

**TEL AVIV UNIVERSITY
SACKLER SCHOOL OF MEDICINE
DEPARTMENT OF HUMAN MOLECULAR GENETICS AND
BIOCHEMISTRY**

**Genome-wide dissection of cellular responses to DNA
double-strand breaks using integrated analysis of
expression microarrays and bioinformatics**

Submitted by

Seagull Shavit ID: 034233205

This work was performed in partial fulfillment of the requirements for a degree of
"Master in Sciences" in the Sackler School of Medicine, Tel-Aviv University

This research work was carried out under the supervision of

Professor Yosef Shiloh

Professor Ron Shamir

Department of Human Molecular

School of Computer Science

Genetics and Biochemistry

Sackler Faculty of Exact Sciences

Sackler School of Medicine

September 2007

Acknowledgements

I would like to sincerely thank my supervisors, Prof. Yosef Shiloh and Prof. Ron Shamir, for their devoted guidance throughout this work.

I wish to express my profound gratitude to Dr. Sharon Rashi-Elkeles and Dr. Ran Elkon. I greatly appreciate the direct guidance, the devoted counseling and the knowledge they provided me with.

I would like to thank Yaniv Lerenthal, who gave hand whenever asked, Dr. Yael Ziv, who was always willing to provide consultation and Adi Maron-Katz, for her professional and friendly support in the use of Expander.

Last but not least, I thank all my colleagues from the lab that were always generous in sharing their time and knowledge.

Contents

Abbreviations	3
Abstract	4
1 Introduction	7
1.1 DNA damage response networks	7
1.2 The DSB response concerted by ATM	9
1.3 Systems biology and functional genomics	12
1.4 Functional genomics technologies	14
1.5 Gene expression microarrays: platforms	16
2 Research Goals and Specific Aims	19
3 Materials and Methods	20
3.1 Materials	20
3.1.1 Solvents, powders and general reagents	20
3.1.2 Cell culture media	20
3.1.3 Cell culture growth materials	21
3.1.4 Cell lines	21
3.1.5 Materials for analysis of RNA	22
3.1.6 Miscellaneous Supplies	23
3.1.7 Instruments	23
3.2 Methods	24
3.2.1 Cell culture	24
3.2.2 Induction of DNA double strand breaks (DSBs)	25
3.2.3 Total RNA isolation	25
3.2.4 Quantification and quality assessment of RNA	27
3.2.5 Preparation of the RNA samples for hybridization	28
3.2.6 Target Hybridization	28
3.2.7 Analysis of gene expression data	30
4 Results	39
4.1 Calibration of the experimental system, using human fibroblasts exposed to NCS	39
4.1.1 Intra-laboratory comparison between biological replicates	40
4.1.2 Testing the side-effects of experimental procedure on gene expression	45
4.1.3 Selecting informative time points for the following experiment	49
4.2 Delineation of transcriptional responses induced by IR in five human cell lines	53
4.2.1 Hierarchical clustering of the cell lines	54
4.2.2 Analysis of differentially expressed genes in specific cell lines	57
4.2.3 Integrative analysis of the five cell lines	74
5 Discussion	90
5.1 Calibration of the experimental system using human fibroblasts exposed to NCS	90
5.2 Delineation of transcriptional responses induced by IR in five human cell lines	91
5.2.1 Hierarchical clustering of the cell lines	91
5.2.2 Analysis of differentially expressed genes in specific cell lines	92
5.2.3 Integrative analysis of the five cell lines	94
5.3 Future prospects	99
References	100

Abbreviations

A-T	ataxia-telangiectasia
BS	binding site
CGH	comparative genomic hybridization
DB	database
DDR	DNA damage response
DSB	double strand break
FC	fold change
IR	ionizing radiation
MM	mismatch
NCS	neocarzinostatin
NF- κ B	nuclear factor kappa B
PCR	polymerase chain reaction
PM	perfect match
PWM	position weight matrix
RNAi	RNA interference
SAGE	serial analysis of gene expression
SD	standard deviation
SNP	single nucleotide polymorphism
TF	transcription factor
TSS	transcription start site

Abstract

Background: DNA damage poses one of the greatest threats to the function and life of the cell and the organism and therefore cells acquired intricate mechanisms to sense and handle such challenges. The efficiency and quality of cellular responses to DNA damage determine whether this insult will be repaired with no lasting effect on cellular life, divert the cell from normal growth to programmed cell death (apoptosis), or end up in neoplastic transformation. Understanding of DNA damage responses has broad implications for basic life processes such as cell cycle control, aging, tissue development and degeneration. It is highly relevant for human health, primarily to coping with environmental hazards, cancer formation, and many neurodegenerative disorders.

Goals: The major goal of our research was to examine the difference between the transcriptional responses to DNA double strand breaks (DSBs) in several human cell lines using experimental and computational analysis. Specific goals were to identify by computational means transcription factors that control the transcriptional response induced by DNA damage and dissect the damage response network into arms mediated by these regulators. Another goal was to find the common denominator of DNA damage responses in five different human cell lines in addition to the differences in damage response among them.

Methods: We used a high-throughput molecular technology - DNA microarrays - to build gene expression profiles. This technology enables detecting genes that are expressed differentially in different cell populations. The experiments were conducted on five different human cell lines, derived from both normal and cancerous tissues. In each experiment, we constructed gene expression profiles in different biological conditions including treatment with DNA DSBs agents, for various types of cell lines. DNA DSBs were induced by treating the cells with the radiomimetic drug neocarzinostatin (NCS), or by exposing them to ionizing radiation (IR).

Mining meaningful biological insights from the enormous amounts of raw data obtained by microarray experiments poses a major bioinformatic challenge. To meet this challenge we adopted an integrative data analysis approach that starts with the initial preprocessing steps of signal extraction, normalization and filtering, and continues through partition analysis (clustering or biclustering) to high-level statistical analyses that seek enriched functional categories and cis-regulatory promoter elements in the clusters/biclusters.

Results: The experimental part of this thesis included two microarray experiments. In the first experiment, which served for calibration, we exposed human BJ-hTert cells to NCS and recorded global transcriptional responses 2, 4, 8 and 12 hr later, with untreated controls at 0 and 8 hr. In this experiment we compared biological replicates prepared by two researchers, examined the effect of the experimental manipulation on gene expression and identified informative time points for the subsequent, more comprehensive, experiment. We concluded that there was very good correlation between the intra-laboratory replicates, yielding reliable and significant results. We noted that the manipulation of culture dishes during the experimental procedure affected gene expression profiles in the cells. Finally, we identified the time point of 6 hr to be the most informative.

In the second experiment, global transcriptional responses were recorded in BJ-hTert (fibroblasts), G361 (melanoma), HepG2 (hepatocellular carcinoma), TK6 (lymphoblast) and U2OS (osteosarcoma) untreated cells, and 3 and 6 hr after IR dose of 5 Gy. Based on the observation of side-effects due to the experimental procedure, we added a chronological control to each time point. When clustering the profiles for each cell line separately, the largest clusters of each cell line depicted up- and down-regulation in response to IR. Computational promoter analysis identified significant enrichment of the binding site signatures of p53 and NF-Y among promoters of up- and down-regulated genes, respectively, pointing to the major role of these two transcription factors in mediating the transcriptional response of the irradiated cells. All cell lines had significant enrichment of functions in the down-regulated gene clusters. Importantly,

regulation of cell cycle, particularly mitosis, was the most widely enriched category in all cell lines, pointing to the activation of cell cycle checkpoints. Surprisingly, we did not find any common denominator among the cancerous cell lines that distinguishes them from the non-cancerous cell lines. A possible explanation is the abnormal nature of "non-cancerous" cultures, which are not equivalent to normal tissues.

Conclusions: In this work, we applied gene expression microarrays combined with computational tools to delineate transcriptional responses induced by IR in five human cell lines and to analyze the differences among them. The differences in the transcriptional response of cell lines to DNA damage show that it is crucial to compare results of several experimental models and avoid reaching global conclusions based on a study of a single cell line or tissue.

Our results emphasize the importance of the transcription factors p53 and NF-Y as key regulators of these responses. Potential new targets of p53 have been suggested in this study.

Key words: Functional genomics, DNA damage response, ionizing radiation, gene expression microarrays, high-density oligonucleotide arrays, clustering, functional analysis, promoter analysis.

1 Introduction

1.1 DNA damage response networks

Cell life is governed by a highly structured network of biochemical pathways that evolved to maintain its metabolism, and in higher organisms also to allow it to carry out specific functions according to tissue context. This carefully laid-out plan of operation may be perturbed by numerous physical and chemical environmental agents that damage cellular constituents. Notable among them are agents that damage the DNA, posing one of the greatest threats to the function and life of the cell and the organism [1, 2]. DNA damage stems from several sources. It inevitably occurs during normal DNA replication (e.g., via replication errors); it is constantly induced by intermediates of normal cellular metabolism, usually by reactive oxygen species formed during cellular respiration or inflammation; and it is inflicted by exposure to environmental physical and chemical agents that induce a large variety of chemical modifications in DNA components or strand breaks [3].

Cells possess intricate mechanisms to sense and handle the challenge posed by DNA damage. The essentiality of these mechanisms for cell life is reflected by the conservation of their core throughout evolution. Elements in the DNA damage response network can be generally divided into a three-layered hierarchy. At the top of the network are specialized DNA surveillance *sensors* that scan the genome for abnormalities. Once the sensors detect damage, they lead to the activation of *transducers* which amplify and convey the alarm message throughout the cell by modulating the activity of downstream *effectors* that in turn affect the biological endpoints of the damage response [4]. The efficiency and quality of cellular responses to DNA damage determine whether this insult will be repaired with no lasting effect on cellular life, or divert the cell from normal growth to programmed cell death, or end up in neoplastic transformation [3]. Understanding of DNA damage responses has broad implications for basic

life processes such as cell cycle control, aging, tissue development and degeneration. It is highly relevant for human health, primarily to coping with environmental hazards, cancer formation [5, 6], and many neurodegenerative disorders. Strong evidence for this is provided by genetic disorders caused by defects in cellular responses to DNA damage. Patients with such disorders exhibit acute predisposition to cancer, degenerative changes in specific tissues, premature aging, and body malformations (e.g., Bloom syndrome [7], xeroderma pigmentosum (XP) [8], hereditary non-polyposis colorectal carcinoma (HNPCC) [9, 10], ataxia-telangiectasia (A-T) [11], and Nijmegen breakage syndrome [12]). The central nervous system (CNS) seems to be especially sensitive to defects in DNA damage response [13-15], possibly due to its high oxidative stress and lack of cellular turnover.

Cellular responses to DNA damage have long been viewed mainly in terms of the concerted activation of DNA repair mechanisms and cell cycle checkpoints that are activated in order to prevent cell death during DNA replication or fixation of genetic alterations at the damage sites or transmission of unbalanced genetic content to daughter cells [16, 17]. However, studies that applied functional genomics approaches demonstrate that the damage-invoked network is much broader than DNA repair and cell cycle control [18-22]. These recent studies show that DNA damage sets off a wide array of signaling pathways that cover most aspects of cellular physiology, ranging from metabolic pathways to changes in protein turnover, cellular trafficking and cell-to-cell signaling. The biological mechanisms and the significance of most parts of this network are barely understood.

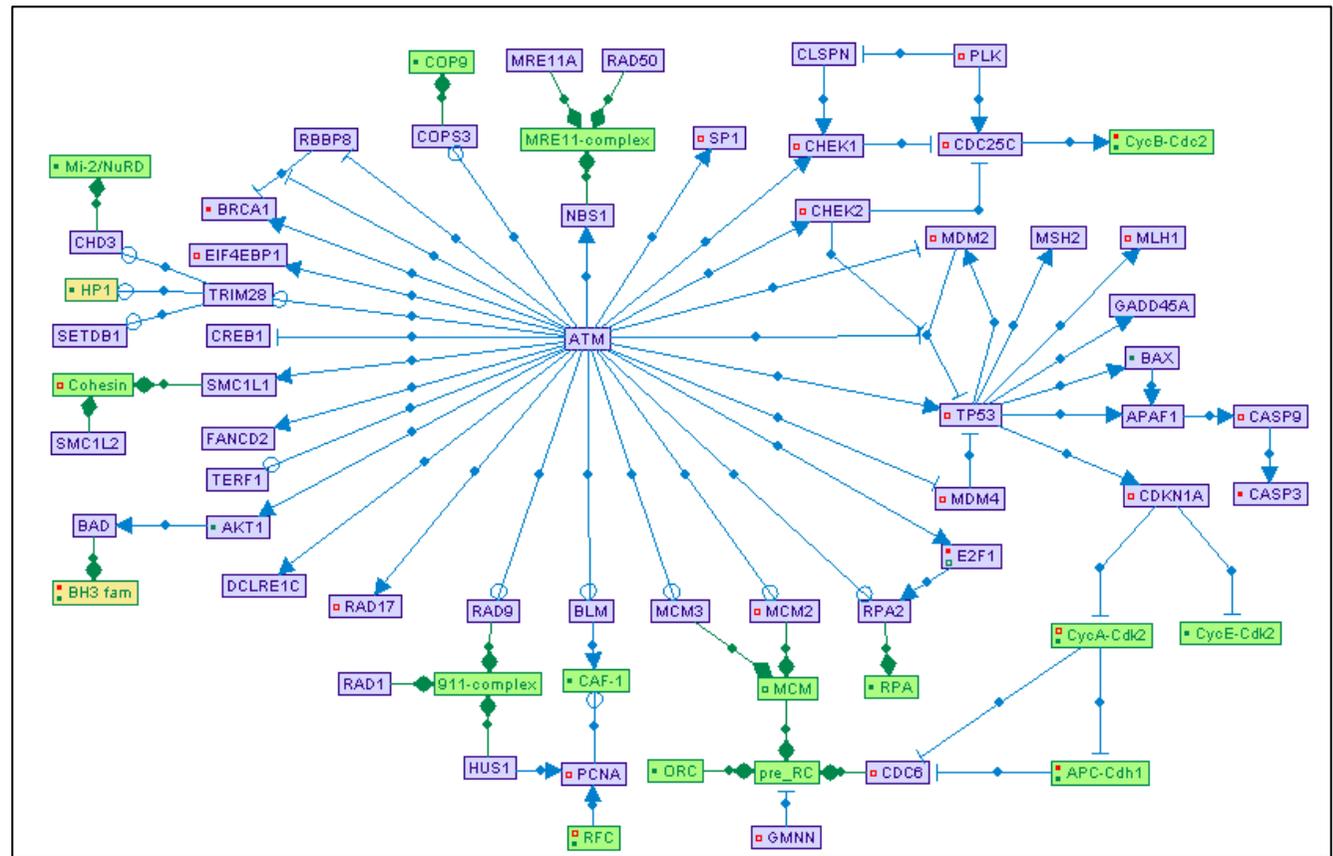
It is also becoming apparent that different tissues induce distinct damage responses, adding to the complexity of the DNA damage response network. In general, tissue sensitivity to DNA damage may be linked to its proliferation rate: terminally differentiated, post-mitotic cells tend to be more resistant (e.g., adult brain, muscle), while tissues with high cellular turnover are more sensitive (e.g., epithelia, bone marrow, spermatogonia and hair follicles) [23-27]. This model fits well with the general radiosensitivity of tumors, which are made up of actively proliferating

cells; however, it fails to explain the radiosensitivity of some of the most radiosensitive tissues – spleen and thymus, which in adults consist mainly of non-dividing cells, as well as the high radiosensitivity of bone-marrow haematopoietic stem cells, which are predominantly quiescent [23].

1.2 The DSB response concerted by ATM

The nuclear protein kinase ATM is positioned at the center of a physiological junction from which the cell activates a vast array of pathways in response to a specific DNA lesion, the double strand break (DSB). When DSBs crop up in the DNA, the cell activates an intricate web of pathways that includes DNA repair mechanisms, cell cycle checkpoints, and numerous other stress responses. ATM plays a pivotal role in the activation of all these branches of the damage response [6] by phosphorylating key players in each of the pathways. To date, more than 20 direct substrates of ATM have been studied in detail and found to control the signaling cascades that execute many physiological processes (**Fig. 1.2-1**). More than 800 candidate substrates have recently been identified [28, 29].

Figure 1.2-1. A portion of the ATM-regulated network. ATM is a master regulator of an intricate web of cellular responses induced by DNA double strand breaks. In the presence of such lesions in the DNA, ATM sets off a wide array of signaling pathways by directly phosphorylating numerous substrates. Among the processes modulated by ATM are cell-cycle checkpoints, apoptosis, DNA repair, gene transcription, and protein degradation. Only a fraction of the interactions are shown in the map: the entire network of all ATM interacting proteins plus their own documented interactions contains more than a hundred proteins with dozens of interconnections. This figure was generated using our SPIKE tool (see methods section). Briefly, violet nodes correspond to proteins; green nodes to protein complexes, and yellow nodes to protein families. Blue edges represent regulations: arrows correspond to activation; T shape edges – to inhibition; and open circles denote regulations whose effect is still not clear. Green edges represent association between nodes (e.g., association between a protein complex and its components). Red and green dots within a node indicate that not all the regulations and associations stored in SPIKE database for the node are displayed in the map.



Notable examples for ATM substrates are p53 that mediates the activation of G1/S checkpoint, DNA repair and apoptosis; the phosphorylation of the cell cycle checkpoints CHK1 and CHK2; and BRCA1 (for a recent review see [30]). A recurrent mode of operation in the tactic taken by ATM in modulating downstream pathways is its parallel regulation of several players within a target pathway. For example, ATM stimulates p53 activity by its phosphorylation on Ser 15 and augments this activation by directly phosphorylating its essential inhibitors Hdm2 and Hdmx [31-34]. In an analogous manner, ATM phosphorylates both BRCA1 and its inhibitor CtIP to achieve robust activation of this response arm [35, 36].

We are only just beginning to understand the very early events that lead to ATM activation in response to DSB. Recent evidence suggests that the complex containing the DNA repair proteins Mre11, Rad50 and Nbs1 (MRN complex) acts as sensor of DSB and is responsible for the recruitment and activation of ATM at DNA damage sites [37, 38]. In a proposed model, Mre11 and Rad50 form structural bridges between free DNA ends at DSB sites via the coiled-coil arms of Rad50 dimers, and NBS1 facilitates the recruitment of ATM and other downstream effectors by protein-protein interactions [39]. The nuclease activity of the MRN complex exerted by MRE11, which resects broken DNA ends to produce single-strand ends, was shown to be required for the activation of ATM.

The gene that encodes ATM is mutated in patients with the autosomal recessive disorder ataxia-telangiectasia (A-T). A-T is a devastating multifaceted disorder characterized by progressive degeneration of the cerebellum leading to severe neuromotor dysfunction; immunodeficiency that stems from compromised functioning of both the B- and T-cell arms of the immune system; premature aging; growth retardation; and extreme predisposition to cancer, mainly of lymphoid origin [11]. Significantly, A-T patients show acute sensitivity to ionizing radiation and other radiomimetic chemicals and therefore cannot be treated effectively for cancer using radiotherapy and commonly used chemotherapies.

The pleiotropic nature of A-T points to the high complexity of the DNA damage network and its essential role in the maintenance of proper functioning cells and tissues. The connection between compromised DNA damage response and cerebellar degeneration in A-T was intensely debated for many years. It was even suggested that in contrast to its role as a nuclear DNA damage protein in proliferating tissues, ATM plays other roles as a cytoplasmic protein in post-mitotic neurons of the CNS [40, 41]. Recent findings brought this debate to an end. One finding is associated with the discovery of the genetic disorder ATLD (AT-like disease), which shares many common features with A-T, including the cerebellar degeneration [42]. The responsible gene encodes the MRE11 DNA nuclease protein. This similarity between the two diseases suggested that the MRN complex is required for ATM activation by DNA damage. This was experimentally confirmed by Uziel et al. [38], strongly suggesting that the cerebellar degeneration in both A-T and ATLD results from the defective DNA damage response. Moreover, Frappart et al. [43] recently showed that while inactivation of the murine *Nbs1* gene, which encodes the Nbs1 subunit of the MRN-complex, is embryonic lethal, its conditional inactivation in neural tissues results in a combination of neurological anomalies, including cerebellar defects and ataxia.

1.3 Systems biology and functional genomics

Life sciences and biomedical research have undergone a revolutionary change in the last few years, with the emergence of a new paradigm, termed *systems biology*, which aims at a systems-level understanding of biological networks [44-50]. Biological research has traditionally applied reductionist experimental approaches whereby cellular systems are deconstructed into their elementary components (genes, proteins) and particular, isolated parts of the system are characterized. The transition to a new experimental paradigm in biology, often called the '*post-genome era*', was triggered by the rapid advance in the human genome project and in large-scale sequencing projects in other model organisms.

The availability of sequences of complete genomes allows one, in principle, to identify all the genes in an organism (and thereby also the entire collection of encoded proteins) — analogous to listing all the parts of a mechanical system. While such a catalog of individual components is invaluable for studying the system, it is not sufficient by itself for understanding the system's function. In the case of the living organism, we need to decipher how the components dynamically interact and regulate each other to form highly intricate physiological systems. This is the goal of the new field called *functional genomics*. More than merely assigning genes into functional categories, functional genomics aims at a comprehensive understanding of genetic networks: how gene products interact and regulate each other to produce coherent and coordinated physiological processes during the organism's development and in response to homeostatic challenges [51].

In contrast to the reductionist approach, functional genomics takes a holistic approach in which the cellular system is analyzed as a whole [52]. This systems-level approach has become feasible in biomedical research thanks to the combination of three indispensable factors. First, as noted above, the completion of the sequencing of genomes of various organisms providing us with entire blueprints of the 'program of life' in these species. Second, the maturation of novel high-throughput biotechnologies for large-scale analysis of cellular constituents that yield comprehensive views of life in a cell, a tissue, and ultimately the whole organism. Third, the development of powerful computational algorithms and data analysis tools. The large-scale sequencing projects and the novel high-throughput biotechnologies have transformed biology into an information-rich science. Experimental biological data are being generated at an unprecedented pace. Mining meaningful biological knowledge out of the huge volume of accumulated data is critically dependent on the availability of supporting bioinformatics tools. Therefore, this novel research paradigm is multidisciplinary and necessitates intimate collaboration between biologists and computational scientists.

1.4 Functional genomics technologies

The novel high-throughput functional genomics technologies analyze cellular constituents at various layers: from the level of the DNA sequence (the *genome* tier), through the expressed RNA molecules (the cellular *transcriptome* tier) to the level of the proteins (the *proteome* tier).

Prominent functional genomics tools for the study of cells at the DNA level are SNP (single nucleotide polymorphism) genotyping and CGH (comparative genomic hybridization) microarrays. SNP arrays enable scientists to conduct genome-wide linkage and association studies in order to discover genetic variations underlying complex human traits [53-55]. For example, the new generation of SNP-genotyping high density array manufactured by Affymetrix, the GeneChip Mapping 100K Set, allows the genotyping of more than 100,000 distinct human SNPs in a single assay [56]. CGH microarrays have significantly increased the resolution of conventional CGH in detection of DNA copy number aberrations; this greatly improves the ability to characterize the chromosomal imbalances resulting in gain and/or loss of genomic material that are recurrent in human cancers [57, 58].

Major functional genomics technologies for analysis of the cellular transcriptome are gene expression microarrays (see detailed description in Section 1.3 below) and SAGE (serial analysis of gene expression). Gene expression microarrays rely on hybridization between RNA molecules extracted from the examined cells (or cDNA molecules derived from them), and complementary probes deposited or synthesized on the array [59-61]. SAGE is based on the isolation of unique sequence tags (typically 10-14 bp in length) from defined locations in mRNA molecules, and concatenation of these tags in a serial way into long DNA molecules for a lump-sum sequencing indicating the expression level of the corresponding RNA molecules [62]. These technologies allow parallel recording of expression levels of thousands of genes in a single assay, providing genome-wide snapshots of the cellular transcriptome under the examined biological conditions. They have proven to be very powerful tools for molecular characterization of pathological conditions and global delineation of transcriptional programs induced by various stimuli, or

programs associated with physiological processes such as differentiation, cell cycle, aging and neoplastic transformation [63-66].

Another functional genomics approach that greatly enhances the study of transcriptional networks combines chromatin immunoprecipitation (ChIP) and promoter microarrays. This technique, also termed ‘ChIP-on-chip’, enables genome-scale identification of promoters that are bound by specific transcription factors (TFs) under certain conditions, in a single experimental assay [67, 68]. ChIP-on-chip was recently applied in a seminal study by Lee et al. [67] to map all TF-promoter binding relationships in yeast under standard growth conditions. The microarrays used in this study contained probes corresponding to the promoters of all known and predicted genes in *S. cerevisiae*. These arrays were reacted with all known TFs in this organism, yielding a comprehensive map of the transcriptional network controlling yeast life under standard growth conditions. The approach was also applied to mammalian cells to identify genome-wide direct targets of many TFs, including E2F, c-Myc and NF- κ B [69-71].

Mass spectrometric analysis for protein identification is a major technology in systems analysis of the cellular proteome (*proteomics*) [72]. It has been applied in recent years to identify protein-protein interactions on a proteome-wide scale (the *interactome*) in organisms ranging from yeast to human [73-75]. In addition, mass spectrometric techniques have been established for the analysis of post-translational modifications, such as phosphorylation and glycosylation [72, 76]. At the same time, protein microarrays are being developed. They can be divided into two categories according to use. In the first category are chips for profiling protein levels. Dozens of antibodies with high specificity are deposited on the chip to enable comparison of the expressions of the respective protein antigens from different samples. This technology was recently applied to identify proteins that are upregulated in specific cancers [77]. Furthermore, antibodies that recognize specific modified states of their target proteins (e.g., recognizing a target that is phosphorylated on a certain site) are used to measure the level of various post-translational modifications. In the second category are chips for biochemical characterization of

the function of proteins. The proteins themselves are deposited on the chip and assayed in parallel for a specific biochemical reaction. For example, in a recent study [78], a protein chip on which all *S. cerevisiae* proteins were spotted was assayed for interactions with specific substrates to identify, among others, all calmodulin-binding proteins.

Another major advance in functional genomics came with the discovery of RNA interference (RNAi) and its harnessing as a research tool [79, 80]. RNAi has dramatically expanded the scope and versatility of cell culture systems for the analysis of gene function and involvement in biological processes. Prior to this discovery, obtaining cell lines deficient for a specific protein depended largely on the availability of cells from human patients affected with genetic disorders, or from knockout mice. Methods for silencing the expression of specific genes using antisense oligos allowed only transient silencing and their efficiency was limited. The advent of RNAi technology changed this situation. The introduction of short (21 nt) double-stranded RNA (or an RNA oligo that acquires a secondary small hairpin structure that imitates double-stranded RNA) into cells results in the degradation of complementary cellular mRNA molecules via activation of the RISC complex (RNA-induced silencing complex) [79]. Furthermore, it is now possible to express ectopically such small hairpin RNA (shRNA) to generate cell lines that are stably knocked-down for the target gene [81, 82]. Several labs have undertaken the task of systematically constructing cell lines that collectively will be knocked-down for most genes in the human genome and in other model organisms [79, 80, 83, 84]. Initial progress towards this goal was recently reported when a large-scale RNAi screen was carried out in human cells to identify new components of the p53 pathway [85].

1.5 Gene expression microarrays: platforms

During its first few years gene expression microarray technology suffered from 'infancy maladies' [86]. Now that it has reached the stage where it yields accurate and reproducible results, it has become a standard research tool in molecular biology labs [87].

Gene expression microarrays come in three platforms that differ in the nature of the probe molecules used to detect RNA levels, and the method of placing them on the array surface. The three platforms are called cDNA microarrays, high-density oligonucleotide arrays, and long-oligonucleotide arrays.

cDNA microarrays use as probes PCR products (typically several hundred bps in length) of cDNA libraries. These probes are mechanically deposited ('spotted') on the array surface by a robotic arm carrying a pin head [59, 60, 88] in a simple, widely used procedure. The problems of irregular spot size and shape and probe concentration are addressed by co-hybridization to the same chip of two samples, a test sample and a reference sample, each labeled with a different fluorescent dye that allows relative measurements of gene expression levels.

