (2006) Increased mortality from lung cancer and bronchiectasis in young adults after exposure to arsenic in utero and in early childhood. *Environ Health Perspect* 114: 1293–1296.
5. Waalkes MP, Liu J, Ward JM, Diwan BA (2004) Animal models for arsenic carcinogenesis: inorganic arsenic is a transplacental carcinogen in mice. *Toxicol Appl Pharmacol* 198: 377–384.
6. Liu J, Xie Y, Ward JM, Diwan BA, Waalkes MP (2004) Toxicogenomic analysis of aberrant gene expression in liver tumors and nontumorous livers of adult mice exposed in utero to inorganic arsenic. *Toxicol Sci* 77: 249–257.
7. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL (2006) Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* 114: 567–572.
8. Williams M (1996) Arsenic contamination in surface drainage and groundwater in part of the southeast Asian Tin Belt, Nakhon Sri Thammarat Province, southern Thailand. *Environ Geol* 27: 16–33.
9. Mandal BK, Suzuki KT. (2002) Arsenic round the world: a review. *Talanta* 58: 201–235.
10. Lo YM, Lo ES, Watson N, Noakes L, Sargent IL, et al. (1996) Two-way cell traffic between mother and fetus: biologic and clinical implications. *Blood* 88: 4390–4395.
11. Karagas MR, Tosteson TD, Blum J, Klaue B, Weiss JE, et al. (2000) Measurement of low levels of arsenic exposure: a comparison of water and toenail concentrations. *Am J Epidemiol* 152: 84–90.
12. Kile ML, Houseman EA, Rodrigues E, Smith TJ, Quamruzzaman Q, et al. (2005) Toenail arsenic concentrations, GSTT1 gene polymorphisms, and arsenic exposure from drinking water. *Cancer Epidemiol Biomarkers Prev* 14: 2419–2426.
13. Hofer G, Grimmer C, Sukhatme VP, Sterzel RB, Rupprecht HD (1996) Transcription factor Egr-1 regulates glomerular mesangial cell proliferation. *J Biol Chem* 271: 28306–28310.
14. Adamson ED, Mercola D (2002) Egr1 transcription factor: multiple roles in prostate tumor cell growth and survival. *Tumour Biol* 23: 93–102.
15. Krones-Herzig A, Mittal S, Yule K, Liang H, English C, et al. (2005) Early growth response 1 acts as a tumor suppressor in vivo and in vitro via regulation of p53. *Cancer Res* 65: 5133–5143.
16. Klausen P, Pedersen L, Jurlander J, Baumann H (2000) Oncostatin M and interleukin 6 inhibit cell cycle progression by prevention of p27kip1 degradation in HepG2 cells. *Oncogene* 19: 3675–3683.
17. Li M, Zhou JY, Ge Y, Matherly LH, Wu GS (2003) The phosphatase MKP1 is a transcriptional target of p53 involved in cell cycle regulation. *J Biol Chem* 278: 41059–41068.
18. Hammer M, Mages J, Dietrich H, Servatius A, Howells N, et al. (2006) Dual specificity phosphatase 1 (DUSP1) regulates a subset of LPS-induced genes and protects mice from lethal endotoxin shock. *J Exp Med* 203: 15–20.
19. Chelbi-alix MK, Bobe P, Benoit G, Canova A, Pine R (2003) Arsenic enhances the activation of Stat1 by interferon gamma leading to synergistic expression of IRF-1. *Oncogene* 22: 9121–9130.
20. Kamat CD, Green DE, Curilla S, Warnke L, Hamilton JW, et al. (2005) Role of HIF signaling on tumorigenesis in response to chronic low-dose arsenic administration. *Toxicol Sci* 86: 248–257.
21. Tracey KJ, Cerami A (1993) Tumor necrosis factor: an updated review of its biology. *Crit Care Med* 21: S415–422.
22. Rahman M, Vahter M, Wahed MA, Sohel N, Yunus M, et al. (2006) Prevalence of arsenic exposure and skin lesions. A population based survey in Matlab, Bangladesh. *J Epidemiol Community Health* 60: 242–248.
23. Kumagai Y, Sumi D (2006) Arsenic: signal transduction, transcription factor, and biotransformation involved in cellular response and toxicity. *ev Pharmacol Toxicol*.
24. Liu J, Kadiiska MB, Liu Y, Lu T, Qu W, et al. (2001) Stress-related gene expression in mice treated with inorganic arsenicals. *Toxicol Sci* 61: 314–320.
25. Wu MM, Chiou HY, Ho IC, Chen CJ, Lee TC (2003) Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. *Environ Health Perspect* 111: 1429–1438.
26. Haugen AC, Kelley R, Collins JB, Tucker CJ, Deng C, et al. (2004) Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol* 5: R95.
27. Germolec DR, Spalding J, Yu HS, Chen GS, Simeonova PP, et al. (1998) Arsenic enhancement of skin neoplasia by chronic stimulation of growth factors. *Am J Pathol* 153: 1775–1785.
28. Germolec DR, Spalding J, Boorman GA, Wilmer JL, Yoshida T, et al. (1997) Arsenic can mediate skin neoplasia by chronic stimulation of keratinocyte-derived growth factors. *Mutat Res* 386: 209–218.
29. Germolec DR, Yoshida T, Gaido K, Wilmer JL, Simeonova PP, et al. (1996) Arsenic induces overexpression of growth factors in human keratinocytes. *Toxicol Appl Pharmacol* 141: 308–318.
30. Norman C, Runswick M, Pollock R, Treisman R (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55: 989–1003.
31. Luo JL, Maeda S, Hsu LC, Yagita H, Karin M (2004) Inhibition of NF- κ B in cancer cells converts inflammation-induced tumor growth mediated by TNF α to TRAIL-mediated tumor regression. *Cancer Cell* 6: 297–305.
32. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31: e15.
33. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, et al. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 4: R28.
34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545–15550.
35. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP (2007) GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. Epub: 20 July 2007.
36. Shamir R, Maron-Katz A, Tanay A, Linhart C, Steinfeld I, et al. (2005) EXPANDER—an integrative program suite for microarray data analysis. *BMC Bioinformatics* 6: 232.