The high-density oligonucleotide array technology was developed by Affymetrix (Santa Clara, CA). In this platform, called GeneChip, 25-mer oligonucleotide probes are synthesized directly on the array using photolithography and photosensitive oligonucleotide synthesis chemistry [51, 61, 89]. Each target gene in these arrays is represented by a probe set of 11-20 perfect match (PM) oligos complementary to different regions along the respective mRNA molecule. A parallel set of mismatch (MM) oligo probes that differ from the PM probes by a single nucleotide at the central position serves as a control that improves the discrimination between specific and nonspecific hybridization signals. Recent studies questioned the utility of the MM probes as negative controls [90, 91]. The design of the Affymetrix GeneChip requires knowledge of the sequence of the target genes. Affymetrix offers whole- or near-whole genome GeneChip expression arrays for many organisms whose genome sequencing is near completion, including bacteria, yeast, worm, fly, chicken, rat, mouse and human.

The third gene expression platform uses as probes in-situ synthesized long oligonucleotides (60-70 mers) or prefabricated oligonucleotides that are spotted on the arrays. In a study of the dependence of specificity and sensitivity of hybridization signals on probe length, Hughes et al [92] found that 60-mer probes yield optimal results. Long oligonucleotide arrays commercially

manufactured by Agilent Technologies (Palo Alto, CA), improved probe deposition by an ink-jet printing process. Mechanically spotted long oligonucleotide arrays were produced by some academic groups [93].

Several studies compared the performance of different platforms [94, 95]. cDNA microarrays were found to be more prone to cross-hybridizations but are cheap and affordable to academic labs. Oligonucleotide-based arrays have higher flexibility in probe selection, which improves specificity. The 60-mer oligonucleotide arrays are highly sensitive (and were shown to detect genes expressed at levels as low as one copy per cell [92]). The GeneChip results are robust thanks to multiple probes used to measure each target. Extensive efforts are underway to understand the factors affecting inter-platform and inter-lab variability and to set standards for the gene expression technology [87, 96-99].

2 Research Goals and Specific Aims

The major goal of our research was to examine the difference between the transcriptional responses to DSBs in several human cell lines using experimental and computational analysis.

Specific aims:

- Delineate transcriptional responses to DNA DSBs in human cells by applying state-of-the-art computational techniques to our experimental datasets.
- Identify by computational means transcription factors that control the transcriptional response induced by DNA damage and dissect the damage response network into arms mediated by these regulators.
- Elucidate biological endpoints of the transcriptional program induced in response to DNA damage.
- Find the common denominator of DNA damage responses in five different human cell lines.
- Dissect the differences in damage response between the cell lines.

3 Materials and Methods

3.1 Materials

The following buffers were made using ddH₂O, unless otherwise mentioned.

RNA stands for total RNA, unless otherwise mentioned.

3.1.1 Solvents, powders and general reagents

Agarose (Low melting) - Kodak, Rochester, NY, USA

Chloroform - Merck, Darmstadt, Germany

DEPC treated H₂O (Diethylene Pyrocarbonate) - Biological industries, Kibbutz Beit Haemek,
ISRAEL

EDTA (Ethylenediamine-tetraacetic acid) disodium salt, 0.5 M - Ambion, Austin, TX, USA

Ethanol - Merck, Darmstadt, Germany

Ethidium bromide - Amersco Inc., Solon, Ohio, USA

Glycerol - Sigma-Aldrich, Inc., St. Louis, MO, USA

Neocarzinostatin (NCS) - Kayaku Co., Japan

PBS (Dulbecco's Phosphate Buffered Saline; 0.15M NaCl, 10mM phosphate buffer PH 7.4) -
Biological industries, Kibbutz Beit Haemek, ISRAEL

Phenol:Chloroform:Isoamyl Alcohol (25:24:1) - Gibco BRL, Life Technologies, Grand
Island, N.Y, USA

2-propanol- Merck, Darmstadt, Germany

Sodium acetate (NaOAc) - Sigma-Aldrich, Inc., St. Louis, MO, USA

3.1.2 Cell culture media

DMEM (for BJ-hTert, HepG2, U2OS cell lines) - Biological Industries, Kibbutz Beit Haemek,
ISRAEL

McCoy (for G361 cell line) - Biological Industries, Kibbutz Beit Haemek, ISRAEL

RPMI 1640 (for TK6 cell line) - Biological Industries, Kibbutz Beit Haemek, ISRAEL

Each medium was supplemented with:

Fetal calf serum (FCS) 10%* (v/v)

Penicillin/Streptomycin 1% (v/v)

L-glutamine 1% (v/v)

* Medium with 10% FCS was used for growth of all cell lines, except for BJ-hTert cells, where 20% FCS was used.

3.1.3 Cell culture growth materials

DMSO (Dimethylsulfoxide) - Merck, Darmstadt, Germany

Heat-inactivated fetal calf serum (FCS) - Biological industries, Kibbutz Beit Haemek, ISRAEL

L-glutamine Solution - Biological industries, Kibbutz Beit Haemek, ISRAEL

Pen-Strep solution - (Penicillin: 10,000 units/ml; Streptomycin: 10mg/ml) - Biological Industries, Kibbutz Beit Haemek, ISRAEL

Trypsin versene – Biolab, Jerusalem, Israel

3.1.4 Cell lines

BJ-hTert, a normal human fibroblast line immortalized by human telomerase gene transfection [100], was obtained from the lab of Dr. Xiao Fang Wang (Duke University, North Carolina, USA).

G361, a human melanoma cell line, was obtained from ATCC.

HepG2, a human Hepatocellular carcinoma, was obtained from Dr. Hagit Eldar-Finkelman of our department.

TK6, a human lymphoblastoid cell line, was obtained from Prof. Leona D. Samson, (Massachusetts Institute of Technology, Cambridge, MA, USA)

U2OS, a human osteosarcoma cell line, was obtained from Dr. Aart G. Jochemsen, (Leiden University Medical Center, The Netherlands).

3.1.5 Materials for analysis of RNA

ΦX174 DNA/BsuRI (Hae III) marker (Molecular weight ladder) - MBI Fermentas, Hanover, MD, USA

10x TBE buffer - Ambion, Austin, TX, USA

TBE buffer was diluted to 0.5x TBE buffer, and was used for preparation of agarose gels and as a running buffer. When preparing agarose gels, it is necessary to add ethidium bromide to a concentration of 0.5 µg/mL (for DNA and RNA staining).

6x loading dye solution – MBI Fermentas, Hanover, MD, USA

Removal of RNase contamination from glass and plastic surfaces:

RNaseZAP® (RNase decontamination solution) - Ambion, Austin, TX

Total RNA Isolation:

TRIzol® reagent (for RNA isolation) - Gibco BRL, Life Technologies, Grand Island, N.Y, USA

RNeasy Plus Mini Kit - Qiagen, Valencia, CA, USA

RNA precipitation:

GlycoBlue™ coprecipitant - Ambion, Austin, TX, USA

Removal of contaminating DNA:

DNA-free™ (DNase treatment & Removal kit) - Ambion, Austin, TX, USA

Target labeling for expression analysis:

GeneChip® One-Cycle Target Labeling and Control Reagents - Affymetrix Inc., Santa Clara, CA, USA

This package contains:

- GeneChip® Poly-A RNA Control Kit

- GeneChip[®] One-Cycle cDNA Synthesis Kit
- GeneChip[®] IVT Labeling Kit
- GeneChip[®] Sample Cleanup Module
- GeneChip[®] Hybridization Control Kit

GeneChip[®] Hybridization, Wash and Stain Kit Reagents - Affymetrix Inc., Santa Clara, CA, USA

3.1.6 Miscellaneous Supplies

Culture flasks (25cm² ,75cm² ,150cm² surface area flasks) – Corning, Corning, NY, USA

Deep freeze tubes - Corning, Corning, NY, USA

Non-stick RNase-free 1.5 mL Microfuge Tubes - Ambion, Austin, TX, USA

Phase Lock Gel (PLG) 2 ml tubes; "Heavy" - Eppendorf AG, Hamburg, Germany

Polypropylene Centrifuge Tubes (1.5, 15, 50 mL) - Greiner-Frickenhansen, Germany

RNase-free Microfuge Tubes – Ambion, Austin, TX, USA

Sterile pipettes and tips - Costar, Cambridge, MA, USA; Biby Sterilin Ltd., Stonen Staffordshire, UK.

Microarrays:

Human Genome U133A 2.0 - Affymetrix Inc., Santa Clara, CA, USA

Human Genome U133 Plus 2.0 - Affymetrix Inc., Santa Clara, CA, USA

3.1.7 Instruments

Hemocytometer, Hausser Scientific, Horsham, PA, USA

X-irradiation: Faxitron[®] Model RX-650 RADIATION SOURCE (Faxitron X-ray Corporation, Wheeling, IL, USA)

Dosimetry: model 35050A dosimeter (Keithley Instruments, Cleveland, Ohio, USA).

NanoDrop[®] ND-1000 - NanoDrop Technologies, Wilmington, Delaware USA.

Fluidics Station 450 - Affymetrix Inc., Santa Clara, CA, USA. The instrument is located at the Functional Genomic Facility at the Sheba Medical Center.

GeneChip[®] Scanner 3000 - Affymetrix Inc., Santa Clara, CA, USA. The instrument is located at the Functional Genomic Facility at the Sheba Medical Center.

Hybridization Oven 640 - Affymetrix Inc., Santa Clara, CA, USA. The instrument is located at the Functional Genomic Facility at the Sheba Medical Center.

3.2 *Methods*

3.2.1 Cell culture

Growth

The cultures were incubated at 37 °C in 5% CO₂ and 95% humidity atmosphere. In order to maintain the cultures at a logarithmic phase of growth, they were split when reaching approximately 80% confluence (attached cells) or maintained at a density of 2×10^5 - 1×10^7 cells per ml (suspension cultures).

Harvesting adherent cultures

Medium was removed from culture. Trypsin (dissociating agent) was added for 30 seconds, and then removed. Trypsin was added again, and cells were observed under a microscope. When cells become rounded and loosen, fresh medium was added, and detached cells were transferred to new flasks.

Harvesting suspended cultures

Cells were precipitated at 1,200 rpm for 7 minutes and re-suspended with fresh medium.

Freezing

Cells at logarithmic phase of growth were precipitated at 1,200 rpm for 7 minutes and re-suspended with a medium containing 10% DMSO. Aliquots of 1 mL were frozen at a -70 °C for 24 hours and then transferred to liquid nitrogen (-186 °C).

Thawing

Cells were removed from frozen storage and quickly thawed in a 37 °C water bath by gently agitating the vial. As soon as the ice crystals melt, cells were transferred gently into a tube containing pre-warmed growth medium, precipitated at 1,200 rpm for 7 minutes, re-suspended with fresh medium and transferred into a flask.

3.2.2 Induction of DNA double strand breaks (DSBs)

NCS treatment

Cells grown to approximately 70% confluence were exposed in the dark to 200ng/mL neocarzinostatin (NCS), a radiomimetic drug that induces DNA DSBs [101, 102], which had been dissolved in PBS immediately prior to the experiment. After treatment, cells were incubated in the dark at 37 °C for 2, 4, 8 or 12 hours prior to RNA extraction.

Radiation treatment

Cells grown to a density of approximately 3×10^6 - 1×10^7 cells per plate, were irradiated at 5 Gy of X-rays. After irradiation cells were incubated for 3 or 6 hours before RNA extraction.

3.2.3 Total RNA isolation

Two methods for total RNA extraction were employed in this work, based on the TRIzol[®] reagent (Gibco BRL) or the RNeasy Mini Kit (Qiagen).

Total RNA isolation with Trizol:

Total RNA was isolated with TRIzol commercial reagent, a mono-phasic solution of phenol and guanidine isothiocyanate, which improves the single-step RNA isolation method developed by Chomczynski and Sacchi [103]. During sample homogenization or lysis, TRIZOL Reagent maintains the integrity of the RNA, while disrupting cells and dissolving cell components.

Homogenization: Cells were lysed directly in the culture dish by adding 1.5 mL of TRIZOL Reagent per 64 cm² and passing the cell lysate several times through a pipette. The homogenized samples were incubated for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes.

Phase separation, RNA precipitation and RNA wash were performed according to the manufacture's instructions.

Redissolving RNA: The RNA pellet was dried at 55°C for 5-8 minutes. RNA was dissolved in DEPC-treated water to a concentration of 1-2 µg/µL by passing the solution a few times through a pipette tip and incubating it for 5-10 min at 55°C.

Removal of genomic DNA: RNA samples were treated with Ambion's DNafree™ Kit designed to remove contaminating DNA from RNA preparations and subsequently remove the DNase and divalent cations from the sample.

The kit contains RNase-free DNaseI enzyme, an optimized 10X Reaction Buffer (100mM Tris-HCl pH 7.5, 25mM MgCl₂, 5mM CaCl₂), and a DNase Inactivation Reagent. DNafree™ Kit designed to remove trace to moderate amounts of contaminating DNA (up to 50 µg DNA/ml RNA) from purified RNA to a level that is mathematically insignificant by RT-PCR.

Purifying the RNA: Phenol:chloroform:isoamyl alcohol (PCI, in a ratio of 25:24:1, respectively) was added in equal volume to the aqueous phase and samples were shaken vigorously for 15 seconds. The RNA-PCI mix was transferred into Eppendorf Phase Lock Gel™ (PLG) tubes, which separate the aqueous and organic phase with a solid inert barrier. The solid barrier enables increased yields of the aqueous phase and minimal risk of interphase sample

contamination. Samples were centrifuged for 2 minutes at 14,000 rpm at 4°C. The upper aqueous phase was transferred to a fresh tube. This procedure was redone. RNA was precipitated from aqueous phase by mixing with 10% volume of 3M sodium acetate pH 5.2, glycoblue (50 µg/ml) and 80% volume of isopropanol (in respect to the total volume of RNA and sodium acetate). Samples were stored at -20°C for 2-20 hours and then centrifuged at 14,000 rpm for 30 minutes at 4°C. Supernatant was removed and RNA pellet was washed twice with 1ml 75% ethanol, followed by centrifugation at 14,000 rpm for 5 minutes at 4°C. RNA pellet was dried at 55°C for 5-8 minutes, and dissolved in DEPC-treated ddH₂O.

Total RNA isolation with RNeasy Plus Mini Kit:

High-quality total RNA has been successfully isolated from mammalian cells (such as cultured cells and lymphocytes) using the RNeasy Plus Mini Kit from QIAGEN.

The RNeasy Plus Mini Kit is designed to purify RNA from small amounts of animal cells or tissues. Biological samples are first lysed and homogenized in a highly denaturing guanidine-isothiocyanate-containing buffer, which immediately inactivates RNases to ensure isolation of intact RNA. The lysate is then passed through a gDNA Eliminator spin column. This column, in combination with the optimized high-salt buffer, allows efficient removal of genomic DNA.

Ethanol is added to the flow-through to provide appropriate binding conditions for RNA, and the sample is then applied to an RNeasy spin column, where total RNA binds to the membrane and contaminants are efficiently washed away. High-quality RNA is then eluted in 50 µl of water.

An isopropanolic precipitation was then performed in order to concentrate the RNA and remove leftovers of RNeasy buffers.

3.2.4 Quantification and quality assessment of RNA

RNA yield was quantified by spectrophotometric analysis using the convention that 1 absorbance unit at 260 nm equals 40 µg/ml RNA. The absorbance was checked at 260 and 280

nm for determination of sample concentration and purity. The A260/A280 ratio should be close to 2.0 for pure RNA (ratios between 1.9 and 2.1 are acceptable). We used the NanoDrop® ND-1000 spectrophotometer, which measures 1 µl samples with high accuracy and reproducibility.

The quality of the RNA was further assured using gel electrophoresis.

3.2.5 Preparation of the RNA samples for hybridization

The preparation of samples for hybridization was done as instructed in the Affymetrix GeneChip® expression analysis technical manual. The main steps are depicted in **Fig. 3.2.5-1**.

Poly(A) RNA controls were added to the 5 µg RNA according to the manual in order to provide exogenous positive controls for monitoring the target labeling process. Each eukaryotic GeneChip probe array contains probesets for several *B. subtilis* genes that are absent in eukaryotic samples: lys, phe, thr, and dap.

3.2.6 Target Hybridization

The fragmented and labeled cRNA was hybridized to Affymetrix high-density microarrays. In the first microarray experiment, cRNA was hybridized to Human Genome U133A 2.0 array. This array includes 18,400 transcripts and variants that represent approximately 14,500 well-characterized human genes. In the second microarray experiment cRNA was hybridized to Human Genome U133 Plus 2.0 array, representing over 47,000 transcripts and variants including 38,500 well-characterized human genes.

The target was hybridized to the array for 16 hours at 45 °C, Rotated at 60 rpm. Immediately after completion of hybridization, the array was washed with various salt concentrations buffers to remove free cRNA molecules. The array was then stained with streptavidin-phycoerythrin (SAPE) that binds to the biotin residues labeling the cRNA, stained with biotinylated anti-streptavidin antibodies, and stained with SAPE again. This process ensures amplification of the fluorescent signal and reduction of background noise. Once the stains were accomplished, the array was scanned with GeneChip® Scanner 3000.

One-Cycle Target Labeling

(for 1-15 µg total RNA or 0.2-2 µg mRNA)

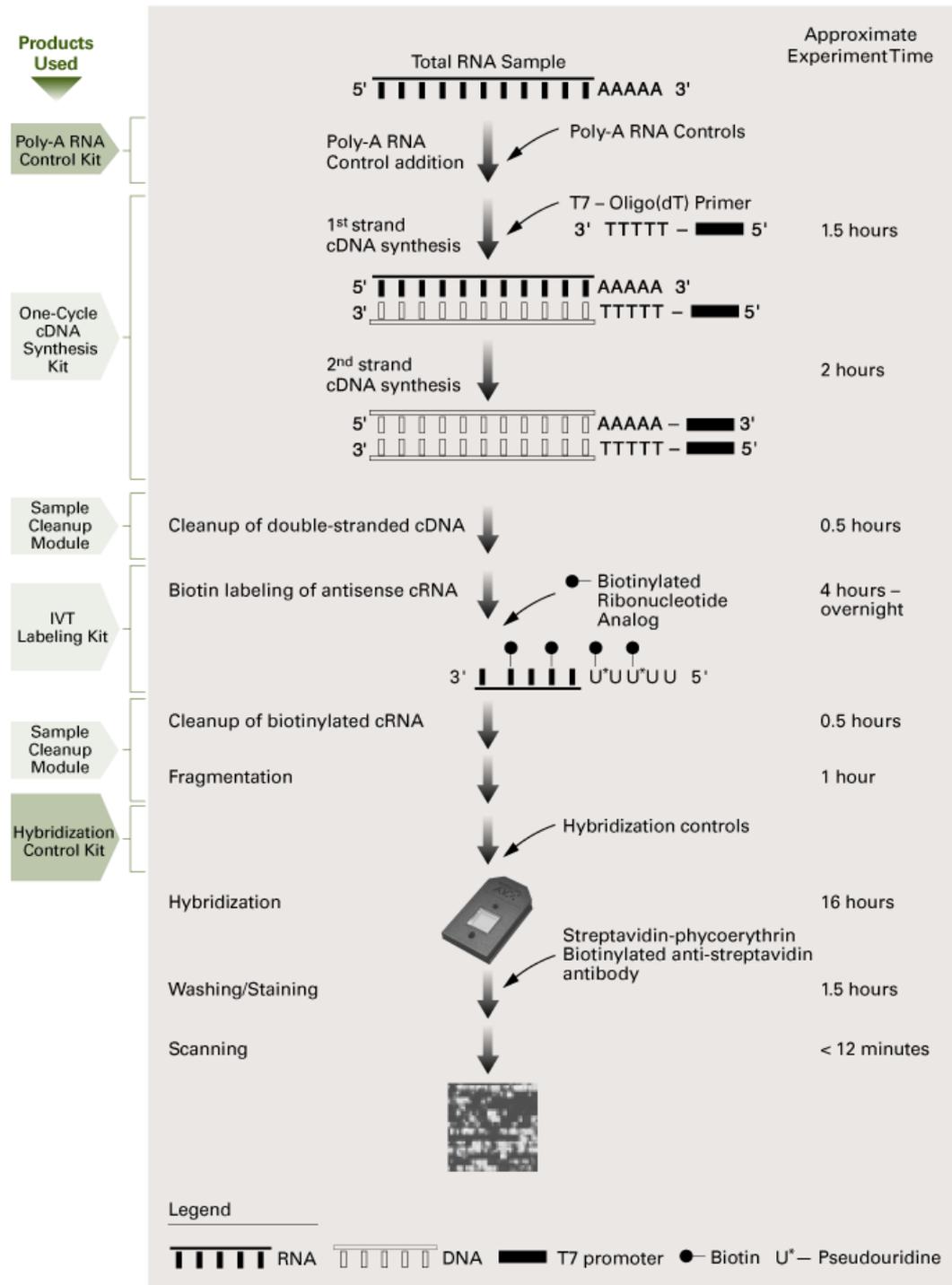


Figure 3.2.5-1. Affymetrix one-cycle target labeling steps

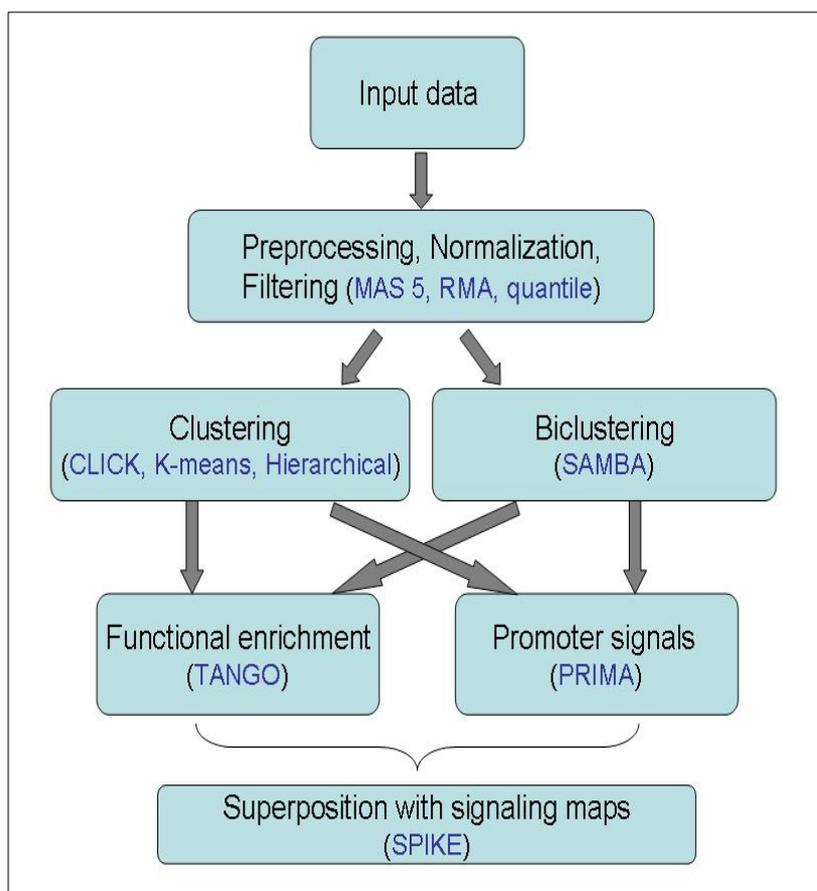
3.2.7 Analysis of gene expression data

Large-scale gene expression microarray experiments yield an enormous amount of data. Mining meaningful biological insights from the raw data poses a major bioinformatic challenge. We adopted an integrative approach for the analysis of the data that includes the following steps:

- Intensity signal extraction.
- Chip normalization.
- Gene clustering and biclustering based on the chip data.
- Enrichment analysis of functional categories and cis-regulatory promoter elements.
- Superposition of the microarray data on signaling maps that reflect current knowledge on cellular signal transduction pathways.

The methods and algorithms applied at each step are described below. The analysis flow is presented in **Fig. 3.2.7-1**. This integrative approach was designed and developed in the labs of Professors Shamir and Shiloh by Dr. Amos Tanay, Chaim Linhart, Dr. Roded Sharan and Dr. Rani Elkou. Each application was first implemented and tested separately, and then integrated into a general analysis package by Adi Maron-Kats and Israel Steinfeld. This integrated package, called *EXPANDER* (EXpression Analyzer and DisplayER) [104], serves as the central platform for the integration of all the microarray data analysis algorithms developed in our lab (<http://acgt.cs.tau.ac.il/expander/>).

Figure 3.2.7-1. Analysis flow of gene expression datasets. A flow chart illustrating the analysis of gene expression datasets. This integrated approach, implemented in the EXPANDER package, starts with the initial preprocessing steps of signal extraction, normalization and filtering, and continues through subgroup analysis (clustering or biclustering) to high-level statistical analyses that seek enriched functional categories and cis-regulatory promoter elements in the clusters/biclusters. In the last step, we superimpose the results on signaling maps that reflect current biological knowledge, using the SPIKE knowledge-base of signaling pathways, which is under development in our labs. Algorithms and tools applied in each step are indicated. Currently, part of the preprocessing and filtering steps, as well as the SPIKE tool, are not yet integrated in the Expander package.



Extraction of expression signals

Our microarray project used the Affymetrix GeneChip technology. In these arrays each target is probed by a set of 11-20 pairs of perfect-match (PM) and mismatch (MM) probes, each of which is 25 bp long and is complementary to a different region along the respective mRNA molecule. Arrays were scanned using an Affymetrix scanner, and images were analyzed using Affymetrix image analysis software to yield 'CEL files' that assign an intensity level for each probe ('cell') in the chip. Several methods are used for summarizing the intensity levels obtained for probes in the same probe set into a representative signal that is indicative of the expression level of the respective target gene [91].

Affymetrix Microarray Analysis Suite version 5 (MAS 5) assumes the following model to describe the relationship between the intensity of signals measured by the probes in a specific

probe set, and the expression level of their target gene (i.e., the concentration of the gene's RNA):

$$PM_{ij} - MM'_{ij} = \epsilon_{ij} * \theta_i$$

Where:

$i=1, \dots, I$ is the index of the chip (in an experiment that includes I chips)

$j=1, \dots, J$ is the index of the probe within a probe set.

PM_{ij} is the signal measured by PM probe j of the probe set in chip i .

MM'_{ij} is the signal measured by MM probe j of the probe set in chip i , manipulated in cases where it is above perfect-match values to prevent negative values in the log transformation.

θ_i is the expression level of the target gene in the sample probed by chip i .

$(1 - \epsilon_{ij})$ is a term that reflects random error, observed to be generally proportional to the probe signal, and therefore introduced as a multiplicative error term.

The measurement error is therefore: $(\epsilon_{ij} - 1) * \theta_i$. We can write equivalently:

$$\log(PM_{ij} - MM'_{ij}) = \log(\theta_i) + \log(\epsilon_{ij});$$

Here it is assumed that the error terms ϵ_{ij} follow log-normal distribution.

The summary E_i of the probe set signals that is appropriate to this model is a (weighted) average over logarithm of the signals.

$$\text{Log}(E_i) = \sum w_j * \text{Log}(PM_{ij} - MM_{ij})$$

The weights w_j are computed using Tukey biweight's function, in which the distance of each data point from the median determines how each value is weighted. Outliers far from the median contribute little to the average making the summary resistant to them.

Recently, the Robust Multi-array Average (RMA) method was introduced and was demonstrated by several studies to outperform the previous methods [91, 105]. Notably, these studies questioned the utility of the MM probes as negative controls and recommended ignoring them in the computation of intensity signals (the MM signals are still utilized in estimation of

global background signals, which we do not discuss here). RMA differs from MAS 5 also in the affinity parameter incorporated into its model, relating to the observed large variability among signals measured by different probes within the same probe set [106]. RMA assumes the following model:

$$\text{Log}(\text{PM}'_{ij}) = \log(\alpha_j \theta_i) + \varepsilon_{ij}$$

where PM' denotes PM values after background correction and normalization; α_j is the affinity coefficient of probe j ; and the error terms ε_{ij} are assumed to follow a normal distribution. RMA estimates θ_i using a robust linear fitting procedure.

In the initial analysis of this project, both the MAS 5.0 and RMA methods were utilized for extraction of intensity signals. There was no significant difference between the results obtained by the two methods, and we chose to continue the analysis with the RMA method, in its implementation provided by the BioConductor project [107]. We preferred to use RMA over Affymetrix' MAS5 for two reasons: First, several studies indicated that the mismatch signals are correlated with the mRNA concentration of their corresponding gene; i.e., they themselves contain information on the expression level of the genes, hence subtracting their signals from the perfect-match ones, as MAS5 does, may add noise to the measurement and therefore be counterproductive [91]. RMA ignores the mismatch probes and computes expression levels based only on perfect match signals. Second, while MAS5 uses global scaling to normalize between arrays, RMA applies the quantile normalization that was demonstrated to perform better [105].

Arrays normalization

The comparison of expression levels measured under different conditions should be preceded by removal of systematic biases between arrays. The process of removing such biases is called normalization, and several methods were developed for this task. The normalization scheme implemented by Affymetrix analysis software is a basic one that computes a single scaling factor

per chip. Multiplying all intensity signals in an array by the scaling factor brings the average signal to a fixed predefined level.

Several studies pointed out that systematic variation between chips is often intensity- or spatially-dependent [105, 108]. Such non-linear biases cannot be removed by global scaling and necessitate more advanced approaches. Yang et al. [108] introduced the lowess normalization scheme that computes a normalization function using local regression. Using lowess, the mean signal is equalized among the arrays over the entire range of intensities (neutralizing non-linear, intensity-dependent biases), and between all spatial sectors on the chip. An even more stringent normalization scheme, quantile normalization, was introduced by Bolstad et al. [105]. This scheme forces the distribution of signals in all analyzed chips to be identical. A recent comparative study reported that quantile normalization outperforms the other methods in removing systematic biases while retaining true biological variation. We adopted quantile normalization in our project.

Filtering

After normalizing the arrays to a common scale, we focused our high-level analyses on the set of *responding genes*; that is, we filtered out genes that are either not expressed in the test conditions or do not respond to the examined perturbations.

Affymetrix assigns a Detection call to each probeset. This call indicates whether a transcript is reliably detected ('Present'), not reliably detected ('Marginal') or not detected ('Absent'). We used the detection calls to filter out probe sets that get 'Absent' or 'Marginal' calls by Affymetrix software across more than two thirds of the arrays in the dataset. Typically, signals of such probe sets are at the lower tail of the intensity distribution, close to the level of background noise.

We then scanned the data for genes that show a variation above a predefined threshold across the test conditions. The number of replicates used in our experiments (1-3 repeats) usually does not allow sufficient statistical power to robustly detect differentially expressed genes; rigorous statistical tests at this stage usually pass too few genes for subsequent analyses. Therefore, at this

initial filtering step, when data consisted of one replicate only, we applied a naive fold-change (FC) filtering criterion of 1.5, 1.75 or 2. Statistical tests with controlled rate of false positive discoveries are applied in our downstream analyses in order to detect global phenomena within the set of responding genes. We used a lower FC threshold than that usually used in microarray analysis — because the RMA method significantly narrows the distribution of expression levels and of the fold changes compared to Affymetrix MAS5 package [109].

Biological replicates increase the power to detect biologically significant gene expression differences. When trying to identify differences between a treatment and a control group, accurate estimates of the biological variability within the groups are essential to determine if the differences between the groups are meaningful [110]. When data consisted of 3 replicates, we used one-way Analysis of Variance (ANOVA) to filter out probesets for which the variance of values between conditions was smaller than the variance within replicates. We then applied a smaller FC threshold, to exclude probesets for which the change in expression between the tested conditions and the control was not sufficiently large. The exact value of the threshold depended on the particular analysis goals.

Identification of major expression patterns in the dataset

In the next step, we subject the set of responding genes to cluster analysis. Clustering algorithms that are applied to gene expression data partition the genes into distinct groups according to their expression patterns over a set of experimental conditions. Such partition should assign genes with similar expression patterns to the same cluster (keeping the high *homogeneity* level of genes in the same cluster) while retaining the distinct expression pattern of each cluster (ensuring high *separation* between clusters). Cluster analysis eases the interpretation of the data by reducing its complexity and revealing the major underlying expression patterns. We used several clustering algorithms, including K-means [111], hierarchical clustering [112], and *CLICK* [113]. The last one is a graph theory-based algorithm developed in Shamir's group

by Dr. Roded Sharan. *CLICK* was demonstrated to outperform other algorithms according to several objective figures of merit [114].

As expression data accumulate and profiles over dozens of different biological conditions are readily available, clustering becomes too restrictive. Clustering algorithms globally partition the genes into disjoint sets according to the overall similarity in their expression patterns; i.e., they search for genes that exhibit similar expression levels over all the measured conditions. Such a requirement is appropriate when small to medium size datasets from one or a few related experiments are analyzed, as it provides statistical robustness and produces results that are easily visualized and comprehended. But when larger datasets are analyzed, a more flexible approach is needed. A *bicluster* is defined as a set of genes that exhibit significant similarity over a subset of the conditions. A biclustering algorithm can dissect a large gene expression dataset into a collection of biclusters, where genes or conditions can take part in more than one bicluster. A set of biclusters can thus characterize a combined, multifaceted gene expression dataset [115]. For this analysis we utilize *SAMBA* (Statistical-Algorithmic Method for Bicluster Analysis), an algorithm that was developed in Shamir's group by Dr. Amos Tanay and Dr. Roded Sharan [116].

Functional enrichment analysis

After identifying the main co-expressed gene groups in the data (either by clustering or biclustering), one of the major challenges is to ascribe to them a biological significance. To this end, we applied statistical analysis that seeks specific functional categories that are significantly over-represented in the analyzed gene groups with respect to a given background set of genes. This analysis utilizes functional annotation files that associate genes with functional categories using the standard vocabulary defined by the Gene Ontology (GO) consortium (<http://www.geneontology.org/>) [117]. The statistical significance of the enrichment of a specific cluster for genes from a particular functional category is determined by computing the upper tail of the hypergeometric distribution (see, for example, [111]), taking into account the number of

genes represented on the chip that are associated with this functional category. While certain genes are represented by several probe sets, to avoid biases, each gene is counted only once.

An improved module for functional enrichment analysis was recently integrated in *EXPANDER*. The *TANGO* (Tool for ANalysis of GO enrichments) algorithm developed by Dr. Amos Tanay accounts better for extensive multiple testing typically done in such analysis (hundreds of categories are typically tested for enrichment). While standard methods for accounting for multiple testing assume independent tests (e.g., Bonferroni, False Discovery Rate), the hierarchical tree-like structure of the GO ontology induces strong dependencies among the categories. *TANGO* accounts for these dependencies, thus yielding more reliable p-value estimations.

Cis-regulatory element analysis

Microarray measurements provide snapshots of cellular transcriptional programs that take place in the examined biological conditions. These measurements do not, however, directly reveal the regulatory networks that underlie the observed transcriptional activity, i.e., the transcription factors (TFs) that control the expression of the responding genes. Computational promoter analysis can shed light on the regulators layer of the network. For this task, the *PRIMA* tool was developed in the Shamir and Shiloh labs [118]. In short, given target sets and a background set of promoters, *PRIMA* performs statistical tests aimed at identifying TFs whose binding site signatures are significantly more prevalent in any of the target sets than in the background set. Typically, sets of co-expressed genes identified using either cluster or bicluster analysis serve as target sets, and the entire collection of promoters of genes present on the microarray serves as the background set. We applied *PRIMA* to gene expression datasets to discover TFs that control the observed alteration in the cellular transcriptome.

Superposition of gene expression data on signaling maps

Knowledge-bases for signaling pathways are becoming highly instrumental in the analysis of high-throughput data in general and gene expression in particular. A simple way to integrate gene expression data with extant biological knowledge is to present microarray results on signaling maps. This can be done, for example, by coloring genes in the maps according to their expression levels. Such coloring points to sub-regions in the network that are turned on or shut down in response to the examined perturbations. Such sub-regions will correspond to subgraphs densely populated with genes that are induced or repressed in the dataset. We use the *SPIKE* (Signaling Pathway Integrated Knowledge Engine, <http://www.cs.tau.ac.il/~spike>) tool for this task. *SPIKE*, which is developed in the Shamir and Shiloh labs, is a comprehensive, up-to-date knowledge-base of cancer-related signaling pathways. *SPIKE* includes three main software components: (1) A database (DB) for biological interactions with an interface that supports data uploading and querying by multiple end-users. Large public data sources and carefully curated pathways constitute distinct tiers of the DB. (2) A visualization package that allows interactive graphic representations of the biological interactions stored in the DB, dynamic layout and navigation through the networks, and superposition of functional genomics data on the interaction maps. (3) An algorithmic inference engine that analyzes the data and suggests new links and interplays between selected components of the network.

Analysis overview

Applying this integrative approach to analysis of gene expression datasets allows us to systematically identify major expression patterns in the data (by applying cluster or bicluster analysis), assign clusters with putative functional roles (based on the enriched functional categories), and reveal transcription factors that underlie the transcriptional response exhibited by the clusters (using *PRIMA*). Superposition of the data on signaling maps using *SPIKE* helps us to identify active pathways and to generate hypothetical mechanistic models for cellular networks that respond to various stresses.

4 Results

The experimental part of this thesis includes two microarray experiments. In the first experiment, which served for calibration, BJ-hTert human cells were exposed to NCS. In the second experiment, we applied our experimental and computational strategy to delineate transcriptional responses to ionizing radiation (IR) in five human cell lines, which represent a variety of tissues: BJ-hTert (fibroblasts), G361 (melanoma), HepG2 (hepatocellular carcinoma), TK6 (lymphoblast) and U2OS (osteosarcoma).

The experiments were carried out in collaboration with Dr. Sharon Rashi-Elkeles and Dr. Ran Elkon of Shiloh group; and Dr. Ninette Amariglio and Prof. Gideon Rechavi of the Functional Genomics Unit at the Sheba Medical Center.

4.1 Calibration of the experimental system, using human fibroblasts exposed to NCS

In this study we applied gene expression microarrays to BJ-hTert foreskin fibroblast human cells exposed to NCS. This study served as a calibration experiment for the next study we have done (see next section). We used the data collected here to compare biological replicates prepared by two researchers; to examine the effect of an experimental manipulation on gene expression; and to choose informative time points for the next, more comprehensive, experiment.

Global transcriptional responses were recorded in BJ-hTert untreated cells, and at 2, 4, 8 and 12 hours after exposure to 200 ng/mL NCS. Untreated cells were harvested at the beginning of the experiment, and 8 hours later. Two out of the six conditions had duplicates: 0 and 4 time points. At the beginning of the experiment, all nine cell culture dishes were taken out of the incubator to the hood in the adjacent room, where NCS was added to seven of the dishes. Then, all cell cultures were returned to the incubator, except for the two dishes of untreated control of 0 hr. Affymetrix GeneChip HGU133A 2.0 arrays (18,400 transcripts and variants that represent

approximately 14,500 well-characterized human genes) were used. RNA samples and chip hybridizations for all time points were prepared by Dr. Sharon Rashi-Elkeles, and samples and chip hybridizations of the two replicates (0 and 4 hr) were independently prepared by this author.

Signal intensities were computed using the RMA method [91] that was run from the BioConductor package (<http://www.bioconductor.org/>), and which applies the quantile normalization. The analysis reported here was also done using Affymetrix MAS 5.0 software, and the results reported below remained the same.

4.1.1 Intra-laboratory comparison between biological replicates

A key step in bringing gene expression technology into clinical practice is performance of large studies to confirm preliminary models. The execution of such confirmatory studies and the transition to clinical practice requires that microarray data from different laboratories are comparable and reproducible [119]. Most important, we expect intra-laboratory microarray data to be comparable and reproducible. Yauk et al. (2004) reported a correlation of around 0.9 between biological replicates for Affymetrix arrays [120]. Our first goal was therefore to assess the reproducibility of the biological replicates, produced independently by two researchers in our lab (Dr. Rashi-Elkeles and this author), following identical handling of the samples based on a single protocol.

We produced replicates for two conditions: untreated cells (harvested in the beginning of the experiment) and treated cells, harvested 4 hr after exposure to NCS. For each condition, one replicate was prepared by Dr. Rashi-Elkeles, and the other by this author. In all, we had four samples for the reproducibility testing (two conditions, each with two replicates). The two samples prepared by Dr. Rashi-Elkeles (investigator A) were designated “set A”, and the two prepared by this author (investigator B) were designated “set B”.

Affymetrix assigns a Detection call to each probeset, indicating whether a transcript is reliably detected ('Present'), not reliably detected ('Marginal') or not detected ('Absent') (see

section 3.2.7). Probesets that registered no 'Present' calls in any of the four chips were excluded. Some 60% of the probesets in the array, 14,179 out of 22,277 probesets, passed this filter.

Intra-laboratory comparison between biological replicates

In order to examine the correlation between sets A and B, we computed the Pearson correlation coefficient between the expression values 4 hr after NCS treatment, revealing a very strong correlation of 0.996.

Next, we examined the agreement between fold changes (4 hr after treatment compared to 0 hr) in both replicates, with a fold-change threshold. We defined three sets of 'responding probesets', consisting of probesets whose expression level was changed by at least 1.5, 1.75 or 2 fold across the two sets of tested conditions (**Table 4.1.1-1**). Some 2% of the probesets present on the array, 237-344 out of 14,179, met the criterion of 1.5 fold change in expression in one of the sets. About half of the responding probesets in set A were found responding in set B as well. This overlap is clearly not random. We used the hyper-geometric distribution to compute the p-value corresponding to these numbers. The p-value of having 160 probesets common to sets A and B out of 14,179 probesets in total is 6.63E-215. **Table 4.1.1-1** shows the p-values for the data corresponding to 1.5-fold or higher changes.

Fold change	1.5	1.75	2
Set A	344	125	63
Set B	237	104	57
Intersection	160	71	37
p-value	6.63E-215	2.04E-130	1.47E-78

Table 4.1.1-1. Number of probesets whose expression values changed by a specified fold-level 4 hr after NCS exposure compared to untreated control, with the corresponding p-values. This criterion was applied separately to each set of samples (A, B), and the number of probesets that satisfied the criterion in both sets was calculated (Intersection).

We chose the 1.5 fold filter for further analysis, because it provided the largest number of probes to work with. The Pearson correlation coefficient between the fold-changes of the replicates was $r=0.97$, which indicated a very strong correlation. Plotting the log fold-change of expression levels of the intersection between the two sets, reassured the strong correlation (see **Fig. 4.1.1-1**).

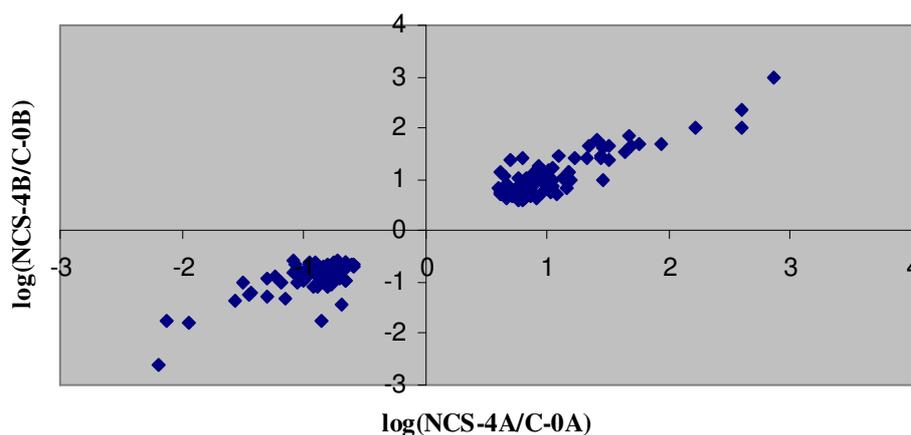


Figure 4.1.1-1: Comparison between the fold change in expression prior to and 4 hr after NCS treatment in set A and set B. Plotted here are the log of fold-changes of the 160 probes that showed at least 1.5 fold change in both sets (the intersection group) between treated samples harvested 4 hr after NCS exposure (NCS-4) and the control (C-0).

Examination of the added value of biological replicates of two investigators

To confirm that the use of replicates improves the reliability and sensitivity of the analysis, we compared data derived from two independent replicates of two investigators with data derived from one replicate, by conducting further analysis on the sets obtained using the 1.5 fold filter. We built three groups, each with 70 probesets: A-only, B-only and AB, representing set A, set B and the intersection set, respectively. Each group consisted of 35 up-regulated and 35 down-regulated probesets. Specifically, group A-only consisted of the 70 probesets that responded by at least 1.5 fold in set A and were least responding in set B. Group B-only was built in the same fashion, and group AB consisted of the most responsive probesets. The groups consisted of the exact same number of probesets, to avoid bias in the analysis. Further explanations follow:

To create group A-only, the 344 probe-sets of set A that passed the 1.5 fold filter were divided into up-regulated and down-regulated probesets. The up-regulated probes were sorted by their fold change in set B, and the 35 probes that were least up-regulated in set B were selected. The down-regulated probes in set A were sorted by their fold-change in set B and the 35 probes with least down-regulation were selected. The union of the two sets formed group A-only. Group B-only was constructed in the same manner.

Group AB was built in two steps: first, for each probeset contained in both sets A and B, $\max(\text{fold change A, fold change B})$ or $\min(\text{fold change in set A, fold change in set B})$ were calculated for up- or down-regulated genes, respectively. The 35 probesets with largest computed values and the 35 probesets with smallest computed values were combined to make the group of the most responsive probesets, group AB.

Since each group was built of up- and down-regulated probes, we actually had six sets of probesets. These sets were loaded to EXPANDER for further analysis.

Functional categories within gene sets. To examine the functional categories of each group, we applied tests aimed at identifying functional categories that are statistically enriched in the groups.

Table 4.1.1-2 depicts the functional categories enriched in the up/down-regulated genes with a very strict p-value ≤ 0.001 , after correction for multiple testing. The corrected p-value could not be smaller than 0.001, since it is based on 1000 random sampling (performed by the algorithm). No categories were enriched for in group A-only. One category was enriched for in group B-only (regulation of progression through cell cycle, with $p=0.008$), but did not pass the p-value threshold of 0.001. Examination of the categories enriched in genes that responded to NCS in both sets (group AB), indicated that the network activated following NCS concerns mainly arrest of the cell cycle. These results clearly demonstrate that combining data from replicates decreases the noise of the measurements and yields more informative and statistically significant (better p-values) results.

Regulation	Functional category, GO ID	Group AB Raw p-val
Up	regulation of progression through cell cycle, GO:0000074	1.75E-7
	negative regulation of cell proliferation, GO:0008285	7.63E-9
Down	microtubule-based process, GO:0007017	3.60E-7
	mitosis, GO:0007067	8.95E-16

Table 4.1.1-2: Functional categories enriched for in up/down-regulated probesets of group AB (raw p-val presented, corrected p-val ≤ 0.001 in all cases).

Computational search for mediating transcriptional regulators. Next, we compared the three groups by identifying the transcriptional regulators that control the observed modulation in gene expression 4 hr after exposure to NCS in each group. To this end, we applied PRIMA. Each group was considered a target set, and the collection of genes present on the chips and expressed in the cells was the background set. Human promoter sequences used here were downloaded from Ensembl project (v27, December 2004 release). Analysis was done on the region from 1,000 bp upstream to 200 bp downstream to genes' putative transcription start sites (TSS). In addition, a long distance scan, from 2,000 bp upstream to 1000 bp downstream, was done, because of the relatively long distance of p53's binding sites from the TSS.

PRIMA identified two well known DNA damage response (DDR) TFs whose binding site signatures were significantly over-represented in group AB, containing the strongest responding genes according to both replicates (**Table 4.1.1-3**). p53, whose binding site signature was found in the up-regulated genes, mediates the activation of G1/S checkpoint, DNA repair and apoptosis [121]. NF-Y, whose binding site signature was enriched for in the down-regulated genes, was demonstrated to control the expression of several key regulators of the G2/M phases of the cell cycle [122-125]. p53 and NF-Y were not detected in A-only and B-only groups. No enrichments passing the p-value threshold of 5×10^{-4} were found in group B-only. Nkx6-2 binding site was enriched for in the promoters of genes in group A-only. This transcription factor has a role in neurogenesis, and is required for alpha- and beta-cell development in the pancreas [126-128].

We conclude that there is a very good correlation between the intra-laboratory replicates, and that the use of data from both replicates yields more reliable and more significant results.

Transcription factor [TRANSFAC Accession number]	Regulation	Group A-only		Group B-only		Group AB	
		EF	p-value	EF	p-value	EF	p-value
p53 [M00034]	Up	-	-	-	-	3.47	1.31E-4
NF-Y [M00287]	Down	-	-	-	-	2.67	4.27E-5
Nkx6-2 [M00489]	Up	3.8	7.3E-5	-	-	-	-

Table 4.1.1-3. TF binding site signatures enriched for in up- and down-regulated genes in groups A-only, B-only and AB (p-val<0.0005). EF denotes the enrichment factor - The ratio between the prevalence of transcription factor hits found by PRIMA in promoters of the genes contained in the cluster and in the background set of all the genes defined as present on the arrays. Analysis was done on the region from 1,000 bp upstream to 200 bp downstream to genes' putative TSS, except for p53, with scan from -2,000 to 1000.

4.1.2 Testing the side-effects of experimental procedure on gene expression

Analysis of microarray data is difficult because of the variability inherent to these data. This variability results from a large number of disparate factors operating at different times and levels during the course of a typical experiment. These numerous factors can be broken down into two major categories, biological variability and experimental (technical) variability [129].

It is important to consider the side-effects of the experimental procedure on gene expression and mRNA stability. For example, it is often reported that the cells to be analyzed are harvested by centrifugation and frozen for RNA extraction at a later time. If the cells encounter a temperature and pH shift during the centrifugation step, even for a short time, this could cause a change in gene expression profiles. Each of these and other experimentally caused gene expression changes will confound the interpretation of the experiment. Obviously, these are not easy variables to control [129].

In this study, we attempted to check whether the manipulation of culture dishes during the experimental procedure contributes to the experimental noise. This procedure was as follows: culture dishes were taken out of the incubator and brought into the hood in an adjacent room, where NCS was added to all dishes except the controls. The treated cells were then put in the incubator for various time periods, while the time 0 control was immediately harvested.

We were concerned that the expression profile of the cells might be affected not only by the NCS treatment, but also by the decrease in temperature and in CO₂ levels caused by this procedure. Therefore, in addition to the control harvested at the beginning of the experiment, we added as another control untreated cells that were harvested 8 hr after the treatment procedure. This latter control was termed 'chronological control', and the control harvested at the beginning of the experiment, was termed 'absolute control'.

Since the samples harvested 8 hr after NCS treatment were processed by Dr. Rashi-Elkeles, we used in this section only expression data obtained from the seven samples prepared by her, to reduce variability and focus on the factor we were testing.

Probesets that registered 'Present' flags by Affymetrix MAS 5.0 software in less than 2 conditions were excluded. Some 60% of the probesets in the array, 13,645 out of 22,277, passed this filter, yielding a 13,645x7 data matrix, with rows corresponding to probes. To examine the similarity between the gene expression profiles of the conditions, we applied average-linkage hierarchical clustering to the dataset (**Fig. 4.1.2-1**). Hierarchical clustering does not partition the genes into subsets. Instead it creates a hierarchy of the elements that can be represented by a dendrogram. In Expander, this is done using the 'agglomerative' method [112], which starts with an initial partition into single element clusters and successively merges clusters until all elements belong to the same 'cluster'. The dendrogram was calculated using the average-linkage scheme.

Surprisingly, we noticed that gene expression profiles in the "chronological control", C-8 (not treated with NCS), are similar to those of the treated samples harvested 8 and 12 hr after

exposure to NCS (NCS-8 and NCS-12, respectively) more than to the absolute control, C-0 (that is, closer nodes in the tree).

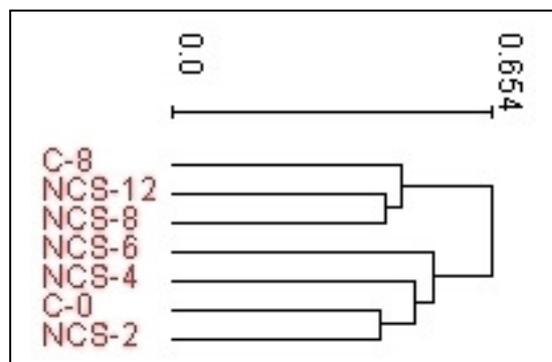


Figure 4.1.2-1. Examination of similarity between the probed conditions. To examine similarity, we applied hierarchical clustering of the probed conditions: two controls (C-0, C-8) and samples at 2, 4, 6, 8 and 12 hr after exposure to NCS. 13,645 probesets were reliably detected in the experiment, yielding a 13,645x7 data matrix, with the rows corresponding to genes. The matrix columns were subjected to hierarchical clustering after normalizing the rows to have mean=0 and SD=1. 'C' stands for control and 'NCS' for NCS exposure. The numbers represent the time points. The scale next to the tree indicates the relative distance between the branches of the tree, which correspond to the conditions. Note that C-8 is grouped with NCS-8 and NCS-12 rather than C-0.

A closer look confirmed the resemblance seen in the hierarchical clustering. We counted the probesets whose expression level changed by at least 1.5 fold between the absolute, the chronological control and 8 hr from after exposure to NCS: (1) C-8 vs. C-0, (2) NCS-8 vs. C-0, (3) NCS-8 vs. C-8. We found that 349 probesets were found up-regulated 8 hr post NCS from the absolute control, while 281 probesets were induced 8 hr from the beginning of the experiment, without exposure to NCS (**Table 4.1.2-1**). We counted the number of probesets in the intersection groups, and calculated the p-value for each overlap, using a hypergeometric distribution. 108 probesets were up-regulated 8 hr from the beginning of the experiment, both with and without exposure to NCS, with a p-value of 6.2E-102. This high concordance is not random. Only 2 induced probesets were common to groups 1 (C-8 vs. C-0) and 3 (NCS-8 vs. C-8), with insignificant p-value. It seems that the 'noisy' probesets that are induced due to the non-NCS experimental manipulation, are different from the NCS-responding probesets. The same principle persists in the down-regulated genes, as observed in **Table 4.1.2-1**.

	FC =1.5	
	Up	Down
(1) C-8 vs. C-0	281	122
(2) NCS-8 vs. C-0	349	351
(3) NCS-8 vs. C-8	127	212
(1) AND (2)	108 (6.2E-102)	70 (1.3E-80)
(1) AND (3)	2 (0.25)	17 (5.7E-12)
(2) AND (3)	83 (2.7E-103)	130 (1.8E-159)

Table 4.1.2-1. Comparison of responding probesets 8 hr from the beginning of the experiment. Number of probesets whose expression level was up- or down-regulated by at least 1.5-fold in the following comparisons: (1) absolute control and chronological 8 hr control; (2) absolute control and 8 hr after exposure to NCS; (3) chronological control and 8 hr after exposure to NCS. The next lines give the number of common probesets responding in each possible pair of groups 1, 2 and 3, along with the corresponding p-value in parentheses.

Functional categories within gene sets. To examine the functional categories of each group, we applied tests aimed at identifying functional categories that are statistically enriched in the groups. The down-regulated genes encode cell-cycle proteins in all three groups, but with better p-values for groups 2 and 3. The up-regulated genes were not associated with a specific function in groups 1 and 2. Group 3 was enriched with the following functions: negative regulation of cellular proliferation (corrected p-value 0.017) and regulation of cell cycle (corrected p-value 0.023).

Computational search for mediating transcriptional regulators. We compared the three groups by identifying the transcriptional regulators that control the observed modulation in gene expression (**Table 4.1.2-2**). Down-regulation: NF-Y was enriched in the three groups, but the enrichment was more significant in groups 2 and 3. E2F was enriched in groups 2 and 3 only, and E2F-1 was enriched solely in group 3. Up-regulation: p53 was enriched only in group 3. IPF1 (insulin promoter factor 1), which is not related to the DNA damage response, was detected in groups 1 (borderline significance) and 2.

Regulation	TF [Accession in TRANSFAC DB]	(1) C-8 vs. C-0	(2) NCS-8 vs. C-0	(3) NCS-8 vs. C-8
Down	NF-Y [M00287]	2.1E-5	4.67E-15	4.88E-10
	E2F [M00918]	-	4.55E-4	1.76E-6
	E2F-1 [M00940]	-	-	1.14E-4
	FOXD3 [M00130]	-	-	4.83E-5
Up	p53 [M00034]	-	-	7.2E-7
	IPF1 [M00436]	(5.52E-4)	2.41E-4	-

Table 4.1.2-2. TF binding site signatures enriched in up- and down-regulated genes in (1) absolute control and chronological 8 hr control; (2) absolute control and 8 hr after exposure to NCS; (3) chronological control and 8 hr after exposure to NCS. P-val<5E-4. The minus sign is used when the TF binding site signature is not enriched. Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene, except for p53, for which the area from -2,000 to 1,000 bp was scanned.

We concluded that the manipulation of culture dishes during the experimental procedure by itself was sufficient to cause changes in gene expression profiles of the cultures. We took this finding into consideration when designing the subsequent experiment: we had chronological controls for each time point, in addition to the absolute control.

4.1.3 Selecting informative time points for the following experiment

We now examined the time points considering two aspects: First, we looked at the kinetics of the response to NCS as a function of time. Then, we focused on each time point separately, to identify time points at which we may get the most definitive regulatory information.

The initial filtering of the data based on detection calls (indicating the reliability of transcript detection) was done as described in the previous section, yielding 13,645 probesets that passed the filter, out of 22,277.

Secondly, we defined the set of 'responding genes', consisting of genes whose expression level was changed by at least 1.75-fold compared to the absolute control at any of the time points. This threshold was chosen after examination of a few thresholds, since it gave the most

significant results (better p-values for functional and promoter analysis). 553 probesets fit that criterion.

We subjected the set of responding genes to CLICK, a clustering algorithm that seeks a balance of intra-cluster homogeneity and inter-cluster separation [113]. Prior to clustering, expression levels of each gene were standardized to have mean equal to zero and variance equal to one; hence, genes clustered together share expression patterns across the tested conditions, but might differ in the magnitude of their response.

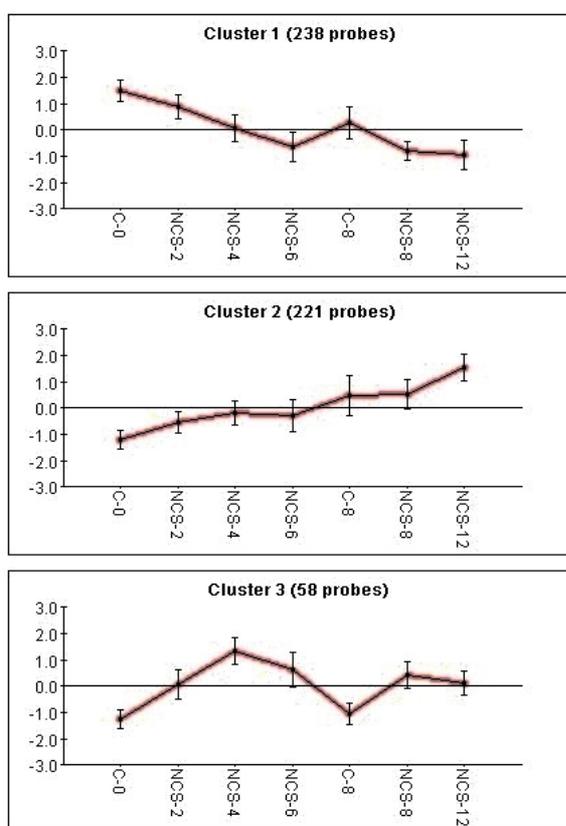


Figure 4.1.3-1: Major clusters identified by CLICK in the set of responding genes to NCS in the following time points: 0, 2, 4, 6, 8 and 12 hr after exposure to NCS, and 8 hr without exposure (C-8). Each cluster represents a set of genes with a similar expression pattern. Prior to clustering the expression levels of each gene were standardized to have a mean value of 0 and variance of 1. The Y axis represents these standardized values. The X axis corresponds to the tested conditions: untreated samples at 0 and 8 hr time points (C-0 and C-8) and 2, 4, 8 and 12 hr after exposure to NCS (NCS-2, 4, 6, 8 and 12). Shown for each cluster is the mean expression pattern calculated over all the genes contained in it. Error bars represent +/- one S.D. The total number of genes in each cluster is indicated.

Three clusters were identified by CLICK (**Fig. 4.1.3-1**). The largest cluster represented down-regulated genes. The strongest down-regulation was apparent 6 hr after exposure to NCS,

and remained stable 8 and 12 hr after treatment. Note that the genes were down-regulated 8 hr after the beginning of the experiment even in the non-treated sample (C-8), pointing to additional factors, not related to DNA damage, involved in the observed down-regulation.

The second and third clusters represented up-regulated genes. At cluster 2, a first wave of expression induction was apparent at 4 and 6 hr, and a second wave was noticed 12 hr after treatment. The two samples harvested 8 hr after the beginning of damage induction depicted the same level of expression, even though one of them was exposed to NCS and the other was not. The third cluster showed a strong up-regulation 4 hr after NCS, and the expression levels seemed to be affected purely by DSB induction, since the control at 8 hr showed the same level of expression as the absolute control.

We conclude that the kinetics shown by the clusters suggests that time-points 4 and 6 hr capture best the up- and down-regulation, respectively.

Next, we examined each time point separately, in relation to the absolute control, in order to see which of the time points might provide more information on the transcriptional responses to DSBs. For each time point we defined the set of 'responding genes', consisting of genes whose expression level was changed by at least 1.75-fold at a specific time point compared to the absolute control. The responding genes were divided into two sets, representing down-regulated and up-regulated genes (**Table 4.1.3-1**). Generally, the number of responding genes increased with time in both directions (up- and down-regulation). The exception was the untreated sample harvested at the 8 hr time-point (C-8), in which the number of down-regulated genes was smaller by several-fold compared to cells 8 hr after NCS exposure.

Regulation	NCS-2	NCS-4	NCS-6	C-8	NCS-8	NCS-12
Down	9	51	130	34	135	177
Up	30	66	61	63	87	199

Table 4.1.3-1: Number of probesets whose expression values changed by at least 1.75-fold 2, 4, 6, 8 and 12 hr after exposure to NCS, and at 8 hr without NCS exposure (C-8).

Using PRIMA in the EXPANDER package, we identified TFs whose binding site signatures were significantly over-represented in the responding gene sets. In **Table 4.1.3-2** we compare the p-values given by PRIMA for the detection of well known DDR TFs: E2F and NF-Y for down-regulated genes, and p53 for up-regulated genes.

Regulation	TF [Accession Num. in TRANSFAC DB]	NCS-2	NCS-4	NCS-6	C-8	NCS-8	NCS-12
Down	E2F [M00918]	---	---	2.0E-06	---	2.2E-05	1.3E-05
	NF-Y [M00287]	---	1.7E-08	1.9E-16	5.2E-04	1.3E-16	7.1E-15
Up	p53 [M00034]	---	5.1E-5	8.7E-07	---	---	---

Table 4.1.3-2. TF binding site signatures enriched in up- and down-regulated genes in each time point: 2, 4, 6, 8 and 12 hr after exposure to NCS, and 8 hr without exposure (C-8). The minus sign is used when the TF binding site signature is not enriched in the corresponding time point. Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene, except for p53, for which the area from -2,000 to 1,000 bp was scanned.

Inspection of **Table 4.1.3-2** revealed that 6 hr after exposure to NCS the examined TFs were detected with the best p-values. Two hr after exposure, none of these TFs were enriched, possibly due to the low number of responding genes at this time point.

Taking into account both the clustering kinetics and the regulatory analysis of each time point, we realized that the time point of 6 hr was the most informative.

It is well known that the DNA damage response network is activated instantaneously after DSB induction which triggers the rapid activation of ATM kinase [130, 131]. In order to cover this early wave of response, we added another time point in the next experiment - 3 hr after exposure to the DNA damaging agent.

4.2 Delineation of transcriptional responses induced by IR in five human cell lines

In this study we applied gene expression microarrays combined with a computational battery to compare the transcriptional responses induced by IR in five human cell lines: BJ-hTert (foreskin fibroblasts), G361 (melanoma), HepG2 (hepatocellular carcinoma), TK6 (lymphoblastoids) and U2OS (osteosarcoma).

Typically, the response to radiation varies between cell lines that are commonly used for research. We were interested in delineating the variability among such cell lines, which is often disregarded by investigators.

Global transcriptional responses were recorded in BJ-hTert, G361, HepG2, TK6 and U2OS untreated cells, and 3 and 6 hr after IR dose of 5 Gy. Based on the observation of side-effects of the experimental procedure, we decided to add a chronological control to each time point: untreated cells were harvested at the beginning of the experiment, and 3 and 6 hr later. Each condition (treated or untreated) was measured in triplicates, two of which were prepared by Dr. Rashi-Elkeles, and the third one - by this author, independently. All together, we had five cell lines under five different conditions (three controls and two irradiated samples), with three biological repeats. In total, we prepared 75 samples.

In each experiment, all culture dishes were brought to the irradiation instrument, the relevant dishes were irradiated, and all cultures were returned to the incubator except for the absolute control.

Affymetrix GeneChip HGU133 Plus 2.0 arrays were used (over 47,000 transcripts and variants, including 38,500 well-characterized human genes). Signal intensities were computed using the RMA method [91] operated from the BioConductor package (<http://www.bioconductor.org/>), which applies the quantile normalization.

Probesets that received 'Absent' or 'Marginal' calls in at least two thirds of the chips were filtered out. This filtering was done for each cell line separately (removing probesets that received 'Absent' or 'Marginal' calls more than 10 out of 15 chips) and for all cell lines together (removing probesets that received 'Absent' or 'Marginal' calls in more than 50 out of 75 chips), yielding approximately the same number of probesets (about 45% of the probesets in the array) for subsequent analysis (**Table 4.2-1**).

Cell line	Num. of present probesets
BJ-hTert	25,283
G361	24,429
HepG2	24,541
TK6	23,400
U2OS	24,937
All 5 cell lines	25,470

Table 4.2-1. Number of probesets expressed in each cell line and in the five cell lines together, which were left after filtering out 'Absent' and 'Marginal' ones.

One-way ANOVA was performed on gene expression measurements in order to filter out probesets for which the variance of values between conditions was smaller than the variance within replicates. We identified genes that had significantly different expression levels ($p < \text{threshold}$, specified below) at each of the 5 conditions in every cell line. A representative expression level for each probeset in each of the tested conditions was then computed by averaging the probeset's signal intensities in the replicate arrays. The ANOVA statistical test was applied in two different manners.

4.2.1 Hierarchical clustering of the cell lines

One-way ANOVA was applied to the dataset containing all 75 measurements. When performing this filter with $p < 0.001$, 23,296 probesets (out of 25,470 present in the arrays) remained for further analysis. Many probesets passed this filter due to the great variance among the different cell lines. Averaging expression levels over replicates, our dataset contained measurements for twenty-five conditions.

We divided the expression level of each of the probesets at the 3 and 6 hr time point by its expression level in the absolute control of the same cell line, yielding a 23,296x20 data matrix, with rows corresponding to probes. We standardized each row to mean=0 and standard deviation (SD) = 1, and subjected the columns of the standardized matrix to average-linkage hierarchical clustering.

We found that the most prominent factor that influenced the expression profile was the identity of the cell line (**Fig. 4.2-1**). In the dendrogram, the conditions were first divided into five groups according to cell lines. Looking closely at the tree of hierarchy, we noted that the root diverged into two branches: TK6, and all the other cell lines. The branch of the four cell lines then diverged into two branches - G361 and BJ-hTert in one branch, and HepG2 and U2OS in the other. The partition inside the TK6 branch was in accordance with the IR treatment: the 3 and 6 hr time points were adjacent, and so were the control 3 and 6 hr samples. In the rest of the cell lines the partition is not strictly according to IR response, implying that further manipulation of the data was required to see the actual response to IR.

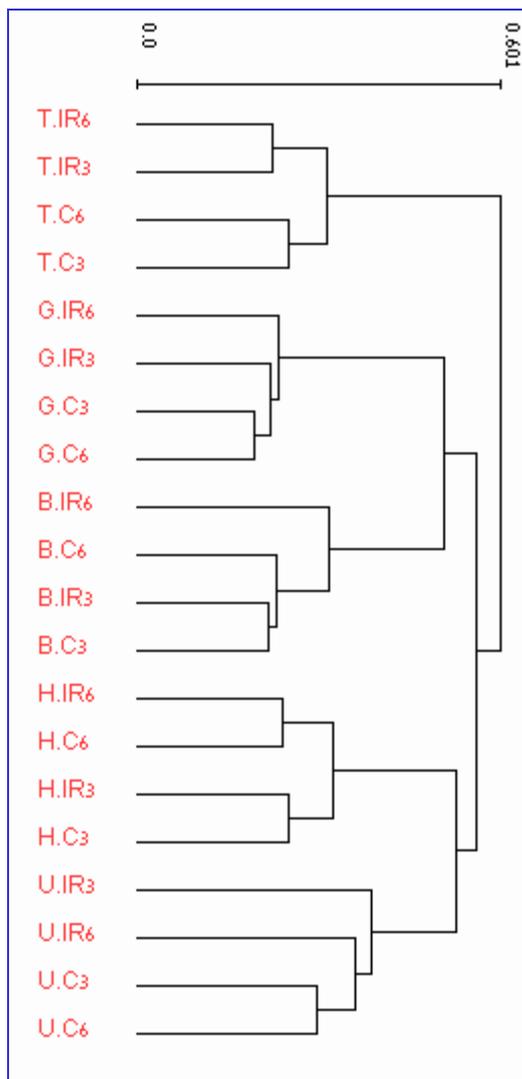


Figure 4.2.1-1. Hierarchical clustering of untreated and treated samples of five cell lines. For each of the 23,296 probesets that passed the ANOVA test, the expression level at 3 and 6 hr was divided by the level of the absolute control of the cell line, yielding a 23,296x20 data matrix, with the rows corresponding to genes. The matrix columns were subjected to hierarchical clustering after normalizing the rows to have mean=0 and SD=1. The conditions' names contain the initial of the corresponding cell line. 'C' stands for control and 'IR' for an irradiated sample. The numbers 3 and 6 represent the time points. The scale next to the tree indicates the relative distance between the branches of the tree, which correspond to the conditions.

4.2.2 Analysis of differentially expressed genes in specific cell lines

One-way ANOVA was applied separately to each cell line over 15 measurements (five conditions, in triplicates): three untreated time points (0, 3 and 6 hr) and two irradiated time points (harvested 3 and 6 hr after exposure to IR). **Table 4.2.2-1** depicts the number of probesets that had significantly different ($p < 0.05$ and $p < 0.01$) expression levels over the five conditions. There are differences between the numbers of probesets that passed the filter in each cell line: TK6 had the largest number of surviving probesets, while G361 had the fewest.

The strict p-value of 0.01 was chosen for further analysis for all cell lines except for G361, for which p-value of 0.05 was used, since 46 probesets were too few for the subsequent analysis.

cell line	p<0.05	p<0.01
BJ-hTert	1,484	456
G361	288	46
HepG2	2,732	994
TK6	6,676	3,447
U2OS	1,270	447

Table 4.2.2-1. Number of probesets that passed one-way ANOVA test, no fold change criterion applied (See Appendix A for a table with the lists of probesets).

We then applied a fold-change threshold criterion to exclude probesets for which the change in expression between the tested conditions and the absolute control was not sufficiently large. We defined the responding gene set as "all genes that passed the ANOVA statistical test and whose expression levels changed (at either 3 or 6 hr) by at least 1.3-fold from the absolute control (time 0)". The number of responding genes for each cell line is presented in **Table 4.2.2-2**.

Due to noise and variability of array data, false-positive and false-negative results must be expected [129]. In order to estimate the false-positives, we calculated the number of probesets randomly expected to pass the p-value threshold applied in the one-way ANOVA test, out of the number of probesets that changed by at least 1.3-fold from the absolute control. For p-

value<0.01 we would expect 1% of the probesets to randomly pass the test. For example, 40 probesets out of 4,027 probesets were expected to randomly pass the test in BJ-hTert cells.

Cell line	1.3 FC	False Positives expected*	Responding genes: ANOVA + 1.3 FC	Expected % false-positives in responding genes set
BJ-hTert	4,027	40	402	10%
G361*	2,715	136	234	58%
HepG2	3,896	39	736	5%
TK6	9,716	97	3,126	3%
U2OS	6,647	66	424	16%

Table 4.2.2-2: Numbers of probesets that showed at least 1.3-fold change from absolute zero control, compared to the number of those that passed both one-way ANOVA and 1.3-fold change, with calculation of expected false-positives.

*Computation of false positives and the ANOVA test were performed with $p < 0.01$ for all cell lines, except for G361, where $p < 0.05$ was used.

We applied the same analysis to each cell line in order to make the results comparable. The analysis flow will be presented in details for BJ-hTert cells, and the results for the other four cell lines will follow.

Analysis of BJ-hTert IR response

Identification of major expression patterns in the dataset. We subjected the set of responding genes to the clustering algorithm CLICK [113]. Prior to clustering, expression levels of each gene were standardized to have mean equal to zero and variance equal to one; hence, genes clustered together share expression patterns across the tested conditions, but might differ in the magnitude of their response.

Five clusters were identified by CLICK (**Fig. 4.2.2-1**). The two major clusters represent two main themes: down-regulation in response to IR (cluster 1, 162 probesets) and up-regulation (cluster 2, 119 probesets). These two clusters were the largest of the five clusters that were identified and were found to be enriched for several functional categories and transcriptional regulators.

Cluster 1 represents not only down-regulation in response to IR, but also down-regulation in response to the generic stress caused by the experimental manipulation described in the previous

section. We can see that not only the irradiated samples but also the unirradiated ones show down-regulation of gene expression.

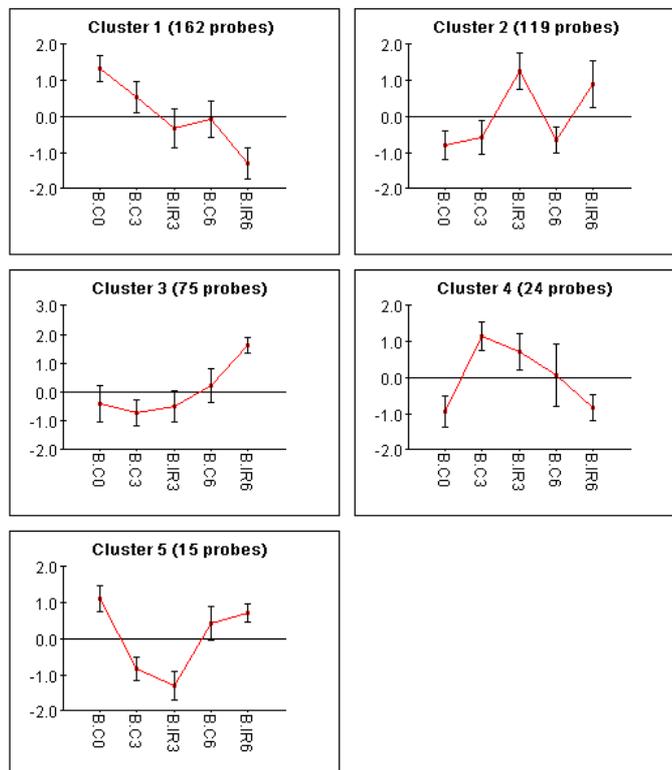


Figure 4.2.2-1. Clusters identified by CLICK in the set of responding genes of BJ-hTert cells. Each cluster represents a set of genes with a similar expression pattern. Prior to clustering, the expression levels of each gene were standardized to have a mean value of 0 and variance of 1. The Y axis represents these standardized values. The X axis corresponds to the tested conditions: unirradiated BJ-hTert samples at 0, 3 and 6 hr time points (B.C0, B.C3 and B.C6) and 3 and 6 hr after irradiation (B.IR3 and B.IR6). Shown for each cluster is the mean expression pattern calculated over all the genes contained in it. Error bars represent +/- one S.D. The total number of genes in each cluster is indicated.

Functional categories within gene clusters. Examination of the genes that responded to IR indicated that the network activated following IR spans many biological processes covering most aspects of the cellular physiology. In an attempt to systematically characterize this network, we applied tests aimed at identifying functional categories that are statistically enriched in the clusters. We utilized the standard vocabulary introduced by the Gene Ontology (GO) consortium [117]. Each gene cluster was considered a target set, and the genes present on the chips and expressed in BJ-hTert cells were the background set. Each enrichment was given a p-value before and after correction for multiple testing (raw and corrected, respectively). The corrected

p-value could not be smaller than 0.001, since 1000 random samplings were performed by the algorithm.

Enriched functional categories ($p < 0.01$ after multiple tests correction) were identified in cluster 1 (**Table 4.2.2-3**). Importantly, regulation of cell cycle was among the categories enriched in this down-regulation cluster, pointing to the prominent impact of IR damage on the cellular life cycle. Cellular responses to DNA damage have long been known to involve concerted activation of cell cycle checkpoints in order to prevent cell death during DNA replication, fixation of genetic alterations at the damage sites, or transmission of unbalanced genetic content to daughter cells [16].

Functional category, GO ID	# probesets associated with the category	Raw p (corrected p)
Spindle organization and biogenesis, GO:0007051	7	1.6E-9 (0.001)
Microtubule-based process, GO:0007017	13	9.2E-11 (0.001)
Cell cycle, GO:0007049	34	2.8E-16 (0.001)
Mitosis, GO:0007067	20	1.9E-17 (0.001)
Regulation of cellular process, GO:0050794	47	1.4E-6 (0.004)
Regulation of transferase activity, GO:0051338	10	2.9E-7 (0.003)
Cell division, GO:0051301	17	5.8E-13 (0.001)
ATP binding, GO:0005524	30	4.07E-8 (0.001)
second-messenger-mediated signaling, GO:0019932	8	3.1E-6 (0.007)
cell cycle checkpoint, GO:0000075	6	2.5E-6 (0.006)

Table 4.2.2-3. Functional categories enriched for in down-regulated genes in BJ-hTert cells (cluster 1), with corrected p -val < 0.01.

Computational search for mediating transcriptional regulators. Next, we sought to identify transcriptional regulators that control the observed modulation in gene expression following IR. To this end, we applied PRIMA. Each gene cluster was considered a target set, and the entire collection of putative human promoters corresponding to genes present on the chips and expressed in BJ-hTert cells was the background set. Human promoter sequences used here were

downloaded from Ensembl project (v27, December 2004 release). Analysis was done on the region from 1,000 bp upstream to 200 bp downstream to the putative transcription start sites (TSS) of the genes. In addition, a long distance scan, from 2,000 bp upstream to 1000 bp downstream, was done because of the relatively long distance of p53's binding sites from the TSS.

Cluster [Regulation]	Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor ¹	p-value
1 [Down]	E2F [M00918]	1.92	2.24E-4
	NF-Y [M00287]	2.21	2.15E-9
2 [Up]	TATA [M00252]	3.58	1.55E-4
	p53 ² [M00034]	3.22	5.34E-7

Table 4.2.2-4. TF binding site signatures enriched in the two major clusters identified by CLICK in BJ-hTert cells ($p < 0.0005$).

¹ Enrichment factor - the ratio between the prevalence of transcription factor hits found by PRIMA in promoters of the genes contained in the cluster and in the background set of promoters of all the genes defined as present on the arrays.

² Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene, except for p53, for which the area from -2,000 to 1,000 bp was scanned.

Identification of genes with maximal response. We detected the genes with the largest response to IR in both directions (up- and down-regulation). For each probeset, we calculated the fold-change from the respective chronological control and sorted the probesets according to max (IR3/C3, IR6/C6) to find the maximally up-regulated genes, and according to min (IR3/C3, IR6/C6) to detect the most down-regulated genes. In addition, we required that the fold-change between the treated samples and the absolute control would pass a 1.5 threshold.

The ten genes that were maximally responding to IR in both directions (up and down) are presented in **Table 4.2.2-5**. The five induced genes are members of cluster 2, and the five repressed genes are members of cluster 1, as expected.

Regulation	Gene Symbol	FC	Description of the protein product
Up	GDF15	4.70	A bone morphogenetic protein, member of the transforming growth factor-beta superfamily that regulates tissue differentiation and maintenance. Kis et al. detected GDF15 as one of the signaling proteins up-regulated following DNA damage [132]
	BTG2	4.54	A member of the BTG/Tob family, which consists of structurally related proteins that appear to have antiproliferative properties. This encoded protein is involved in the regulation of the G1/S transition of the cell cycle.
	CPM	4.31	A membrane-bound arginine/lysine carboxypeptidase. Its expression is associated with monocyte to macrophage differentiation.
	FAM84B	3.86	This protein is over-expressed in esophageal carcinomas [133].
	TNFRSF10A	3.27	A member of the TNF-receptor superfamily. This receptor is activated by tumor necrosis factor-related apoptosis inducing ligand (TNFSF10/TRAIL), and thus transduces cell death signal and induces cell apoptosis. Studies with FADD-deficient mice suggested that FADD, a death domain containing adaptor protein, is required for the apoptosis mediated by this protein.
Down	CCNE2	0.20	A member of the highly conserved cyclin family. Cyclins function as regulators of CDK kinases. This cyclin forms a complex with and functions as a regulatory subunit of CDK2. It has been shown to specifically interact with CIP/KIP family of CDK inhibitors, and plays a role in cell cycle G1/S transition. A significantly increased expression level of this gene was observed in tumor-derived cells.
	ZNF367	0.20	Contains a Cys2His2 zinc-finger motif. Gevaert et al. suggested it as one of the proteins affecting breast cancer [134].
	CENPA	0.28	A centromere protein that contains a histone H3 related histone fold domain that is required for targeting to the centromere. CENPA is proposed to be a component of a nucleosome-like structure, in which it replaces one or both copies of conventional histone H3 in the (H3-H4) ₂ tetrameric core of the nucleosome particle.
	DTL	0.30	A nuclear protein with centrosome targeting in mitosis, and plays important roles in DNA synthesis, cell cycle progression, cytokinesis, proliferation, and differentiation [135]. Banks et al. reported that DTL, PCNA and the DDB1-CUL4A complex play critical and differential roles in regulating the protein stability of p53 and MDM2/HDM2 in unstressed and stressed cells [136].
	SPBC25	0.30	This gene may be involved in kinetochore-microtubule interaction and spindle checkpoint activity.

Table 4.2.2-5: Maximally responding genes in BJ-hTert cells, induced or repressed. The fold-change (FC) was measured from the respective chronological control, at either 3 or 6 hr following IR. Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

Analysis of G361 IR response

G361 cells seem less responsive than the other four cell lines in our study. As shown in **Table 4.2.2-1**, only 46 probesets passed one-way ANOVA statistical test with $p < 0.01$. Neither functional categories, nor TF binding site signatures were found to be enriched in the clustering

solution of these 46 probesets. Therefore, one-way ANOVA test with $p < 0.05$ was used, followed by fold-change criterion of 1.3, for the creation of the responding genes set of G361 cells, consisting of 234 probesets.

Two clusters were detected, but only the down regulated set of genes was found enriched for functional categories and TF binding sites.

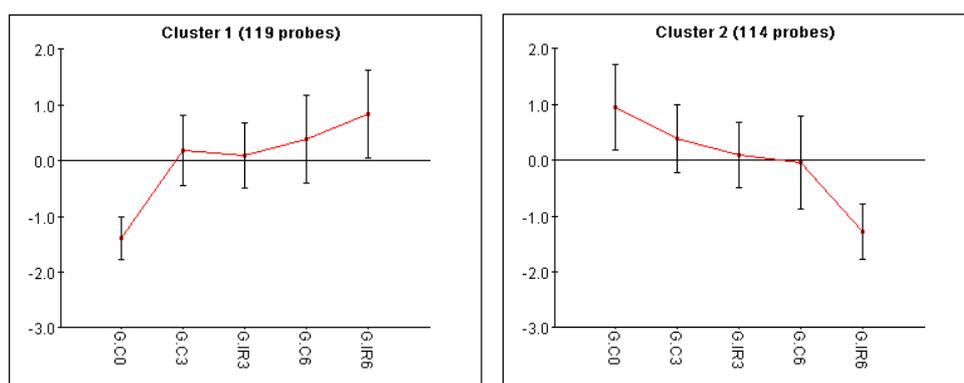


Figure 4.2.2-2. Clusters identified by CLICK in the set of 234 responding probesets of G361 cells. Clustering method and technical details are as in Figure 4.2.2-1.

Functional categories within gene clusters. Two functional categories related to cell cycle were detected in the down-regulated cluster, as observed in **Table 4.2.2-6**.

Computational search for mediating transcriptional regulators. NF-Y, which controls the expression of several key regulators of the G2/M phases of the cell cycle, was enriched in the cluster of down-regulated genes, as demonstrated in **Table 4.2.2-7**.

Functional category, GO id	Num. of genes associated with the category	Raw p (corrected p)
Cell division, GO:0051301	11	1.16E-8 (0.001)
Mitotic cell cycle, GO:0000278	12	5.48E-9 (0.001)

Table 4.2.2-6. Functional categories enriched in the down-regulated genes set (cluster 2) of G361 cells.

Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor	p-value
NF-Y [M00287]	1.78	8.64E-4

Table 4.2.2-7. TF binding site signatures enriched in down-regulated genes set (cluster 2) of G361 cells. Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene.

Identification of genes with maximal response. We detected the ten genes that were maximally responding to IR in both directions (five up and five down), as described above for BJ-hTert. The five induced genes were members of cluster 1, and the five repressed genes were members of cluster 2, as expected.

Regulation	Gene Symbol	FC	Description of the protein product
Up	LOC340109	2.80	Hypothetical protein, no domains were found in its sequence [137].
	TP53INP1	2.66	In response to double-strand DNA breaks, the encoded protein is induced in a p53-dependent way, induces cell cycle arrest in G1 and enhances p53 mediated apoptosis.
	LOC401131	2.21	This protein has a 2Fe-2S ferredoxins iron-sulfur binding region signature and a signal-peptide domain (which indicates the protein will be secreted) [137].
	PAPPA	2.20	A secreted metalloproteinase which cleaves insulin-like growth factor binding proteins (IGFBPs). It is thought to be involved in local proliferative processes such as wound healing and bone remodeling. Low plasma level of this protein has been suggested as a biochemical marker for pregnancies with aneuploid fetuses.
	LOC375010	2.09	This protein has a signal-peptide domain (which indicates the protein will be secreted) [137].
Down	PLK1	0.34	A protein kinase with a crucial role in cell division. Plk1 is an important regulator of several events during mitosis, and is also involved in activating the DNA damage check point in G2 phase [138].
	CCNB1	0.41	A regulatory protein involved in mitosis. The gene product complexes with p34 (cdc2) to form the maturation-promoting factor (MPF). Two alternative transcripts have been found, a constitutively expressed transcript and a cell cycle-regulated transcript, which is expressed predominantly during G2/M phase.
	CENPE	0.44	A kinesin-like motor protein that accumulates in the G2 phase of the cell cycle. Unlike other centrosome-associated proteins, it is not present during interphase and first appears at the centromere region of chromosomes during prometaphase. CENPE is proposed to be one of the motors responsible for mammalian chromosome movement and/or spindle elongation.
	EGR1	0.46	A nuclear protein that belongs to the EGR family of C2H2-type zinc-finger proteins and functions as transcriptional regulator. The products of its target genes are required for differentiation and mitogenesis. Studies suggest this is a cancer suppressor gene.
	DEPDC1	0.48	Contains a DEP domain and a Rho-GTP domain. The DEP domain is found in more than 50 proteins involved in G protein signaling pathways. It has been proposed that the DEP domain could play a selective role in targeting DEP domain-containing proteins to specific subcellular membranous sites (Burchett 2000). Small G proteins of the Rho family, which includes Rho, Rac and Cdc42, regulate phosphorylation pathways that control a range of biological functions including cytoskeleton formation and cell proliferation.

Table 4.2.2-8. Maximally responding genes in G361 cells, induced or repressed. The fold-change (FC) was based on the comparison to the respective chronological control, at either 3 or 6 hr following IR. Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

Analysis of HepG2 IR response

Four clusters were identified by CLICK (Fig. 4.2.2-3). The two major clusters represent two main themes: down-regulation in response to IR (cluster 1, 320 probesets) and up-regulation (cluster 2, 300 probesets).

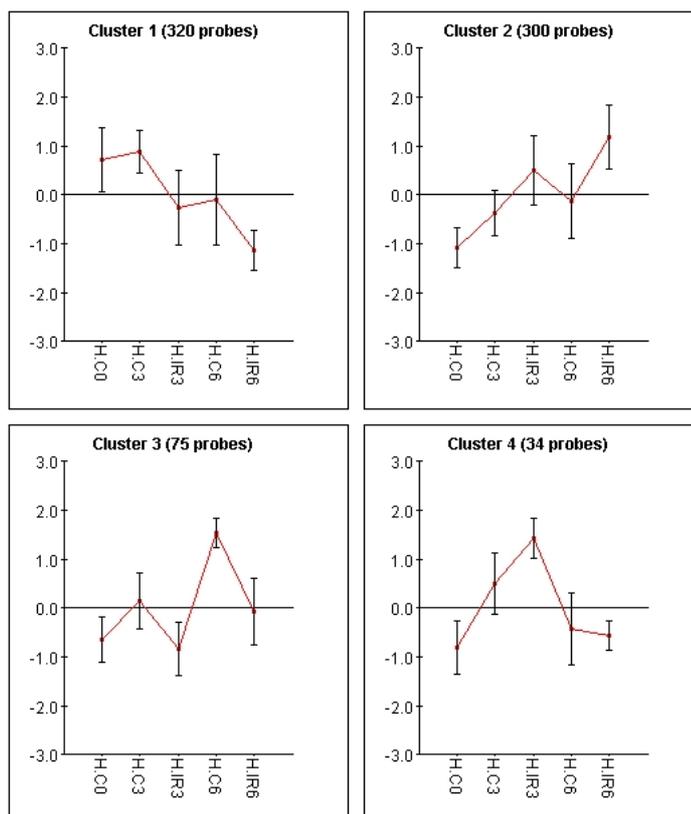


Figure 4.2.2-3. Clusters identified by CLICK in the set of responding genes of HepG2 cells. Clustering method and technical details are as in Figure 4.2.2-1.

Functional categories within gene clusters. Enriched functional categories related to cell cycle were detected in the down-regulated cluster.

Functional category	Num. of Genes associated with the category	Raw p (corrected p)
microtubule-based process, GO:0007017	13	1.141E-7 (0.0020)
cell division, GO:0051301	13	1.846E-5 (0.032)
Mitosis, GO:0007067	17	4.149E-9 (0.001)

Table 4.2.2-9. Functional categories enriched in down-regulated genes set (cluster 1) of HepG2 cells.

Computational search for mediating transcriptional regulators. Transcriptional regulators

that control the observed modulation in gene expression following IR were detected in both down and up-regulated clusters.

Cluster [Regulation]	Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor	p-value
1 [Down]	E2F [M00918]	1.7	3.84E-4
	NF-Y [M00287]	1.8	8.3E-9
	IRF-7 [M00453]	1.8	4.67E-4
2 [Up]	E2F [M00803]	1.1	2.2E-4
	p53 [M00034]	2.36	1.28E-8

Table 4.2.2-10. TF binding site signatures enriched in down- and up-regulated gene sets of HepG2 cells (p<0.0005). Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene, except for p53, for which the area from -2,000 to 1,000 bp was scanned.

Identification of genes with maximal response. We detected the ten genes that were maximally responding to IR in both directions (up and down), as described above for BJ-hTert. The five induced genes were members of cluster 2, and the five repressed genes were members of cluster 1, as expected.

Regulation	Gene Symbol	FC	Description of the protein product
Up	FUCA1	6.05	Defective FUCA1 causes accumulation of fucose in the tissues, leading to Fucosidosis disease, an autosomal recessive lysosomal storage disease. Different phenotypes include clinical features such as neurologic deterioration, growth retardation, progressive psychomotor deterioration and angiokeratoma.
	TP53INP1	4.57	In response to double-strand DNA breaks, the encoded protein is induced in a p53-dependent way, induces cell cycle arrest in G1 and enhances p53 mediated apoptosis.
	FETUB	4.55	A member of the fetuin family, part of the cystatin superfamily of cysteine protease inhibitors. Fetuins have been implicated in several diverse functions, including osteogenesis and bone resorption, regulation of the insulin and hepatocyte growth factor receptors, and response to systemic inflammation.
	LOC340109	3.86	Hypothetical protein, no domains were found in its sequence [137]
	CPM	3.80	A membrane-bound arginine/lysine carboxypeptidase. Its expression is associated with monocyte to macrophage differentiation.
Down	FLJ22624	0.32	The protein encoded by this gene is required for the activation of Aurora-A (AURKA) at the onset of mitosis.
	C15orf20	0.35	A human helicase, also known as PIF1, is cell cycle regulated [139] and inhibits telomerase activity [140].
	CENPE	0.36	A kinesin-like motor protein that accumulates in the G2 phase of the cell cycle. Unlike other centrosome-associated proteins, it is not present during interphase and first appears at the centromere region of chromosomes during prometaphase. CENPE is proposed to be one of the motors responsible for mammalian chromosome movement and/or spindle elongation.
	CENPA	0.37	A centromere protein that contains a histone H3 related histone fold domain that is required for targeting to the centromere. CENPA is proposed to be a component of a nucleosome-like structure.
	FAM83D	0.38	A member of the FAM83 family, a mitotic spindle protein [141].

Table 4.2.2-11: Maximally responding genes in HepG2 cells, induced or repressed. The fold-change (FC) was measured from the respective chronological control, at either 3 or 6 hr following IR. Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

Analysis of TK6 IR response

Fifteen clusters were identified by CLICK (**Fig. 4.2.2-4**). Of special interest are the four major clusters, and cluster 7 because it seems very 'clean' (the response to IR is clearly seen), and is controlled by an important regulator. Clusters 1, 4 and 7 represent down-regulation, and clusters 2 and 3 - up-regulation.

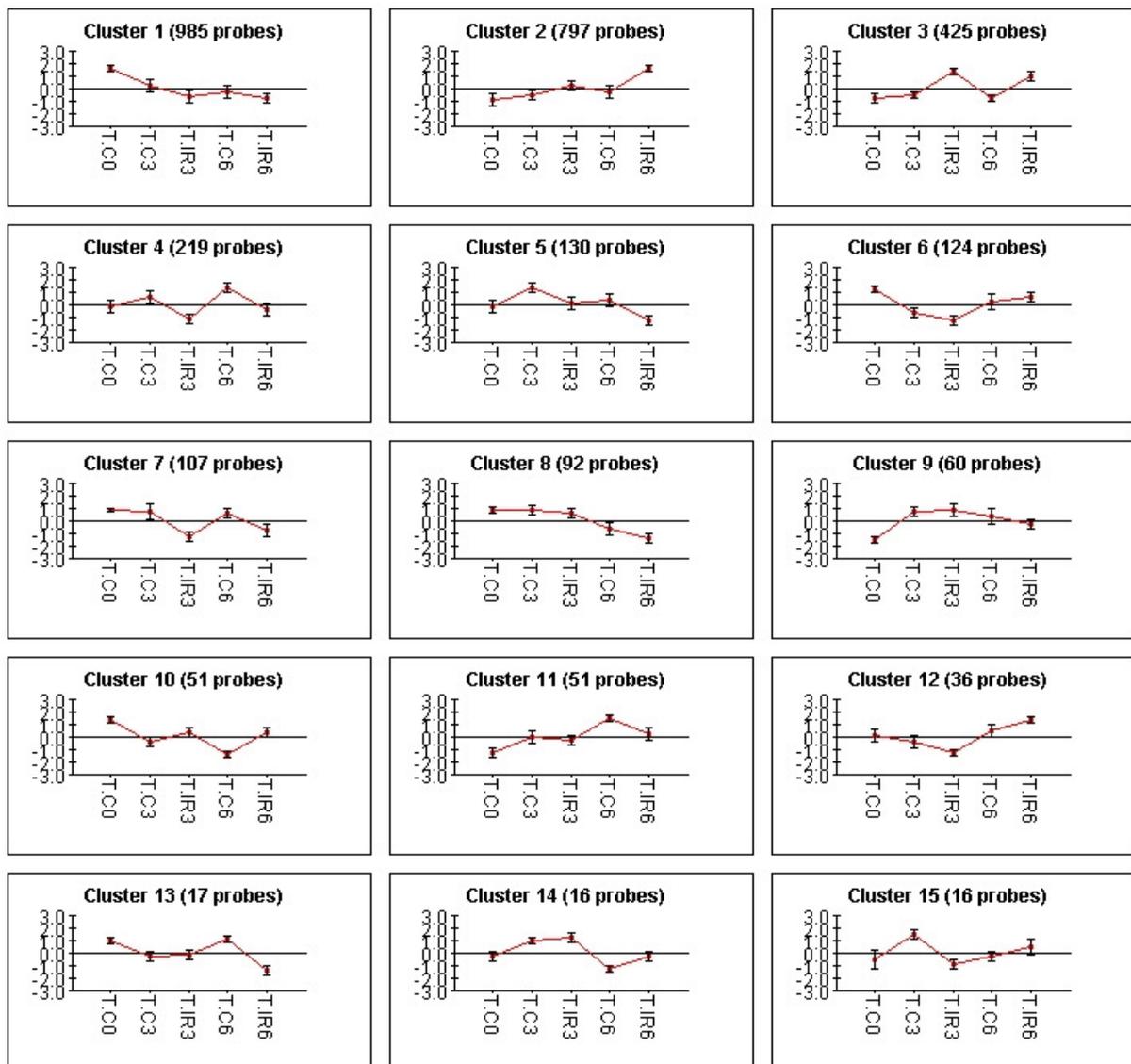


Figure 4.2.2-4. Clusters identified by CLICK in the set of responding genes of TK6 cells. Clustering method and technical details are as in Figure 4.2.2-1.

Functional categories within gene clusters. Enriched functional categories, mostly related to cell cycle, were detected in the down-regulated clusters 1 and 4.

Cluster [Regulation]	Functional category, GO ID	Number of Genes associated with the category	Raw p (corrected p)
1 [Down]	regulation of nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism, GO:0019219	147	2.771E-13 (0.001)
	DNA binding, GO:0003677	127	1.068E-8 (0.001)
	regulation of physiological process, GO:0050791	199	1.208E-12 (0.001)
	nucleobase\, nucleoside\, nucleotide and nucleic acid metabolism, GO:0006139	207	1.478E-10 (0.001)
	transcription\, DNA-dependent, GO:0006351	140	1.279E-12 (0.001)
	ion binding, GO:0043167	188	1.088E-9 (0.001)
	zinc ion binding, GO:0008270	136	3.290E-10 (0.001)
	nucleic acid binding, GO:0003676	182	8.445E-8 (0.002)
4 [Down]	mitotic cell cycle, GO:0000278	16	1.773E-8 (0.001)

Table 4.2.2-12. Functional categories enriched in down-regulated gene sets (cluster 1 and 4) of TK6 cells.

Computational search for mediating transcriptional regulators. Transcriptional regulators that control the observed modulation in gene expression following IR (and experimental noise) were detected in both down and up-regulated clusters.

Cluster [Regulation]	Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor	p-value
1 [Down]	ETF [M00695]	1.16	1.32E-6
	HIF-1 [M00797]	1.25	1.12E-4
3 [Up]	p53 [M00034]	2.23	8.45E-11
	TATA [M00252]	2.51	4.58E-6
4 [Down]	CACCC-binding factor [M00721]	1.76	1.16E-4
7 [Down]	NF-Y [M00287]	1.81	5.59E-5

Table 4.2.2-13. TF binding site signatures enriched in down- and up-regulated gene sets of TK6 cells (p<0.0005). Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene, except for p53, for which the area from -2,000 to 1,000 bp was scanned.

Identification of genes with maximal response. We detected the ten genes that were maximally responding to IR in both directions (up and down), as described above for BJ-hTert. The five induced genes were members of the up-regulated clusters 2 and 3, as expected. The five repressed genes were members of the down-regulated cluster 4 and 7. None of them was from the down-regulated cluster 1. Inspection of cluster 1 can explain why, since the down-regulation in it is mainly with respect to the absolute control, and not to the chronological controls.

Regulation	Gene Symbol	FC	Description of the protein product
Up	FUCA1	24.12	Defective FUCA1 causes accumulation of fucose in the tissues, leading to Fucosidosis disease, an autosomal recessive lysosomal storage disease. Different phenotypes include clinical features such as neurologic deterioration, growth retardation, progressive psychomotor deterioration and angiokeratoma.
	C4orf18	11.48	Hypothetical protein, function unknown.
	TP53I3	11.43	This gene is induced by p53 and is thought to be involved in p53-mediated cell death. The protein encoded by this gene is similar to oxidoreductases, which are enzymes involved in cellular responses to oxidative stresses and irradiation.
	ASTN2	10.50	A member of the Astrotactin family. Mori et al. reported ASTN2 to be up-regulated following irradiation in primary CD4+ T lymphocytes [142].
	TP53INP1	9.67	In response to double-strand DNA breaks, the encoded protein is induced in a p53-dependent way, induces cell cycle arrest in G1 and enhances p53 mediated apoptosis.
Down	PSRC1	0.11	A proline-rich protein. Studies of the related mouse gene suggest that this gene is regulated by p53 and may participate in p53-mediated growth suppression.
	CENPE	0.20	A kinesin-like motor protein that accumulates in the G2 phase of the cell cycle. Unlike other centrosome-associated proteins, it is not present during interphase and first appears at the centromere region of chromosomes during prometaphase. CENPE is proposed to be one of the motors responsible for mammalian chromosome movement and/or spindle elongation.
	KIF14	0.24	A kinesin-like motor protein, which contains an ATP-binding motif and a microtubule-binding part. RNA interference-mediated silencing of KIF14 disrupts cell cycle progression and induces cytokinesis failure[143], and overexpression of KIF14 is associated with cancer [144].
	AURKA	0.27	A cell cycle-regulated kinase that appears to be involved in microtubule formation and/or stabilization at the spindle pole during chromosome segregation. Found at the centrosome in interphase cells and at the spindle poles in mitosis. May play a role in tumor development and progression.
	C15orf20	0.28	A human helicase, also known as PIF1, is cell cycle regulated [139] and inhibits telomerase activity [140]

Table 4.2.2-14: Maximally responding genes in TK6 cells, induced or repressed. The fold-change (FC) was measured from the respective chronological control, at either 3 or 6 hr following IR. Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

Analysis of U2OS IR response

Two clusters were identified by CLICK (**Fig. 4.2.2-5**). These two clusters represent two main themes: up-regulation in response to IR (cluster 1, 214 probesets) and down-regulation (cluster 2, 210 probesets).

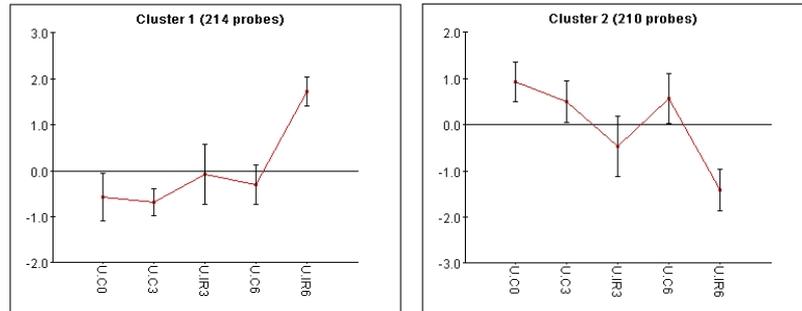


Figure 4.2.2-5: Clusters identified by CLICK in the set of responding genes of U2OS cells. Clustering method and technical details are as in Figure 4.2.2-1.

Functional categories within gene clusters. Functional categories related to cell cycle were detected in the down-regulated cluster, as observed in **Table 4.2.2-15**.

Functional category	Num. of Genes associated with the category	Raw p (corrected p)
cell cycle, GO:0007049	34	1.365E-12 (0.001)
microtubule-based process, GO:0007017	14	2.126E-10 (0.001)
spindle organization and biogenesis, GO:0007051	8	2.436E-10 (0.001)
cell division, GO:0051301	17	5.376E-11 (0.001)
Mitosis, GO:0007067	22	2.136E-17 (0.001)
chromosome segregation, GO:0007059	6	2.329E-6 (0.004)

Table 4.2.2-15: Functional categories enriched in down-regulated genes set (cluster 2) of U2OS cells (corrected p<0.01).

Computational search for mediating transcriptional regulators. The mediator of activation of G1/S checkpoint, DNA repair and apoptosis, p53, was enriched in the up-regulated genes. NF- κ B, which controls the expression of several key regulators of the G2/M phases of the cell cycle, was enriched in the cluster of down-regulated genes (**Table 4.2.2-16**).

Cluster [Regulation]	Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor	p-value
1 [Up]	p53 [M00034]	2.62	5.99E-8
2 [Down]	NF-Y [M00287]	2.17	1.5E-9

Table 4.2.2-16: TF binding site signatures enriched in down and up-regulated gene sets of U2OS cells ($p < 0.0005$). Analysis was done on the region extending 1,000 bp upstream to 200 bp downstream to the putative TSS of each gene, except for p53, for which the area from -2,000 to 1,000 bp was scanned.

Identification of genes with maximal response. We detected ten genes that were maximally responding to IR in both directions (up and down), as described above for BJ-hTert. The five induced genes were members of cluster 1, and the five repressed genes were members of cluster 2, as expected.

Regulation	Gene Symbol	FC	Description of the protein product
Up	GDF15	6.95	A bone morphogenetic protein, member of the transforming growth factor-beta superfamily that regulates tissue differentiation and maintenance. Kis et al. detected GDF15 as one of the signaling proteins up-regulated following DNA damage [132]
	TP53INP1	5.49	In response to double-strand DNA breaks, the encoded protein is induced in a p53-dependent way, induces cell cycle arrest in G1 and enhances p53 mediated apoptosis.
	KITLG	4.94	This gene encodes the ligand of the tyrosine-kinase receptor encoded by the KIT locus. This ligand is a pleiotropic factor that acts in utero in germ cell and neural cell development, and hematopoiesis, all believed to reflect a role in cell migration. In adults, it functions pleiotropically, while mostly noted for its continued requirement in hematopoiesis.
	BTG2	4.24	A member of the BTG/Tob family, which consists of structurally related proteins that appear to have antiproliferative properties. This encoded protein is involved in the regulation of the G1/S transition of the cell cycle.
	FAS	4.01	A member of the TNF-receptor superfamily. This receptor contains a death domain. It has been shown to play a central role in the physiological regulation of programmed cell death, and has been implicated in the pathogenesis of various malignancies and diseases of the immune system. This receptor was found to be involved in transducing the proliferating signals in normal diploid fibroblast and T cells.
Down	C15orf20	0.20	A human helicase, also known as PIF1, is cell cycle regulated [139] and inhibits telomerase activity [140]
	GAS2L3	0.29	Growth arrest-specific 2 like 3 is a member of the GAS2 family, and contains a calponin homology domain (CH-domain), a superfamily of actin-binding domains found in both cytoskeletal proteins and signal transduction proteins [145]. However, in calponins, there is evidence that the CH domain is not involved in its actin binding activity [146].
	KIF14	0.31	A kinesin-like motor protein, which contains an ATP-binding motif and a microtubule-binding part. RNA interference-mediated silencing of KIF14 disrupts cell cycle progression and induces cytokinesis failure [143], and overexpression of KIF14 is associated with cancer [144].
	CCNB1	0.34	A regulatory protein involved in mitosis. The gene product complexes with p34 (cdc2) to form the maturation-promoting factor (MPF). Two alternative transcripts have been found, a constitutively expressed transcript and a cell cycle-regulated transcript, which is expressed predominantly during G2/M phase.
	PLK1	0.34	The protein encoded by this gene is required for cell division. Jang et al. reported that Plk1 is an important regulator of several events during mitosis, and is also involved in DNA damage check-point in G2 phase [138].

Table 4.2.2-17: Maximally responding genes in U2OS cells, induced or repressed. The fold-change (FC) was calculated based on the respective chronological control, at either 3 or 6 hr after irradiation. Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

4.2.3 Integrative analysis of the five cell lines

The hierarchical clustering of all cell lines revealed that the samples are most influenced by cell line identity. The analysis of the IR response separately for each cell line demonstrated resemblance in the IR response of the different cell lines: mainly a decrease in expression of cell cycle genes. In this section we will try to identify the genes that respond universally to IR, and on the other hand, identify the cell line-specific responding genes.

One-way ANOVA ($p < 0.001$) was applied on the dataset containing all 75 measurements, yielding 23,296 probesets (out of 25,470 'present' probesets) for further analysis. As mentioned in a previous section, many probesets passed this filter due to the great variance among the cell types. Averaging expression levels over replicates, our dataset contained measurements for twenty-five conditions.

Analysis of genes responding to IR in more than one cell line

In this analysis we tried to filter out genes that responded merely to the side effect of the experimental manipulation, and not to IR. For each probeset, we required that the expression value measured 3 or 6 hr after irradiation would pass a threshold of fold-change from time 0, and also pass a lower threshold fold-change from the respective chronological control. We defined the damage-responding gene set as all genes whose expression levels changed (at either 3 or 6 hr) by at least 1.5-fold from the absolute control and at least 1.2-fold of the respective chronological control, in the same direction (up- or down-regulation).

This dual filter was applied several times on the data, each time with specific aim: finding universally responding genes (which respond in all cell lines); genes that respond in four of the cell lines, three of them, and so on. The solution received is actually a biclustering solution.

We obtained sets of damage-responding genes according to the direction of response (increasing or decreasing levels of expression) and according to the number of cell lines in which

the genes were responding. **Table 4.2.3-1** shows the number of genes whose expression levels were found to increase or decrease in different subsets of cell lines.

Num. of cell lines	Num. of up-regulated genes	Num. of down-regulated genes
5	13	15
4	30	25
3	48	56
2	152	217

Table 4.2.3-1. Number of genes that were found responding to IR in all five, or part of the cell lines. We defined the 'responding genes' as genes whose expression levels changed after IR (at either 3 or 6 hr) by at least 1.5-fold from the absolute control and at least 1.2-fold of the respective chronological control, in the same direction (up- or down-regulation).

We created a partial heat map of **Table 4.2.3-1**, depicting visually three prominent expression patterns of the biclustering solution, for each direction (up- and down-regulation): The upper-, middle-, lower-clusters in each of the matrices in **Figure 4.2.3-1** represent genes that are induced (reduced) in 5, 4 and 3 of the lines.

Figure 4.2.3-1. Major expression patterns in the damage-induced and -reduced gene set in an integrative analysis of the 5 cell lines. The letters 'B', 'G', 'H', 'T' and 'U' stand for BJ-hTert, G361, HepG2, TK6 and U2OS cell lines. For each of the damage-induced (reduced) probes, \log_2 induction (reduction) fold of expression level 3 or 6 hr after IR treatment was computed in each cell line. The highest fold (at either 3 or 6 hr) was taken, yielding a 5 column data matrix, with the rows corresponding to probes and the columns to cell lines. The heat map visually represents the clustered matrix. Red, green and black entries represent above-, below- and near-average fold of induction, respectively. In each of the matrices, three prominent expression patterns are evident: The upper-, middle-, lower-clusters represent genes that are induced (reduced) in 5, 4 and 3 of the lines.



Exploring the genes that universally respond to IR. In order to gather information about the similarity of response to IR in the five cell lines, we analyzed those genes that were found significantly responding in more than one cell line.

It is of great interest to explore the genes that showed strong response in all cell lines of this study. The up-regulated genes are presented in **Table 4.2.3-2**, along with their major known function. These genes were those involved in apoptosis, cell cycle checkpoints, ROS protection, metabolism and DNA repair. Of note, an unknown transcript that has similarity to an ALU sequence as been up-regulated universally in our study.

Gene/Protein Name	Description of the protein product
230330_at*	This target has similarity to an ALU sequence. It has a Thiolase active site, 2Fe-2S ferredoxins iron-sulfur binding region signature and a signal-peptide domain (which indicates the protein will be secreted) [137].
BTG2	A member of the BTG/Tob family, which consists of structurally related proteins that appear to have antiproliferative properties. This encoded protein is involved in the regulation of the G1/S transition of the cell cycle.
C12orf5	Regulated as part of the p53 tumor suppressor pathway. The protein functions by blocking glycolysis and directing the pathway into the pentose phosphate shunt. Expression of this protein also protects cells from DNA damaging reactive oxygen species and provides some protection from DNA damage-induced apoptosis.
CDKN1A	Functions as a regulator of cell cycle progression at G1, and plays a regulatory role in S phase DNA replication and DNA damage repair. The expression of this gene is tightly controlled by the tumor suppressor protein p53, through which this protein mediates the p53-dependent cell cycle G1 phase arrest in response to a variety of stress stimuli.
CPM	A membrane-bound arginine/lysine carboxypeptidase. Its expression is associated with monocyte to macrophage differentiation.
DDB2	The smaller subunit of a heterodimeric protein implicated in the etiology of xeroderma pigmentosum group E (XPE), which functions in Nucleotide Excision Repair (NER).
FAS	A member of the TNF-receptor superfamily. This receptor contains a death domain. It has been shown to play a central role in the physiological regulation of programmed cell death, and has been implicated in the pathogenesis of various malignancies and diseases of the immune system. This receptor was found to be involved in transducing the proliferating signals in normal diploid fibroblast and T cells.
FDXR	A mitochondrial flavoprotein that initiates electron transport for cytochromes P450 receiving electrons from NADPH. The expression of this gene is regulated by the p53 family and sensitizes cells to oxidative stress-induced apoptosis [147].

MDM2	A target gene of the transcription factor tumor protein p53. The encoded protein is a nuclear phosphoprotein that binds and inhibits transactivation by tumor protein p53, as part of an autoregulatory negative feedback loop. This protein has E3 ubiquitin ligase activity, which targets tumor protein p53 for proteasomal degradation. This protein also affects the cell cycle, apoptosis, and tumorigenesis through interactions with other proteins, including retinoblastoma 1 and ribosomal protein L5.
hypothetical protein MGC5370	This unknown protein has a signal-peptide domain (which indicates the protein will be secreted) and trans-membrane regions [137].
PPM1D	A member of the PP2C family of Ser/Thr protein phosphatases. PP2C family members are known to be negative regulators of cell stress response pathways. The expression of this gene is induced in a p53-dependent manner in response to various environmental stresses. This phosphatase negatively regulates the activity of p38 MAP kinase, MAPK/p38, through which it reduces the phosphorylation of p53, and in turn suppresses p53-mediated transcription and apoptosis. This gene is located in a chromosomal region known to be amplified in breast cancer.
SESN1	A member of the Sestrins protein family, whose expression is modulated by p53. The encoded protein is a potential regulator of cellular growth.
TP53INP1	In response to double-strand DNA breaks, the encoded protein is induced in a p53-dependent way, and induces cell cycle arrest in G1 and enhances p53 mediated apoptosis.

Table 4.2.3-2. Major functions of genes up-regulated universally (in all five cell lines). Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

* Probeset ID of Affymetrix was used, where there was no protein name available.

We used the SPIKE tool to search for DNA damage sub-network maps, in which this set of universally up-regulated genes was enriched. SPIKE is a comprehensive, up-to-date knowledge-base of cancer-related signaling pathways. The entire SPIKE database (containing all annotated human genes) served as background. The set of universally up-regulated genes was enriched in the p53 map with a p-value of 8.99E-23. Strikingly, 10 out of the 13 induced genes are known to be p53-regulated genes, as observed in **Figure 4.2.3-2**.

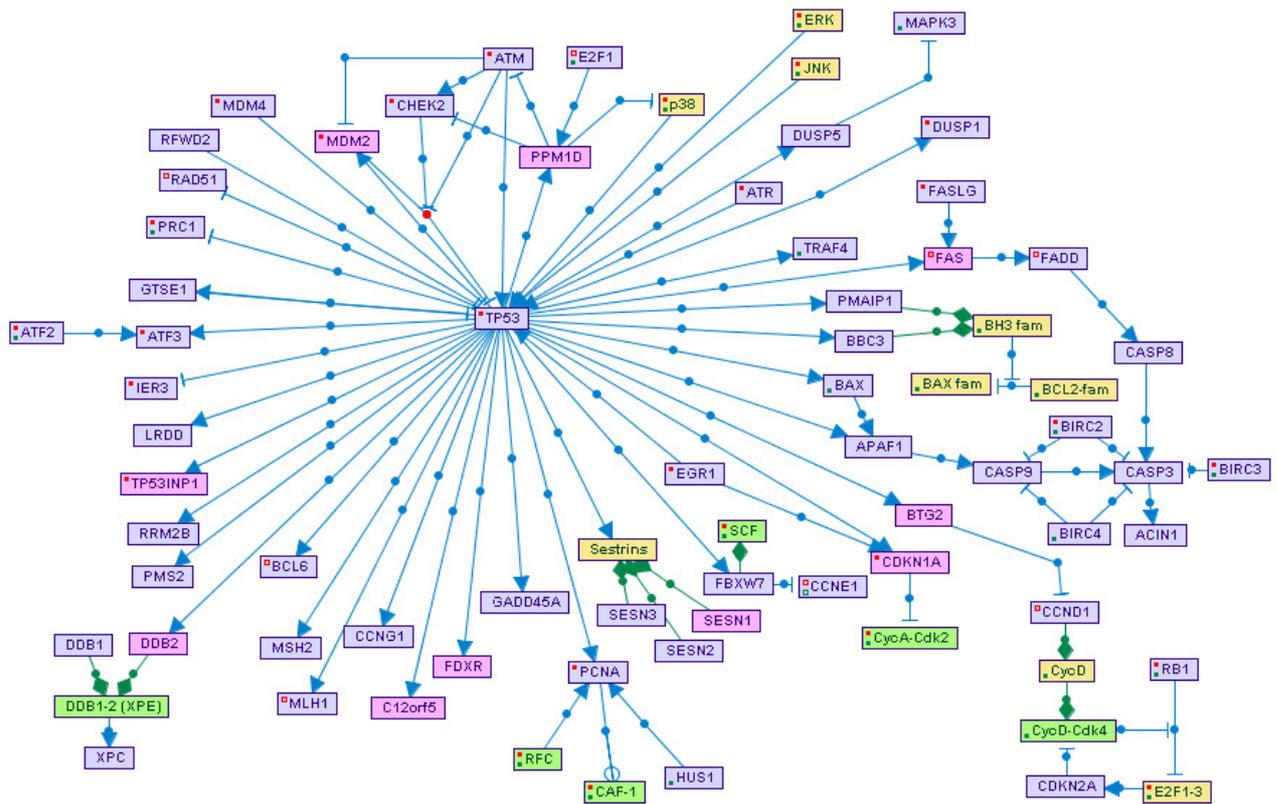


Figure 4.2.3-2. Superposition of the cluster of genes induced in all cell lines on p53-regulated network. This figure was generated using the SPIKE tool. The genes that were found up-regulated in all cell lines are pinkish in color. The white nodes correspond to proteins; green nodes represent protein complexes, and yellow nodes - protein families. Blue edges represent regulations: arrows correspond to activation; T shape edges – to inhibition; and open circles denote regulations whose effect is still unclear. Green edges represent association between nodes (e.g., association between a protein complex and its components). Red and green dots within a node indicate that not all the regulations and associations stored in SPIKE database for the node are displayed in the map. Superimposing the data on this map clearly shows that the p53 network was activated by the examined stress.

The genes whose expression was significantly down-regulated at 3 or 6 hr in all five lines, are listed in **Table 4.2.3-3** Most of these genes facilitate transit through the cell cycle.

Gene Name	Gene Function
ASPM	The human ortholog of the <i>Drosophila melanogaster</i> 'abnormal spindle' gene (<i>asp</i>), which is essential for normal mitotic spindle function in embryonic neuroblasts[148]. The mouse gene <i>Aspm</i> is expressed specifically in the primary sites of prenatal cerebral cortical neurogenesis.
AURKA	A cell cycle-regulated kinase that appears to be involved in microtubule formation and/or stabilization at the spindle pole during chromosome segregation. Found at the centrosome in interphase cells and at the spindle poles in mitosis. May play a role in tumor development and progression.
AURKB	A cell cycle-regulated kinase that associates with microtubules during chromosome movement and segregation. Localizes to microtubules near kinetochores, specifically to the specialized microtubules called K-fibers.

BUB1	A kinase involved in spindle checkpoint function. The kinase functions in part by phosphorylating a member of the mitotic checkpoint complex and activating the spindle checkpoint. Mutations in this gene have been associated with aneuploidy and several forms of cancer.
CCNB1	A regulatory protein involved in mitosis. The gene product complexes with p34 (cdc2) to form the maturation-promoting factor (MPF). Two alternative transcripts have been found, a constitutively expressed transcript and a cell cycle-regulated transcript, which is expressed predominantly during G2/M phase.
CDCA8	A component of a chromosomal passenger complex required for stability of the bipolar mitotic spindle.
CENPA	A centromere protein which contains a histone H3 related histone fold domain that is required for targeting to the centromere. CENPA is proposed to be a component of a nucleosome-like structure, in which it replaces one or both copies of conventional histone H3 in the (H3-H4) ₂ tetrameric core of the nucleosome particle.
CENPE	A kinesin-like motor protein that accumulates in the G2 phase of the cell cycle. Unlike other centrosome-associated proteins, it is not present during interphase and first appears at the centromere region of chromosomes during prometaphase. CENPE is proposed to be one of the motors responsible for mammalian chromosome movement and/or spindle elongation.
DEPDC1	Contains a DEP domain and a Rho-GTP domain. The DEP domain is found in more than 50 proteins involved in G protein signaling pathways. It has been proposed that the DEP domain could play a selective role in targeting DEP domain-containing proteins to specific subcellular membranous sites [149]. Small G proteins of the Rho family, which includes Rho, Rac and Cdc42, regulate phosphorylation pathways that control a range of biological functions including cytoskeleton formation and cell proliferation.
KIAA1333	A G2/M-specific gene with DNA damage responsive expression. Brooks et al. suggest a possible role in cell cycle regulation and the cellular response to DNA damage [150]. Contains a HECT domain, pointing to an ubiquitin-protein ligase activity.
KIF14	A kinesin-like motor protein, which contains an ATP-binding motif and a microtubule-binding part. RNA interference-mediated silencing of KIF14 disrupts cell cycle progression and induces cytokinesis failure [143], and overexpression of KIF14 is associated with cancer [144].
KIF18A	A kinesin-like motor protein, which contains an ATP-binding motif and a microtubule-binding part.
PSRC1	A proline-rich protein. Studies of the related mouse gene suggest that this gene is regulated by p53 and may participate in p53-mediated growth suppression.
SGOL2	An inner centromere protein that is essential for centromere cohesion, whose localization depends on the BUB1 and Aurora B kinases [151]. This protein is proposed to contribute to the spatial regulation of MCAK activity, which corrects defective kinetochore attachments.
TROAP	Plays an important role in mammalian cells by associating with the microtubular cytoskeleton [152]

Table 4.2.3-3. Major functions of genes down-regulated universally (in all five cell lines). Gene description was obtained from Entrez-Gene (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Otherwise, references are provided.

We used SPIKE to search for DNA damage sub-network maps, in which this set of universally down-regulated genes was enriched. It was enriched in the G2-M map with a p-value of 4.83E-17. Eight out of the 15 suppressed genes are involved in G2-M phase of the cell cycle, as observed in **Figure 4.2.3-3**.

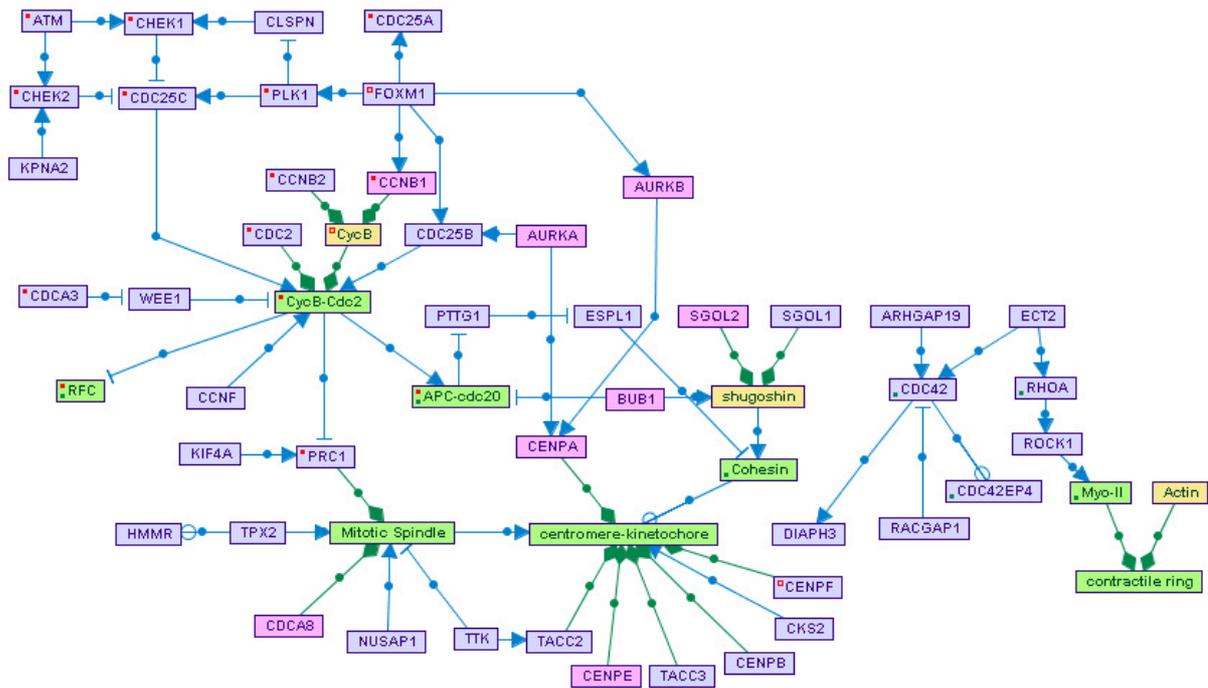


Figure 4.2.3-3. Superposition of the genes down-regulated in all cell lines on G2-M network. This figure was generated using SPIKE. The genes that were found to be down-regulated in all cell lines are pinkish in color. Symbols and colors are as in **Figure 4.2.3-2**. Superimposing the data on this map clearly shows that G2-M network was suppressed by the examined stress.

The genes that were up-regulated in four of the cell lines were found enriched in the p53 map, with p-value of 1.86E-4. The genes down-regulated in four of the cell lines were found mostly enriched in the G2-M map, with p-value of 2.79E-11.

Identification of major expression patterns in the groups of genes responding in more than one cell line. The genes that were found significantly responding in more than one cell line (**Table 4.2.3-1**), were loaded to EXPANDER as clusters. In all, eight clusters were loaded: four clusters of up-regulated genes, and four of down-regulated genes. Each cluster, except for the one representing genes that were universally responding, was a union of a few groups. For example, the cluster of genes up-regulated in four cell lines, was a union of genes responding in all lines except for one (either BJ-hTert, G361, HepG2, TK6 or U2OS). Analysis of these clusters helped us recognize the similarity between the cell lines.

Analysis was done also on non-unified clusters, e.g. all genes that respond in all lines except for BJ-hTert. No additional information was added from the non-unified analysis, in comparison to the unified analysis, and the results were less significant. Therefore, only results of the unified analysis are presented below.

Functional categories within gene clusters. Enriched functional categories (corrected $p < 0.05$) were identified in three down-regulated clusters and only one up-regulated cluster (**Table 4.2.3-4**). This is of no surprise, since in the analysis of each cell line separately, no functional enrichment was found for the up-regulated clusters. The down-regulated clusters are again enriched with cell cycle related functions.

Cluster [Regulation]	Functional category	Num. of Genes associated with the category	Raw p (corrected p)
5 cell lines [Up]	negative regulation of cell proliferation, GO:0008285	5	2.15E-8 (0.001)
4 cell lines [Down]	mitotic cell cycle, GO:0000278	10	2.98E-12 (0.001)
3 cell lines [Down]	cell division, GO:0051301	8	4.61E-7 (0.005)
	M phase, GO:0000279	9	6.28E-8 (0.001)
	mitotic cell cycle, GO:0000278	10	8.81E-9 (0.001)
2 cell lines [Down]	cell division, GO:0051301	13	8.27E-7 (0.008)
	cell cycle, GO:0007049	32	5.43E-10 (0.001)
	mitotic cell cycle, GO:0000278	18	5.07E-10 (0.001)
	regulation of cellular process, GO:0050794	65	1.86E-7 (0.002)

Table 4.2.3-4. Functional categories enriched in gene sets up and down-regulated in more than one cell line (corrected $p < 0.01$). Each cluster, except for the one representing genes that were universally responding, is a union of a few groups.

Computational search for mediating transcriptional regulators. Enriched binding sites signatures ($p < 1E-3$) were identified in all clusters (**Table 4.2.3-5**).

Cluster [Regulation]	Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor	p-value
5 cell lines [Up]	p53 [M00034]	5.61	4.84E-5
4 cell lines [Up]	p53 [M00034]	5.76	8.37E-9
	TBP [M00252]	7.81	1.05E-6
3 cell lines [Up]	TBP [M00252]	5.00	4.08E-5
2 cell lines [Up]	p53 [M00034]	2.02	3.64E-4
5 cell lines [Down]	NF-Y [M00287]	3.51	8.27E-5
4 cell lines [Down]	NF-Y [M00287]	3.66	5.21E-8
3 cell lines [Down]	NF-Y [M00287]	2.89	1.82E-9
2 cell lines [Down]	NF-Y [M00287]	1.54	2.0E-4

Table 4.2.3-5. TF binding site signatures enriched in gene sets up and down-regulated in more than one cell line ($p < 0.0005$). Each cluster, except for the one representing genes that were universally responding, is a union of a few groups. Analysis was done on the region from 1,000 bp upstream to 200 bp downstream to genes' putative TSS, except for p53, with scan from -2000 to 1000.

Analysis of genes that respond to IR uniquely in one cell line

In order to understand the difference in response of cell lines to IR, we looked for probesets that were found significantly responding in only one cell line, i.e., probesets that apparently represent genes that are uniquely responding to IR in one specific cell line. To this end, we applied a dual criterion to define a responding probeset in one of the cell lines, and non-response in the remaining four cell lines.

A probeset was considered responding in a cell line if the expression value measured 3 or 6 hr after irradiation would pass a threshold fold-change of 1.5 from time 0, and a threshold fold-change of 1.2 from the respective chronological control, in the same direction (up or down).

A probeset was considered *not*-responding in a cell line if the expression value measured 3 or 6 hr after irradiation would pass neither a threshold fold-change of 1.45 from time 0, nor a lower

threshold fold-change of 1.15 from the respective chronological control, in the same direction (up- or down-regulation).

Table 4.2.3-6 depicts the number of probesets that were responding to IR in one specific cell line and not in the other cell lines. TK6 seemed to have the largest number of uniquely responding genes, while G361 has the fewest.

cell line	Up	down
BJ-hTert	38	134
G361	7	9
HepG2	20	15
TK6	296	237
U2OS	48	52

Table 4.2.3-6. Number of probesets that were responding to IR in one specific cell line and not in the other cell lines.

The genes that were found significantly responding in one cell line (**Table 4.2.3-6**), were loaded to EXPANDER as clusters. Most of the clusters are relatively small, and therefore the group analysis was not expected to yield results. For the groups that showed no functional enrichment, we looked at the list of genes and noted interesting ones. The list of genes for each cell line, with their fold-change from the absolute control is provided in the Appendix.

Genes that respond to IR uniquely in BJ-hTert cell line

Genes up-regulated uniquely in BJ-hTert.

We noted that four histone genes were uniquely up-regulated in BJ-hTert cell line. The SPEN gene was also up-regulated in BJ only. SPEN is a transcriptional repressor, whose repression activity can occur through interactions with other repressors, by the recruitment of proteins involved in histone deacetylation,

Genes down-regulated uniquely in BJ-hTert.

The down-regulated genes were found enriched with the following functional categories (corrected $p < 0.05$, in Expander), presented in **Table 4.2.3-7**.

Functional category	Num. of probesets associated with the category	Raw p (corrected p)
response to DNA damage stimulus, GO:0006974	17	5.59E-12 (0.001)
cell cycle checkpoint, GO:0000075	5	1.87E-5 (0.028)
cell cycle, GO:0007049	19	7.59E-7 (0.002)
DNA replication, GO:0006260	25	3.81E-26 (0.001)
DNA metabolism, GO:0006259	33	6.71E-21 (0.001)

Table 4.2.3-7. Functional categories enriched in gene sets down regulated uniquely in BJ-hTert cell line.

We then conducted a computational search for mediating transcriptional regulators. Enriched binding sites signatures ($p < 5E-5$) were identified, shown in **Table 4.2.3-8**.

Transcription factor [Accession num. in TRANSFAC DB]	Enrichment factor	p-value
E2F [M00918]	2.22	3.62E-5
NF-Y [M00287]	1.95	3.24E-4

Table 4.2.3-8. TF binding site signatures enriched in gene sets down-regulated in BJ-hTert cell line.

It is surprising that cell cycle genes and transcription factors were down-regulated uniquely in BJ-hTert cells, since we observed a general down-regulation of cell cycle in all cell lines in response to IR. However, inspection of the fold-changes of gene expression in BJ-hTert cells compared to the other cell lines reveals that indeed these genes were responding uniquely in BJ-hTert cells (see Appendix for the list of genes and fold-changes).

We used SPIKE to search for DNA damage sub-network maps, in which this set of universally down-regulated genes was enriched. It was enriched in the G1-S map with a p-value of 5.68E-13, and in the ATM map with p-value 2.38E-10 (Figures 4.2.3-4 and 4.2.3-5).

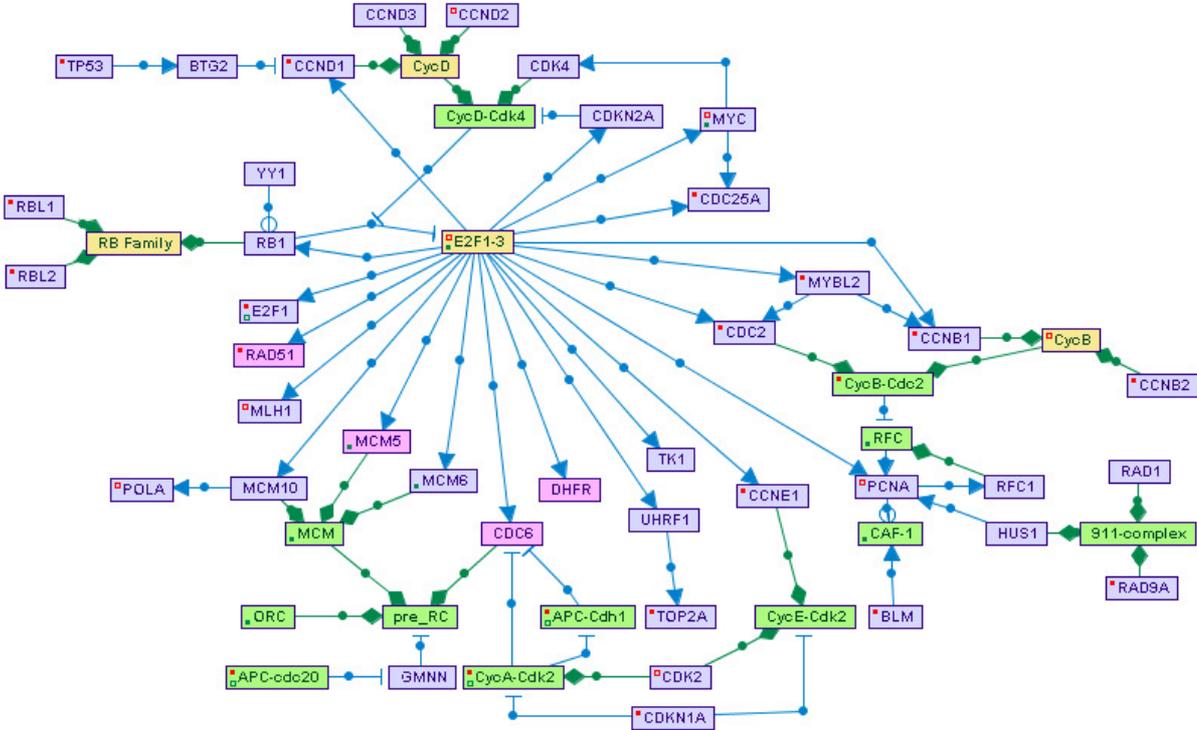


Figure 4.2.3-4. Superposition of the genes down-regulated uniquely in BJ-hTert on G1-S network. This figure was generated using SPIKE. The genes which were found down-regulated in BJ-hTert are pinkish in color. Symbols and colors are as in Figure 4.2.3-2. Superimposing the data on this map clearly shows that the G1-S network was suppressed in BJ-hTert by the examined stress.

septic shock, lung fibrosis, glomerulonephritis, atherosclerosis, and AIDS. In contrast, complete and persistent inhibition of NF-kappa-B has been linked directly to apoptosis, inappropriate immune cell development, and delayed cell growth [153-155].

The unique up-regulation of STC1 in HepG2 (a hepatocellular carcinoma cell line) is interesting. This gene encodes a glycoprotein that is known to have altered expression in hepatocellular, ovarian, and breast cancers.

Genes down-regulated uniquely in HepG2.

CTH was down-regulated in response to IR. This gene encodes a cytoplasmic enzyme in the trans-sulfuration pathway that converts cystathione derived from methionine into cysteine. Glutathione synthesis in the liver is dependent upon the availability of cysteine.

A p53-induced protein, ENC1, which encodes an actin-binding protein, was down-regulated as well in HepG2.

Genes that respond to IR uniquely in TK6 cell line

Genes up-regulated uniquely in TK6.

We noted that the COP1, an inhibitor of apoptosis, is up-regulated only in TK6.

The expression of a few interferons, specifically IFI16, augmented in response to IR. IFI16 interacts with p53 and modulates its function and other cell cycle regulatory factors [156, 157]. Fujiuchi et al. (2004) reported that expression of IFI16 enhanced p53 transcriptional activity in cells exposed to IR, and suggested that loss of IFI16 results in deregulation of p53-mediated apoptosis, leading to cancer development [158].

We noted the up-regulation of RAD52. This gene product was shown to bind single-stranded DNA ends and mediate the DNA-DNA interaction necessary for the annealing of complementary DNA strands. It was also found to interact with DNA repair protein RAD51, which suggested a role for this protein in the homologous recombination pathway of double strand breaks repair.

TRIAD3 encodes a cytoplasmic protein which specifically colocalizes and interacts with the serine/threonine protein kinase, receptor-interacting protein (RIP). Zinc finger domains of the encoded protein are required for its interaction with RIP and for inhibition of TNF- and IL1-induced NF-kappa B activation pathways.

Genes down-regulated uniquely in TK6.

RAD23B and RAD50 appeared down-regulated. The protein encoded by RAD23B is one of two human homologs of *Saccharomyces cerevisiae* Rad23, a protein involved in the nucleotide excision repair (NER). This protein was found to be a component of the protein complex that specifically complements the nucleotide excision repair defect of xeroderma pigmentosum group C (XP-C) cell extracts in vitro [159, 160]. RAD50 is a member of the MRN complex together with MRE11 and NBS1 [161]. The protein complex acts as a sensor of double strand breaks. This complex is important for DNA double-strand break repair, cell cycle checkpoint activation, telomere maintenance, and meiotic recombination [162]. Rad50-knockout mice are not viable, suggesting an essential role for this protein in cell growth and viability [163].

A computational search for mediating transcriptional regulators in the down-regulated genes set, pointed to AP-2alpha [accession M00469], with enrichment factor of 1.74, and p-value of 1.29E-4. AP-2alpha is also one of the genes down-regulated uniquely in TK6. It's a sequence-specific DNA-binding protein that interacts with inducible viral and cellular enhancer elements to regulate transcription of selected genes. AP-2 factors bind to the consensus sequence 5'-GCCNNNGGC-3' and activate genes involved in a large spectrum of important biological functions including developmental process [164-166]. Li et al. (2006) reported that AP-2alpha and AP-2gamma are transcriptional targets of p53 in human breast carcinoma cells [167].

Genes that respond to IR uniquely in U2OS cell line

There was no common denominator to these genes. See the Appendix for a complete list.

5 Discussion

Cellular responses to DNA damage are crucial for maintaining homeostasis and preventing the development of cancer [1]. The response of normal and cancerous cells to the radiation utilized in cancer therapy determines the treatment's success.

Thanks to the advances in functional genomics it is now possible to study biological systems as a whole and obtain large-scale snapshots of cellular transcriptome and proteome.

5.1 Calibration of the experimental system using human fibroblasts exposed to NCS

To calibrate our experimental system we exposed human BJ-hTert cells to 200 ng/mL NCS and recorded global transcriptional responses 2, 4, 8 and 12 hr later, with untreated controls at 0 and 8 hr. In this experiment we compared biological replicates prepared by two researchers, examined the effect of the experimental manipulation on gene expression and identified informative time points for the following, more comprehensive, experiment. We concluded that there was very good correlation between the intra-laboratory replicates, yielding reliable and significant results.

We noted that the manipulation of culture dishes during the experimental procedure affected gene expression profiles in the cells. Subsequently we took this information into consideration while designing the second study: we had "chronological controls" for each time point in the study, in addition to the absolute control.

Taking into account both the clustering kinetics and the regulatory analysis of each time point, we realized that the time point of 6 hr after exposure was the most informative. In order to cover better the early wave of response we added a 3 hr time point in the subsequent experiment[130].

5.2 Delineation of transcriptional responses induced by IR in five human cell lines

We applied gene expression microarrays combined with a computational battery, to delineate transcriptional responses induced by a moderate dose of IR in five human cell lines, three of which were cancerous: BJ-hTert (foreskin fibroblasts), G361 (melanoma), HepG2 (hepatocellular carcinoma), TK6 (lymphoblastoids) and U2OS (osteosarcoma), 3 and 6 hours after irradiation of 5 Gy.

A major obstacle in the search for new cancer drug targets is that the drugs are often toxic to normal tissues and yet are used at high doses in order to kill tumor cells. Therefore cellular targets that can be affected by low treatment doses are interesting since they could be used to reduce the harmful effects of the treatment [168]. For this reason, we used a clinically relevant dosage of several Gy. Studying expression profiles in untreated and irradiated normal and cancerous human cell lines can potentially elucidate molecular determinants that affect sensitivity to IR and other chemotherapy chemicals.

5.2.1 Hierarchical clustering of the cell lines

When performing hierarchical clustering on the expression profiles using a large portion of the probesets (~25k), they were divided primarily into cell lines, and not according to IR treatment (**Fig. 4.2.1-1**). In retrospect, this is expected, since we were looking at a very wide portion of the transcriptional map, which is highly influenced by cell type. Several hundred genes were expected to respond to the treatment.

Cell lines with similar tissues of origin, BJ-hTert and G361 derived from the skin, grouped together. This is in accordance to Ross et al. (2000) [169], who indicated that specific features of gene expression patterns appear to be related to physiological properties of the cell lines, such as their doubling time in culture, drug metabolism or the interferon response.

Apparently the cancerous or 'normal' property of the cells did not have a major affect on the global gene expression profile. According to the hierarchical clustering, BJ-hTert seems closer to the three cancerous cells than to TK6 cells. This may be explained by basic differences between cultured cells and tissues [170] (see below).

5.2.2 Analysis of differentially expressed genes in specific cell lines

We found large differences among the numbers of probesets that responded to IR in each cell line. TK6 showed the largest number of responding probesets, while G361 had the fewest (Tables 4.2.2-1 and 4.2.2-2). G361 had previously been found to show radioresistance compared to other cancer cell lines [171, 172].

Clustering analysis was carried out for each cell line separately. The two largest clusters of each cell line depicted up- and down-regulation, mainly in response to IR. Examination of the largest cluster of down-regulation of each cell line revealed that all cell lines except for TK6 have a stronger down-regulation at 6 hr compared to 3 hr. However, in TK6 there are smaller clusters that do show stronger down-regulation at 6 hr (clusters 8 and 13). Examination of the largest cluster of up-regulation of each cell line revealed that all cell lines except for BJ-hTert have a stronger up-regulation at 6 hr than at 3 hr.

Functional categories

After identifying the main co-expressed gene groups, we used the TANGO tool to ascribe to them biological significance. TANGO applies tests aimed at identifying functional categories that are statistically enriched in gene clusters. No functions were enriched in clusters of genes up-regulated after IR in any of the cell lines.

Maximally up-regulated genes were identified for each cell line independently. For example, GDF15 was the maximally up-regulated gene in both BJ-hTert and U2OS. This gene regulates tissue differentiation and maintenance and was previously reported in a DNA damage context [132]. BTG2, also induced in these two cell lines, has antiproliferative properties. TNFRSF10A,

induced in BJ-hTert, induces cell apoptosis. TP53INP1 was among the five maximally responding genes of four of the cell lines (all except BJ-hTert). In response to double-strand DNA breaks, the encoded protein is induced in a p53-dependent manner, induces cell cycle arrest in G1 and enhances p53-mediated apoptosis. FUCA1 was maximally up-regulated in HepG2 and TK6 cells. When this protein is defective, it leads to a disease with clinical features such as neurologic deterioration, which is a common feature of several genetic disorders caused by defects in DNA damage responses.

In contrast, all cell lines had significant enrichment of functions in the down-regulated gene clusters. Importantly, regulation of cell cycle, particularly mitosis, was the most widely enriched category in all cell lines, pointing to the activation of cell cycle checkpoints.

Inspection of the genes that were maximally down-regulated in each cell line showed that most of them were cell cycle genes, in accordance with the computational ontology. All cell lines except for BJ-hTert have kinesin-like motor proteins among their five most repressed genes: CENPE in G361, HepG2 and TK6, KIF14 in TK6 and U2OS. These proteins are vital to spatial dynamics of chromosomes during mitosis. PLK1, strongly repressed in G361 and U2OS cells, is required for cell division. Jang et al. (2007) proposed a novel pathway, that links the ATM/ATR/Chk axis and the protein kinase Plk1 in DNA damage response in mitosis [138].

Some of the repressed genes are involved in the regulatory mechanism of the induced genes. For example: DTL, down-regulated in BJ-hTert, was reported to play a role in regulating p53 stability [136]. PSRC1, the maximally repressed gene in TK6 cells, may participate in p53-mediated growth suppression [173].

We examined the order of magnitude of the maximal fold changes in the various cell lines, in both directions. Maximal fold changes were achieved in TK6 cells, while the lowest fold changes were observed in G361 cells.

Transcriptional regulators

Using the PRIMA algorithm, we searched the promoters of the genes in the major clusters for over-represented cis-regulatory elements. PRIMA revealed highly significant enrichment for p53 binding signature in the up-regulated genes in all cell lines except G361 (see further discussion in section 5.2.3). NF-Y signature was identified in the down-regulated genes of all cell lines. These results suggest that these two transcription factors are major regulators in the response to IR in the studied cell lines.

The activation and stabilization of the p53 protein play a major role in the DNA damage response and are mediated by ATM-dependent posttranslational modifications of p53 and Mdm2, a ubiquitin ligase of p53. ATM directly phosphorylates p53 as well as its inhibitor and E3 ubiquitin ligase, Mdm2, and activates also the checkpoint kinase Chk2, which in turn phosphorylates p53 on yet another site [4]. A new ATM target – the p53 inhibitor Mdmx – has been recently identified in the p53 control loop. Mdmx undergoes several ATM- and Chk2-dependent phosphorylations that mediate its proteosomal degradation [1, 174, 175].

All these ATM-dependent modifications contribute to the stabilization and rapid accumulation of p53 in response to IR-induced DNA damage.

NF-Y is known to be involved in many different aspects of cellular life, and has a role in the regulation of the cell cycle. NF-Y was demonstrated to control the expression of several key regulators of the G2/M phases of the cell cycle [122-125], in agreement with the enrichment of cell cycle functions which we found in the clusters regulated by NF-Y.

5.2.3 Integrative analysis of the five cell lines

We tried to identify the genes that respond purely to IR, and filtered out genes that responded to the side effect of the experimental manipulation. We conducted an analysis of the five cell lines together, to identify the 'universal' response to IR, and on the other hand, to pinpoint the individual differences between the cell lines.

Analysis of genes responding to IR in more than one of the cell line

The genes that showed common response in all cell lines are of particular interest. There were 13 up-regulated genes (**Table 4.2.3-2**), involved in apoptosis, cell cycle checkpoints, ROS protection and DNA repair. TANGO revealed the functional category "negative regulation of cell proliferation" as significantly overrepresented this group of genes. Significantly, 10 out of the 13 induced genes are known to be p53-regulated, and indeed PRIMA revealed highly significant enrichment for p53 binding site signature in this cluster. We suspect that the remaining three – 230330_at (Affymetrix probe ID), hypothetical protein MGC5370 and CPM – may be potential targets of p53 as well.

In the previous section (5.2.2), we noted that p53 signature was not significantly noticed in the up-regulated cluster of G361 cells. On the other hand, it showed significant appearance in the group of 13 genes up-regulated in all five cell lines, including G361. We would have expected p53 to be revealed in the separate analysis of G361. A second look at the up-regulated genes in G361 (**Fig. 4.2.2-2**) reveals that these genes are considerably up-regulated in the chronological controls (3 and 6 hr after treatment), which means that the cluster is 'noisy'. The filtering of genes in the analysis of each cell line separately did not include specifically the information of the chronological controls, in contrast to the integrative analysis of the cell lines. Since most of the p53 targets did not show up in the 'responding gene set' of G361, it seems that the separate analysis that did not include the chronological controls, was not as clean as the integrative analysis.

One of the universally induced genes was DDB2, which functions in nucleotide excision repair (NER). However, the damage in this study was DSBs. Despras et al. (2007) reported a relationship between DSB repair and NER. Cells in which the XPC DNA repair gene was stably silenced showed increased sensitivity to etoposide, a topoisomerase II inhibitor that creates DSBs [176]. They suggest that XPC deficiency may contribute to impaired DSB repair.

The genes encoding MDM2 and PPM1D, which are involved in inhibition of p53-mediated apoptosis [177, 178], were induced side by side with the apoptotic genes FAS, FDXR and TP53INP1 (**Table 4.2.3-2**). It is interesting to investigate the interactions between these groups of genes, regulated by p53. MDM2 binds p53 and inhibits its transactivation activity and serves also as its E3 ubiquitin ligase. The p53-inducible phosphatase PPM1D negatively regulates the activity of p38 MAPK, through which it reduces the phosphorylation of p53, and in turn suppresses p53-mediated transcription and apoptosis, creating a potential feedback loop [178]. The protein-protein map presented on **Fig. 4.2.3-2** shows that PPM1D inhibits ATM and CHEK2, which inhibit MDM2. PPM1D seems to be critical for regulating the ATM-mediated tumor surveillance network [179]. E2F1, which plays a crucial role in the control of cell cycle and action of tumor suppressor proteins, induces PPM1D and TP53INP1 [180, 181], mediating both cell proliferation and p53-dependent/independent apoptosis. Notably, there is a complicated network of interactions between these genes.

Fifteen genes were down-regulated in all cell lines studied (**Table 4.2.3-3**). Most of these genes facilitate transit through the cell cycle, in agreement with the detection of the signature of NF-Y binding site in this group of genes. Pawlik et al. (2004) suggested that cell cycle regulation might be the most important determinant of sensitivity to ionizing radiation. DNA damage induced by ionizing radiation initiates signals that can ultimately activate either the cell cycle checkpoints that permit time for damage processing or irreversible growth arrest that leads to cell death. Pawlik et al. (2004) noted that the cell cycle phase at the time of damage initiation also determines a cell's relative radiosensitivity: cells are most radiosensitive in G2-M, less sensitive in G1, and least sensitive during the later part of the S phase [182].

KIAA1333 and TROAP, down-regulated in all cell lines studied, did not seem to be involved in cell cycle control at first sight, but a second look revealed evidence for their involvement. Crawford et al. (2001) reported that both KIAA1333 and TROAP genes have G2/M-specific expression, and are down-regulated following DNA damage [183]. Kho et al. (2004) assessed

the transcriptional response of HCT116 colorectal cancer cells during apoptosis induced by the anticancer drug 5-Fluorouracil (5-FU, an antimetabolite that activates p53) as a function of p53 status and identified the down-regulated TROAP gene as a potential target of p53 [184].

An analysis of the genes that were found responding in 4, 3 or 2 of the cell lines did not add much new information, but supported previous results: the down-regulated genes were cell cycle related, and the two dominant transcription factors were p53 (for up-regulated genes) and NF-Y (for down-regulated genes). An additional TF whose binding site was enriched in genes up-regulated in 4 or 3 cell lines was TBP (TATA-box binding protein). TBP is a component of TFIID, which coordinates the activities of more than 70 polypeptides that are required for the initiation of transcription by RNA polymerase II [185]. Wild-type p53 binds directly to TBP and interferes with transcriptional initiation [186]. Mutations in the polyglutamine string of TBP are associated with spinocerebellar ataxia 17, a neurodegenerative disorder classified as a polyglutamine disease [187, 188].

Genes that respond to IR uniquely in one cell line

We wanted to identify differences between IR responses that are cell line-specific. To this end, we looked closely at the probesets that responded significantly in one cell line and not in the others. TK6 had the largest number of uniquely responding genes, while G361 had the fewest (**Table 4.2.3-6**).

We noted that four histone genes were uniquely up-regulated in BJ-hTert cell line. The connection between histone dynamics and the DNA damage responses is well documented [189-192]. A prominent damage-induced histone modification is the phosphorylation of γ H2AX in nucleosomes flanking DSB sites. Phosphorylated H2AX anchors the damage response proteins that are recruited to these sites such as MDC1, 53BP1 and BRCA1 [193]. Methylated lysine residues in histone proteins play a major role in 53BP1 recruitment to the damaged sites [194].

The genes that were down-regulated uniquely in BJ-hTert were enriched with cell cycle functional categories. Since down-regulation of cell cycle progression should be common to all

cell lines following DNA damage, we expected this phenomenon in all cell lines and not just one. RAD51 and BRCA1 were among the genes down-regulated uniquely in BJ-hTert cells. BRCA1 is involved in the early stage of the DNA damage response in an unclear mechanism [195-197]. RAD51 interacts with the ssDNA-binding protein RPA and RAD52 [198, 199], and mediates the homologous recombination pathway of double strand breaks [200, 201].

The unique induction of NFkB2 in HepG2 is interesting, since it is related to the anti-apoptotic response. COP1, an inhibitor of apoptosis [202, 203], is up-regulated only in TK6.

Surprisingly, we did not find any common denominator among the cancerous cell lines that distinguishes them from the non-cancerous cell lines. A possible explanation is the abnormal nature of "non-cancerous" cultures, which are not equivalent to normal tissues [170]. The two 'normal' cell lines in this study, BJ-hTert and TK6, are immortalized. BJ-hTert fibroblast line was immortalized by expression of human telomerase gene [100]. Glaviano et al. (2006) suggest that the type of genomic instability in human cells may depend critically on whether they are telomerase-positive or negative (hTERT+ or hTERT-) [204]. TK6 is a human B lymphoblastoid cell line obtained by spontaneous immortalization [205], which shows chromosomal instability including altered telomere maintenance [206-208]. In general, cell lines derived from tumor tissues are suspected for increased genomic instability [169].

5.3 Future prospects

In this work, we applied gene expression microarrays combined with computational tools to delineate transcriptional responses induced by IR in five human cell lines, and to analyze the differences between them.

The differences in the transcriptional response of cell lines to DNA damage show that it is crucial to compare results of several experimental models and avoid reaching global conclusions based on a study of a single cell line or tissue.

Our results emphasize the importance of the transcription factors p53 and NF-Y as key regulators of these responses. Potential new targets of p53 have been suggested in this study.

It will be interesting to probe the transcriptional response induced by DNA damage in cells knocked-down for p53, NF-Y and both of them. Further studies in our lab are following this direction. This should allow further elucidation of functional links among key players in the DNA damage response network, another step on the way to understand how our cells cope with one of the major threats to cellular homeostasis – DNA damage.

References

1. Shiloh, Y., *The ATM-mediated DNA-damage response: taking shape*. Trends Biochem Sci, 2006. **31**(7): p. 402-10.
2. Wyman, C. and R. Kanaar, *DNA double-strand break repair: all's well that ends well*. Annu Rev Genet, 2006. **40**: p. 363-83.
3. Norbury, C.J. and I.D. Hickson, *Cellular responses to DNA damage*. Annu Rev Pharmacol Toxicol, 2001. **41**: p. 367-401.
4. Shiloh, Y., *ATM and related protein kinases: safeguarding genome integrity*. Nat Rev Cancer, 2003. **3**(3): p. 155-68.
5. Coultas, L. and A. Strasser, *The molecular control of DNA damage-induced cell death*. Apoptosis, 2000. **5**(6): p. 491-507.
6. Shiloh, Y. and M.B. Kastan, *ATM: genome stability, neuronal development, and cancer cross paths*. Adv Cancer Res, 2001. **83**: p. 209-54.
7. Kaneko, H. and N. Kondo, *Clinical features of Bloom syndrome and function of the causative gene, BLM helicase*. Expert Rev Mol Diagn, 2004. **4**(3): p. 393-401.
8. Magnaldo, T. and A. Sarasin, *Xeroderma pigmentosum: from symptoms and genetics to gene-based skin therapy*. Cells Tissues Organs, 2004. **177**(3): p. 189-98.
9. Charames, G.S. and B. Bapat, *Genomic instability and cancer*. Curr Mol Med, 2003. **3**(7): p. 589-96.
10. Jo, W.S. and D.C. Chung, *Genetics of hereditary colorectal cancer*. Semin Oncol, 2005. **32**(1): p. 11-23.
11. Chun, H.H. and R.A. Gatti, *Ataxia-telangiectasia, an evolving phenotype*. DNA Repair (Amst), 2004. **3**(8-9): p. 1187-96.
12. Digweed, M. and K. Sperling, *Nijmegen breakage syndrome: clinical manifestation of defective response to DNA double-strand breaks*. DNA Repair (Amst), 2004. **3**(8-9): p. 1207-17.
13. Abner, C.W. and P.J. McKinnon, *The DNA double-strand break response in the nervous system*. DNA Repair (Amst), 2004. **3**(8-9): p. 1141-7.
14. Caldecott, K.W., *DNA single-strand breaks and neurodegeneration*. DNA Repair (Amst), 2004. **3**(8-9): p. 875-82.
15. Uberti, D., G. Ferrari Toninelli, and M. Memo, *Involvement of DNA damage and repair systems in neurodegenerative process*. Toxicol Lett, 2003. **139**(2-3): p. 99-105.
16. Laiho, M. and L. Latonen, *Cell cycle control, DNA damage checkpoints and cancer*. Ann Med, 2003. **35**(6): p. 391-7.
17. Shimada, M. and M. Nakanishi, *DNA damage checkpoints and cancer*. J Mol Histol, 2006. **37**(5-7): p. 253-60.
18. Jelinsky, S.A., et al., *Regulatory networks revealed by transcriptional profiling of damaged Saccharomyces cerevisiae cells: Rpn4 links base excision repair with proteasomes*. Mol Cell Biol, 2000. **20**(21): p. 8157-67.
19. Jelinsky, S.A. and L.D. Samson, *Global response of Saccharomyces cerevisiae to an alkylating agent*. Proc Natl Acad Sci U S A, 1999. **96**(4): p. 1486-91.
20. Fornace, A.J., Jr., et al., *The complexity of radiation stress responses: analysis by informatics and functional genomics approaches*. Gene Expr, 1999. **7**(4-6): p. 387-400.
21. Koch-Paiz, C.A., et al., *Functional genomics of UV radiation responses in human cells*. Mutat Res, 2004. **549**(1-2): p. 65-78.
22. Amundson, S.A., M. Bittner, and A.J. Fornace, Jr., *Functional genomics as a window on radiation stress signaling*. Oncogene, 2003. **22**(37): p. 5828-33.
23. Gudkov, A.V. and E.A. Komarova, *The role of p53 in determining sensitivity to radiotherapy*. Nat Rev Cancer, 2003. **3**(2): p. 117-29.

24. Komarova, E.A., et al., *Stress-induced secretion of growth inhibitors: a novel tumor suppressor function of p53*. *Oncogene*, 1998. **17**(9): p. 1089-96.
25. Burns, T.F., E.J. Bernhard, and W.S. El-Deiry, *Tissue specific expression of p53 target genes suggests a key role for KILLER/DR5 in p53-dependent apoptosis in vivo*. *Oncogene*, 2001. **20**(34): p. 4601-12.
26. Rosen, E.M., et al., *Biological basis of radiation sensitivity. Part 2: Cellular and molecular determinants of radiosensitivity*. *Oncology (Huntingt)*, 2000. **14**(5): p. 741-57; discussion 757-8, 761-6.
27. Rosen, E.M., et al., *Biological basis of radiation sensitivity. Part 1: Factors governing radiation tolerance*. *Oncology (Huntingt)*, 2000. **14**(4): p. 543-50.
28. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage*. *Science*, 2007. **316**(5828): p. 1160-6.
29. Mu, J.J., et al., *A proteomic analysis of ataxia telangiectasia-mutated (ATM)/ATM-Rad3-related (ATR) substrates identifies the ubiquitin-proteasome system as a regulator for DNA damage checkpoints*. *J Biol Chem*, 2007. **282**(24): p. 17330-4.
30. Kurz, E.U. and S.P. Lees-Miller, *DNA damage-induced activation of ATM and ATM-dependent signaling pathways*. *DNA Repair (Amst)*, 2004. **3**(8-9): p. 889-900.
31. Pereg, Y., et al., *Differential roles of ATM- and Chk2-mediated phosphorylations of Hdmx in response to DNA damage*. *Mol Cell Biol*, 2006. **26**(18): p. 6819-31.
32. Lavin, M.F. and N. Gueven, *The complexity of p53 stabilization and activation*. *Cell Death Differ*, 2006. **13**(6): p. 941-50.
33. Wade, M., et al., *Hdmx modulates the outcome of p53 activation in human tumor cells*. *J Biol Chem*, 2006. **281**(44): p. 33036-44.
34. Kawai, H., et al., *RING domain-mediated interaction is a requirement for MDM2's E3 ligase activity*. *Cancer Res*, 2007. **67**(13): p. 6026-30.
35. Cortez, D., et al., *Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks*. *Science*, 1999. **286**(5442): p. 1162-6.
36. Li, S., et al., *Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response*. *Nature*, 2000. **406**(6792): p. 210-5.
37. Lee, J.H. and T.T. Paull, *ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex*. *Science*, 2005. **308**(5721): p. 551-4.
38. Uziel, T., et al., *Requirement of the MRN complex for ATM activation by DNA damage*. *Embo J*, 2003. **22**(20): p. 5612-21.
39. Abraham, R.T. and R.S. Tibbetts, *Cell biology. Guiding ATM to broken DNA*. *Science*, 2005. **308**(5721): p. 510-1.
40. Oka, A. and S. Takashima, *Expression of the ataxia-telangiectasia gene (ATM) product in human cerebellar neurons during development*. *Neurosci Lett*, 1998. **252**(3): p. 195-8.
41. Barlow, C., et al., *ATM is a cytoplasmic protein in mouse brain required to prevent lysosomal accumulation*. *Proc Natl Acad Sci U S A*, 2000. **97**(2): p. 871-6.
42. Stewart, G.S., et al., *The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder*. *Cell*, 1999. **99**(6): p. 577-87.
43. Frappart, P.O., et al., *An essential function for NBS1 in the prevention of ataxia and cerebellar defects*. *Nat Med*, 2005. **11**(5): p. 538-44.
44. Aggarwal, K. and K.H. Lee, *Functional genomics and proteomics as a foundation for systems biology*. *Brief Funct Genomic Proteomic*, 2003. **2**(3): p. 175-84.
45. Ehrenberg, M., et al., *Systems biology is taking off*. *Genome Res*, 2003. **13**(11): p. 2377-80.
46. Ideker, T., *Systems biology 101--what you need to know*. *Nat Biotechnol*, 2004. **22**(4): p. 473-5.
47. Ideker, T., T. Galitski, and L. Hood, *A new approach to decoding life: systems biology*. *Annu Rev Genomics Hum Genet*, 2001. **2**: p. 343-72.
48. Kitano, H., *Computational systems biology*. *Nature*, 2002. **420**(6912): p. 206-10.

49. Kitano, H., *Systems biology: a brief overview*. Science, 2002. **295**(5560): p. 1662-4.
50. Hood, L., et al., *Systems biology and new technologies enable predictive and preventative medicine*. Science, 2004. **306**(5696): p. 640-3.
51. Lockhart, D.J. and E.A. Winzler, *Genomics, gene expression and DNA arrays*. Nature, 2000. **405**(6788): p. 827-36.
52. Lander, E.S. and R.A. Weinberg, *Genomics: journey to the center of biology*. Science, 2000. **287**(5459): p. 1777-82.
53. Kennedy, G.C., et al., *Large-scale genotyping of complex DNA*. Nat Biotechnol, 2003. **21**(10): p. 1233-7.
54. Middleton, F.A., et al., *Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22*. Am J Hum Genet, 2004. **74**(5): p. 886-97.
55. Mitra, N., et al., *Localization of cancer susceptibility genes by genome-wide single-nucleotide polymorphism linkage-disequilibrium mapping*. Cancer Res, 2004. **64**(21): p. 8116-25.
56. Matsuzaki, H., et al., *Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays*. Nat Methods, 2004. **1**(2): p. 109-11.
57. Mantripragada, K.K., et al., *Genomic microarrays in the spotlight*. Trends Genet, 2004. **20**(2): p. 87-94.
58. Snijders, A.M., D. Pinkel, and D.G. Albertson, *Current status and future prospects of array-based comparative genomic hybridisation*. Brief Funct Genomic Proteomic, 2003. **2**(1): p. 37-45.
59. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. **270**(5235): p. 467-70.
60. Duggan, D.J., et al., *Expression profiling using cDNA microarrays*. Nat Genet, 1999. **21**(1 Suppl): p. 10-4.
61. Lipshutz, R.J., et al., *High density synthetic oligonucleotide arrays*. Nat Genet, 1999. **21**(1 Suppl): p. 20-4.
62. Velculescu, V.E., B. Vogelstein, and K.W. Kinzler, *Analysing uncharted transcriptomes with SAGE*. Trends Genet, 2000. **16**(10): p. 423-5.
63. Tamayo, P., et al., *Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation*. Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2907-12.
64. Whitfield, M.L., et al., *Identification of genes periodically expressed in the human cell cycle and their expression in tumors*. Mol Biol Cell, 2002. **13**(6): p. 1977-2000.
65. Lee, C.K., et al., *Gene expression profile of aging and its retardation by caloric restriction*. Science, 1999. **285**(5432): p. 1390-3.
66. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. Nature, 2000. **403**(6769): p. 503-11.
67. Lee, T.I., et al., *Transcriptional regulatory networks in Saccharomyces cerevisiae*. Science, 2002. **298**(5594): p. 799-804.
68. Ren, B., et al., *Genome-wide location and function of DNA binding proteins*. Science, 2000. **290**(5500): p. 2306-9.
69. Ren, B., et al., *E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints*. Genes Dev, 2002. **16**(2): p. 245-56.
70. Martone, R., et al., *Distribution of NF-kappaB-binding sites across human chromosome 22*. Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12247-52.
71. Li, Z., et al., *A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells*. Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8164-9.
72. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. Nature, 2003. **422**(6928): p. 198-207.

73. Ho, Y., et al., *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry*. Nature, 2002. **415**(6868): p. 180-3.
74. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes*. Nature, 2002. **415**(6868): p. 141-7.
75. Forler, D., et al., *An efficient protein complex purification method for functional proteomics in higher eukaryotes*. Nat Biotechnol, 2003. **21**(1): p. 89-92.
76. Mann, M., et al., *Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome*. Trends Biotechnol, 2002. **20**(6): p. 261-8.
77. Sreekumar, A., et al., *Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins*. Cancer Res, 2001. **61**(20): p. 7585-93.
78. Zhu, H., et al., *Global analysis of protein activities using proteome chips*. Science, 2001. **293**(5537): p. 2101-5.
79. Hannon, G.J., *RNA interference*. Nature, 2002. **418**(6894): p. 244-51.
80. Hannon, G.J. and J.J. Rossi, *Unlocking the potential of the human genome with RNA interference*. Nature, 2004. **431**(7006): p. 371-8.
81. Brummelkamp, T.R., R. Bernards, and R. Agami, *A system for stable expression of short interfering RNAs in mammalian cells*. Science, 2002. **296**(5567): p. 550-3.
82. Brummelkamp, T.R., R. Bernards, and R. Agami, *Stable suppression of tumorigenicity by virus-mediated RNA interference*. Cancer Cell, 2002. **2**(3): p. 243-7.
83. Kamath, R.S., et al., *Systematic functional analysis of the Caenorhabditis elegans genome using RNAi*. Nature, 2003. **421**(6920): p. 231-7.
84. Paddison, P.J., et al., *A resource for large-scale RNA-interference-based screens in mammals*. Nature, 2004. **428**(6981): p. 427-31.
85. Berns, K., et al., *A large-scale RNAi screen in human cells identifies new components of the p53 pathway*. Nature, 2004. **428**(6981): p. 431-7.
86. Knight, J., *When the chips are down*. Nature, 2001. **410**(6831): p. 860-1.
87. Kawasaki, E.S., *The end of the microarray Tower of Babel: will universal standards lead the way?* J Biomol Tech, 2006. **17**(3): p. 200-6.
88. Cheung, V.G., et al., *Making and reading microarrays*. Nat Genet, 1999. **21**(1 Suppl): p. 15-9.
89. Wodicka, L., et al., *Genome-wide expression monitoring in Saccharomyces cerevisiae*. Nat Biotechnol, 1997. **15**(13): p. 1359-67.
90. Naef, F., et al., *DNA hybridization to mismatched templates: a chip study*. Phys Rev E Stat Nonlin Soft Matter Phys, 2002. **65**(4 Pt 1): p. 040902.
91. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data*. Biostatistics, 2003. **4**(2): p. 249-64.
92. Hughes, T.R., et al., *Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer*. Nat Biotechnol, 2001. **19**(4): p. 342-7.
93. Barczak, A., et al., *Spotted long oligonucleotide arrays for human gene expression analysis*. Genome Res, 2003. **13**(7): p. 1775-85.
94. Woo, Y., et al., *A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms*. J Biomol Tech, 2004. **15**(4): p. 276-84.
95. Park, P.J., et al., *Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference*. J Biotechnol, 2004. **112**(3): p. 225-45.
96. Bammler, T., et al., *Standardizing global gene expression analysis between laboratories and across platforms*. Nat Methods, 2005. **2**(5): p. 351-6.
97. Irizarry, R.A., et al., *Multiple-laboratory comparison of microarray platforms*. Nat Methods, 2005. **2**(5): p. 345-50.
98. Larkin, J.E., et al., *Independence and reproducibility across microarray platforms*. Nat Methods, 2005. **2**(5): p. 337-44.
99. Wang, Y., et al., *Inferring gene regulatory networks from multiple microarray datasets*. Bioinformatics, 2006. **22**(19): p. 2413-20.

100. Bodnar, A.G., et al., *Extension of life-span by introduction of telomerase into normal human cells*. Science, 1998. **279**(5349): p. 349-52.
101. Dedon, P.C. and I.H. Goldberg, *Free-radical mechanisms involved in the formation of sequence-dependent bistranded DNA lesions by the antitumor antibiotics bleomycin, neocarzinostatin, and calicheamicin*. Chem Res Toxicol, 1992. **5**(3): p. 311-32.
102. Povirk, L.F., *DNA damage and mutagenesis by radiomimetic DNA-cleaving agents: bleomycin, neocarzinostatin and other enediynes*. Mutat Res, 1996. **355**(1-2): p. 71-89.
103. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction*. Anal Biochem, 1987. **162**(1): p. 156-9.
104. Shamir, R., et al., *EXPANDER--an integrative program suite for microarray data analysis*. BMC Bioinformatics, 2005. **6**: p. 232.
105. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. Bioinformatics, 2003. **19**(2): p. 185-93.
106. Li, C. and W.H. Wong, *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection*. Proc Natl Acad Sci U S A, 2001. **98**(1): p. 31-6.
107. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics*. Genome Biol, 2004. **5**(10): p. R80.
108. Yang, Y.H., et al., *Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation*. Nucleic Acids Res, 2002. **30**(4): p. e15.
109. Elkon, R., et al., *Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis*. Genome Biol, 2005. **6**(5): p. R43.
110. Wei, C., J. Li, and R.E. Bumgarner, *Sample size for detecting differentially expressed genes in microarray experiments*. BMC Genomics, 2004. **5**(1): p. 87.
111. Tavazoie, S., et al., *Systematic determination of genetic network architecture*. Nat Genet, 1999. **22**(3): p. 281-5.
112. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
113. Sharan, R. and R. Shamir, *CLICK: a clustering algorithm with applications to gene expression analysis*. Proc Int Conf Intell Syst Mol Biol, 2000. **8**: p. 307-16.
114. Sharan, R., R. Elkon, and R. Shamir, *Cluster analysis and its applications to gene expression data*. Ernst Schering Res Found Workshop, 2002(38): p. 83-108.
115. Tanay, A., et al., *Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data*. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2981-6.
116. Tanay, A., R. Sharan, and R. Shamir, *Discovering statistically significant biclusters in gene expression data*. Bioinformatics, 2002. **18 Suppl 1**: p. S136-44.
117. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*. Nat Genet, 2000. **25**(1): p. 25-9.
118. Elkon, R., et al., *Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells*. Genome Res, 2003. **13**(5): p. 773-80.
119. Dobbin, K.K., et al., *Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays*. Clin Cancer Res, 2005. **11**(2 Pt 1): p. 565-72.
120. Yauk, C.L., et al., *Comprehensive comparison of six microarray technologies*. Nucleic Acids Res, 2004. **32**(15): p. e124.
121. Fei, P. and W.S. El-Deiry, *P53 and radiation responses*. Oncogene, 2003. **22**(37): p. 5774-83.

122. Manni, I., et al., *NF-Y mediates the transcriptional inhibition of the cyclin B1, cyclin B2, and cdc25C promoters upon induced G2 arrest.* J Biol Chem, 2001. **276**(8): p. 5570-6.
123. Jung, M.S., et al., *p53 and its homologues, p63 and p73, induce a replicative senescence through inactivation of NF-Y transcription factor.* Oncogene, 2001. **20**(41): p. 5818-25.
124. Yun, J., et al., *p53 negatively regulates cdc2 transcription via the CCAAT-binding NF-Y transcription factor.* J Biol Chem, 1999. **274**(42): p. 29677-82.
125. Imbriano, C., et al., *Direct p53 transcriptional repression: in vivo analysis of CCAAT-containing G2/M promoters.* Mol Cell Biol, 2005. **25**(9): p. 3737-51.
126. Nelson, S.B., A.E. Schaffer, and M. Sander, *The transcription factors Nkx6.1 and Nkx6.2 possess equivalent activities in promoting beta-cell fate specification in Pdx1+ pancreatic progenitor cells.* Development, 2007. **134**(13): p. 2491-500.
127. Henseleit, K.D., et al., *NKX6 transcription factor activity is required for alpha- and beta-cell development in the pancreas.* Development, 2005. **132**(13): p. 3139-49.
128. Pattyn, A., et al., *Complementary roles for Nkx6 and Nkx2 class proteins in the establishment of motoneuron identity in the hindbrain.* Development, 2003. **130**(17): p. 4149-59.
129. Hatfield, G.W., S.P. Hung, and P. Baldi, *Differential analysis of DNA microarray gene expression data.* Mol Microbiol, 2003. **47**(4): p. 871-7.
130. Lee, J.H. and T.T. Paull, *Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex.* Science, 2004. **304**(5667): p. 93-6.
131. Bakkenist, C.J. and M.B. Kastan, *DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation.* Nature, 2003. **421**(6922): p. 499-506.
132. Kis, E., et al., *Microarray analysis of radiation response genes in primary human fibroblasts.* Int J Radiat Oncol Biol Phys, 2006. **66**(5): p. 1506-14.
133. Huang, X.P., et al., *Negative implication of C-MYC as an amplification target in esophageal cancer.* Cancer Genet Cytogenet, 2006. **165**(1): p. 20-4.
134. Gevaert, O., et al., *Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks.* Bioinformatics, 2006. **22**(14): p. e184-90.
135. Pan, H.W., et al., *Role of L2DTL, cell cycle-regulated nuclear and centrosome protein, in aggressive hepatocellular carcinoma.* Cell Cycle, 2006. **5**(22): p. 2676-87.
136. Banks, D., et al., *L2DTL/CDT2 and PCNA interact with p53 and regulate p53 polyubiquitination and protein stability through MDM2 and CUL4A/DDB1 complexes.* Cell Cycle, 2006. **5**(15): p. 1719-29.
137. Zdobnov, E.M. and R. Apweiler, *InterProScan--an integration platform for the signature-recognition methods in InterPro.* Bioinformatics, 2001. **17**(9): p. 847-8.
138. Jang, Y.J., et al., *Regulation of Polo-like kinase 1 by DNA damage in mitosis. Inhibition of mitotic PLK-1 by protein phosphatase 2A.* J Biol Chem, 2007. **282**(4): p. 2473-82.
139. Mateyak, M.K. and V.A. Zakian, *Human PIF helicase is cell cycle regulated and associates with telomerase.* Cell Cycle, 2006. **5**(23): p. 2796-804.
140. Zhang, D.H., et al., *The human Pif1 helicase, a potential Escherichia coli RecD homologue, inhibits telomerase activity.* Nucleic Acids Res, 2006. **34**(5): p. 1393-404.
141. Nousiainen, M., et al., *Phosphoproteome analysis of the human mitotic spindle.* Proc Natl Acad Sci U S A, 2006. **103**(14): p. 5391-6.
142. Mori, M., et al., *Transcriptional response to ionizing radiation in lymphocyte subsets.* Cell Mol Life Sci, 2005. **62**(13): p. 1489-501.
143. Carleton, M., et al., *RNA interference-mediated silencing of mitotic kinesin KIF14 disrupts cell cycle progression and induces cytokinesis failure.* Mol Cell Biol, 2006. **26**(10): p. 3853-63.
144. Corson, T.W. and B.L. Gallie, *KIF14 mRNA expression is a predictor of grade and outcome in breast cancer.* Int J Cancer, 2006. **119**(5): p. 1088-94.
145. Castresana, J. and M. Saraste, *Does Vav bind to F-actin through a CH domain?* FEBS Lett, 1995. **374**(2): p. 149-51.

146. Gimona, M. and R. Mital, *The single CH domain of calponin is neither sufficient nor necessary for F-actin binding*. J Cell Sci, 1998. **111 (Pt 13)**: p. 1813-21.
147. Liu, G. and X. Chen, *The ferredoxin reductase gene is regulated by the p53 family and sensitizes cells to oxidative stress-induced apoptosis*. Oncogene, 2002. **21(47)**: p. 7195-204.
148. Bond, J., et al., *ASPM is a major determinant of cerebral cortical size*. Nat Genet, 2002. **32(2)**: p. 316-20.
149. Burchett, S.A., *Regulators of G protein signaling: a bestiary of modular protein binding domains*. J Neurochem, 2000. **75(4)**: p. 1335-51.
150. Brooks, W.S., S. Banerjee, and D.F. Crawford, *G2E3 is a nucleo-cytoplasmic shuttling protein with DNA damage responsive localization*. Exp Cell Res, 2007. **313(4)**: p. 665-76.
151. Huang, H., et al., *Tripin/hSgo2 recruits MCAK to the inner centromere to correct defective kinetochore attachments*. J Cell Biol, 2007. **177(3)**: p. 413-24.
152. Nadano, D., et al., *Human tastin, a proline-rich cytoplasmic protein, associates with the microtubular cytoskeleton*. Biochem J, 2002. **364(Pt 3)**: p. 669-77.
153. Chen, F.E. and G. Ghosh, *Regulation of DNA binding by Rel/NF-kappaB transcription factors: structural views*. Oncogene, 1999. **18(49)**: p. 6845-52.
154. Baldwin, A.S., Jr., *The NF-kappa B and I kappa B proteins: new discoveries and insights*. Annu Rev Immunol, 1996. **14**: p. 649-83.
155. Bernal-Mizrachi, L., C.M. Lovly, and L. Ratner, *The role of NF-{kappa}B-1 and NF-{kappa}B-2-mediated resistance to apoptosis in lymphomas*. Proc Natl Acad Sci U S A, 2006. **103(24)**: p. 9220-5.
156. Gugliesi, F., et al., *Up-regulation of the interferon-inducible IFI16 gene by oxidative stress triggers p53 transcriptional activity in endothelial cells*. J Leukoc Biol, 2005. **77(5)**: p. 820-9.
157. Kwak, J.C., et al., *IFI16 as a negative regulator in the regulation of p53 and p21(Waf1)*. J Biol Chem, 2003. **278(42)**: p. 40899-904.
158. Fujiuchi, N., et al., *Requirement of IFI16 for the maximal activation of p53 induced by ionizing radiation*. J Biol Chem, 2004. **279(19)**: p. 20339-44.
159. Araki, M., et al., *Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair*. J Biol Chem, 2001. **276(22)**: p. 18665-72.
160. Masutani, C., et al., *Identification and characterization of XPC-binding domain of hHR23B*. Mol Cell Biol, 1997. **17(12)**: p. 6915-23.
161. Takemura, H., et al., *Defective Mre11-dependent activation of Chk2 by ataxia telangiectasia mutated in colorectal carcinoma cells in response to replication-dependent DNA double strand breaks*. J Biol Chem, 2006. **281(41)**: p. 30814-23.
162. Dolganov, G.M., et al., *Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair*. Mol Cell Biol, 1996. **16(9)**: p. 4832-41.
163. Bender, C.F., et al., *Cancer predisposition and hematopoietic failure in Rad50(S/S) mice*. Genes Dev, 2002. **16(17)**: p. 2237-51.
164. Huang, Z., H. Xu, and L. Sandell, *Negative regulation of chondrocyte differentiation by transcription factor AP-2alpha*. J Bone Miner Res, 2004. **19(2)**: p. 245-55.
165. O'Brien, E.K., et al., *Transcription factor Ap-2alpha is necessary for development of embryonic melanophores, autonomic neurons and pharyngeal skeleton in zebrafish*. Dev Biol, 2004. **265(1)**: p. 246-61.
166. Mitsunari, T., et al., *Clathrin adaptor AP-2 is essential for early embryonal development*. Mol Cell Biol, 2005. **25(21)**: p. 9318-23.
167. Li, H., et al., *AP-2alpha and AP-2gamma are transcriptional targets of p53 in human breast carcinoma cells*. Oncogene, 2006. **25(39)**: p. 5405-15.

168. Mothersill, C. and C. Seymour, *Radiation-induced bystander and other non-targeted effects: novel intervention points in cancer therapy?* *Curr Cancer Drug Targets*, 2006. **6**(5): p. 447-54.
169. Ross, D.T., et al., *Systematic variation in gene expression patterns in human cancer cell lines*. *Nat Genet*, 2000. **24**(3): p. 227-35.
170. Shi, Z.Z., et al., *Glutathione synthesis is essential for mouse development but not for cell growth in culture*. *Proc Natl Acad Sci U S A*, 2000. **97**(10): p. 5101-6.
171. Matsumura, Y., et al., *Increase in radiation sensitivity of human malignant melanoma cells by expression of wild-type p16 gene*. *Cancer Lett*, 1997. **115**(1): p. 91-6.
172. Yamagishi, N., et al., *Modification of the radiosensitivity of human cells to which simian virus 40 T-antigen was transfected*. *J Radiat Res (Tokyo)*, 1995. **36**(4): p. 239-47.
173. Lo, P.K. and F.F. Wang, *Cloning and characterization of human and mouse DDA3 genes*. *Biochim Biophys Acta*, 2002. **1579**(2-3): p. 214-8.
174. Pereg, Y., et al., *Phosphorylation of Hdmx mediates its Hdm2- and ATM-dependent degradation in response to DNA damage*. *Proc Natl Acad Sci U S A*, 2005. **102**(14): p. 5056-61.
175. Okamoto, K., et al., *DNA damage-induced phosphorylation of MdmX at serine 367 activates p53 by targeting MdmX for Mdm2-dependent degradation*. *Mol Cell Biol*, 2005. **25**(21): p. 9608-20.
176. Despras, E., et al., *Long-term XPC silencing reduces DNA double-strand break repair*. *Cancer Res*, 2007. **67**(6): p. 2526-34.
177. Haupt, Y., Y. Barak, and M. Oren, *Cell type-specific inhibition of p53-mediated apoptosis by mdm2*. *Embo J*, 1996. **15**(7): p. 1596-606.
178. Bulavin, D.V., et al., *Amplification of PPM1D in human tumors abrogates p53 tumor-suppressor activity*. *Nat Genet*, 2002. **31**(2): p. 210-5.
179. Shreeram, S., et al., *Regulation of ATM/p53-dependent suppression of myc-induced lymphomas by Wip1 phosphatase*. *J Exp Med*, 2006. **203**(13): p. 2793-9.
180. Hershko, T., et al., *E2F1 modulates p38 MAPK phosphorylation via transcriptional regulation of ASK1 and Wip1*. *J Biol Chem*, 2006. **281**(42): p. 31309-16.
181. Hershko, T., et al., *Novel link between E2F and p53: proapoptotic cofactors of p53 are transcriptionally upregulated by E2F*. *Cell Death Differ*, 2005. **12**(4): p. 377-83.
182. Pawlik, T.M. and K. Keyomarsi, *Role of cell cycle in mediating sensitivity to radiotherapy*. *Int J Radiat Oncol Biol Phys*, 2004. **59**(4): p. 928-42.
183. Crawford, D.F. and H. Piwnica-Worms, *The G(2) DNA damage checkpoint delays expression of genes encoding mitotic regulators*. *J Biol Chem*, 2001. **276**(40): p. 37166-77.
184. Kho, P.S., et al., *p53-regulated transcriptional program associated with genotoxic stress-induced apoptosis*. *J Biol Chem*, 2004. **279**(20): p. 21183-92.
185. Guermah, M., et al., *The TBN protein, which is essential for early embryonic mouse development, is an inducible TAFII implicated in adipogenesis*. *Mol Cell*, 2003. **12**(4): p. 991-1001.
186. Seto, E., et al., *Wild-type p53 binds to the TATA-binding protein and represses transcription*. *Proc Natl Acad Sci U S A*, 1992. **89**(24): p. 12028-32.
187. Nanda, A., et al., *Case of spinocerebellar ataxia type 17 (SCA17) associated with only 41 repeats of the TATA-binding protein (TBP) gene*. *Mov Disord*, 2007. **22**(3): p. 436.
188. Wu, Y.R., et al., *Genetic testing in spinocerebellar ataxia in Taiwan: expansions of trinucleotide repeats in SCA8 and SCA17 are associated with typical Parkinson's disease*. *Clin Genet*, 2004. **65**(3): p. 209-14.
189. Sanders, S.L., et al., *Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage*. *Cell*, 2004. **119**(5): p. 603-14.
190. Unal, E., et al., *DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain*. *Mol Cell*, 2004. **16**(6): p. 991-1002.

191. Wang, H., et al., *Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage*. Mol Cell, 2006. **22**(3): p. 383-94.
192. Karagiannis, T.C. and A. El-Osta, *Chromatin modifications and DNA double-strand breaks: the current state of play*. Leukemia, 2007. **21**(2): p. 195-200.
193. Bekker-Jensen, S., et al., *Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1*. J Cell Biol, 2005. **170**(2): p. 201-11.
194. Huyen, Y., et al., *Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks*. Nature, 2004. **432**(7015): p. 406-11.
195. Yamane, K., J.E. Schupp, and T.J. Kinsella, *BRCA1 activates a G2-M cell cycle checkpoint following 6-thioguanine-induced DNA mismatch damage*. Cancer Res, 2007. **67**(13): p. 6286-92.
196. Rodriguez, H., et al., *Lymphoblasts of women with BRCA1 mutations are deficient in cellular repair of 8,5'-Cyclopurine-2'-deoxynucleosides and 8-hydroxy-2'-deoxyguanosine*. Biochemistry, 2007. **46**(9): p. 2488-96.
197. Wiltshire, T., et al., *BRCA1 contributes to cell cycle arrest and chemoresistance in response to the anticancer agent irifolven*. Mol Pharmacol, 2007. **71**(4): p. 1051-60.
198. Sugiyama, T. and S.C. Kowalczykowski, *Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation*. J Biol Chem, 2002. **277**(35): p. 31663-72.
199. Shukla, A., et al., *Interaction of hRad51 and hRad52 with MCM complex: a cross-talk between recombination and replication proteins*. Biochem Biophys Res Commun, 2005. **329**(4): p. 1240-5.
200. Ristic, D., et al., *Human Rad51 filaments on double- and single-stranded DNA: correlating regular and irregular forms with recombination function*. Nucleic Acids Res, 2005. **33**(10): p. 3292-302.
201. Kim, P.M., et al., *Overexpression of human RAD51 and RAD52 reduces double-strand break-induced homologous recombination in mammalian cells*. Nucleic Acids Res, 2001. **29**(21): p. 4352-60.
202. Wang, X., et al., *Dysregulation of receptor interacting protein-2 and caspase recruitment domain only protein mediates aberrant caspase-1 activation in Huntington's disease*. J Neurosci, 2005. **25**(50): p. 11645-54.
203. Wang, X., et al., *Protective role of Cop in Rip2/caspase-1/caspase-4-mediated HeLa cell death*. Biochim Biophys Acta, 2006. **1762**(8): p. 742-54.
204. Glaviano, A., et al., *Effects of hTERT on metal ion-induced genomic instability*. Oncogene, 2006. **25**(24): p. 3424-35.
205. Levy, J.A., M. Virolainen, and V. Defendi, *Human lymphoblastoid lines from lymph node and spleen*. Cancer, 1968. **22**(3): p. 517-24.
206. Yandell, D.W. and J.B. Little, *Chromosome 14 marker appearance in a human B lymphoblastoid cell line of nonmalignant origin*. Cancer Genet Cytogenet, 1986. **20**(3-4): p. 231-9.
207. Kodama, Y., et al., *Cytogenetic analysis of spontaneous and 2-cyanoethylene oxide-induced tk-/- mutants in TK6 human lymphoblastoid cultures*. Environ Mol Mutagen, 1989. **14**(3): p. 149-54.
208. Schwartz, J.L., et al., *The TP53 dependence of radiation-induced chromosome instability in human lymphoblastoid cells*. Radiat Res, 2003. **159**(6): p. 730-6.

תקציר

מבוא: נזקי דנ"א מציבים את אחד האימים החמורים לתפקוד התא והאורגניזם, ולכן קיימים בתאים מנגנונים משוכללים כדי לחוש בהם ולהתמודד עם אתגר זה. היעילות שבה מתמודד התא עם נזקי דנ"א קובעת, אם פגיעה זו תתקן בלא שתותיר אפקט מתמשך על חיי התא, או שתגרום למוות תאי מתוכנת, או להתמרה סרטנית. הכרת תגובות התא לנזקי דנ"א היא בעלת השלכות רבות על הבנת תהליכי חיים בסיסיים, כגון בקרת מחזור התא, הזדקנות, והתפתחות רקמות וניוון. הכרה כזו רלוונטית מאוד לבריאות הציבור, בעיקר לגבי גיבוש דרכי התמודדות עם גורמים סביבתיים, המסבים נזק לדנ"א, ועם תהליכים הגורמים להתמרה סרטנית ולמחלות ניווניות שונות.

מטרות: המטרה המרכזית של מחקרנו הייתה לאפיין את ההבדלים בתגובה לשברים דו-גדיליים בדנ"א של מספר שורות תאי אדם, תוך שילוב מערכי דנ"א ממוזערים ושיטות חישוביות מתקדמות. מטרות ספציפיות היו: זיהוי באמצעים חישוביים של פקטורי השיעתוק המניעים את התגובה לנזק בדנ"א ופירוק רשת השיעתוק לזרועותיה השונות, בהתאם לפעולת פקטורי שיעתוק אלו. מטרה נוספת הייתה איתור המכנה המשותף לתגובות לנזק בדנ"א בחמש שורות שונות של תאי אדם, לצד ההבדלים בתגובה הקיימים ביניהם.

שיטות: השתמשנו בטכנולוגיה של מערכי דנ"א ממוזערים, היוצרת כמויות נתונים גדולות ביותר, לבניית פרופילים של ביטוי הגנים. טכנולוגיה זו מאפשרת לזהות גנים המתבטאים באופן שונה באוכלוסיות תאים שונות. הניסויים שבוצעו כללו שימוש בחמש שורות שונות של תאי אדם, שנגזרו מרקמות נורמליות וסרטניות. בכל ניסוי נבנו פרופילים של ביטוי גנים בתנאים ביולוגיים שונים, שכללו השריית נזקים של שברים דו-גדיליים בדנ"א בכמה שורות תאים. השריית הנזקים בדנ"א נעשתה באמצעות הוספת הכימיקל הרדיומימיטי (NCS) neocarzinostatin לתאים, או על ידי הקרנתם בקרינה מייננת (ionizing radiation - IR).

כריית מידע בעל משמעות ביולוגית מתוך כמויות הנתונים הגולמיים הגדולות, הנוצרות מניסויי מערכי דנ"א ממוזערים, מציבה אתגר ביואינפורמטי ניכר. כדי להתמודד עם אתגר זה אימצנו גישת אנליזה מקיפה לניתוח נתונים, המתקבלים מטכנולוגיית המערכים הממוזערים. הגישה פותחת ביישום טכניקות מתקדמות לעיבוד הנתונים ונירמול המערכים, ממשיכה בהפעלת שיטות של קיבוץ (clustering) לזיהוי צברים של גנים המראים דפוס ביטוי דומה, ומסתיימת בהפעלה של ניתוחים סטטיסטיים, המנסים לגלות מחלקות פונקציונליות מועשרות ורצפי בקרת-שיעתוק מועשרים בקרב הצברים שזוהו.

תוצאות: החלק הניסויי של עבודה זו כלל שני ניסויי מערכי דנ"א ממוזערים. בניסוי הראשון, ששימש לכיול, השרינו שברים דו-גדיליים בדנ"א של תאי BJ-hTert (פיברובלסטים) בעזרת הכימיקל הרדיומימיטי NCS ומדדנו

את פרופיל ביטוי הגנים כעבור 2, 4, 8 ו-12 שעות. בתאי ביקורת (שלא טופלו ב-NCS) נרשמו הפרופילים בנקודות הזמן 0 ו-8 שעות. בניסוי זה השווינו בין חזרות ביולוגיות שבוצעו על ידי שתי חוקרות, בחנו את השפעת הלוואי של הפרוצדורה הניסויית על פרופיל ביטוי הגנים וזיהינו נקודות זמן אינפורמטיביות עבור הניסוי המרכזי בעבודה זו.

בניסוי השני נמדדו פרופילים של ביטוי הגנים בתאי אדם מחמישה סוגים: BJ-hTert (פיברובלסטים), G361 (תאי מלנומה), HepG2 (תאי קרצינומה של כבד), TK6 (לימפובלסטואידים) ו-U2OS (תאי סרטן העצם). פרופילי הביטוי נמדדו בתאים לא מטופלים, ובתאים שהוקרנו ב-5 Gy של קרינה מייננת, 3 ו-6 שעות לאחר הקרנה. לאור השפעות הלוואי של הפרוצדורה הניסויית, שנצפו בניסוי הכיול, הוספנו ביקורות של תאים לא מטופלים לכל נקודת זמן. הבחנו, ששני צברי הגנים הגדולים ביותר בכל שורת תאים מתארים שפעול ודיכוי של גנים בתגובה להקרנה. אנליזת פרומוטורים זיהתה העשרה משמעותית של חתימות אתרי הקישור של p53 ו-NF-Y בגנים המשופעלים והמדוכאים בצברים אלו, בהתאמה, והצביעה על התפקיד החשוב של פקטורי שיעתוק אלה בבקרת התגובה לקרינה מייננת. בכל סוגי התאים זוהתה העשרה של קטגוריות פונקציונליות בצברי הגנים המדוכאים בעקבות הקרנה. על הקטגוריות שזוהו נמנות הקטגוריות "ויסות מחזור התא" ובאופן מיוחד "מיטוזה", דבר המצביע על שיפעולם של מנגנוני בקרת מחזור התא. באופן מפתיע, לא נמצא מכנה משותף בין שורות התאים הסרטניים, שיבדיל אותן משורות התאים הלא-סרטניים. הסבר אפשרי לכך הוא, שגם שורות תאים "לא סרטניות" אינן בעלות מאפיינים של תאים נורמלים ברקמה.

מסקנות: בעבודה זו שילבנו מערכי דנ"א ממוזערים עם כלים חישוביים כדי להשיג ניתוח כוללני של תגובת השיעתוק התאית לקרינה מייננת (IR) בחמש שורות תאים, ולזהות את ההבדלים בין תגובותיהם. ההבדלים בביטוי הגנים בתגובה לנזק בדנ"א מעידים על החשיבות של השוואת תוצאות מכמה מערכות ניסוי שונות והימנעות מהסקת מסקנות כוללניות, המבוססות על מחקר בשורת תאים או רקמה בודדת.

התוצאות שקיבלנו מדגישות את חשיבותם של פקטורי השיעתוק p53 ו-NF-Y כבקרים מרכזיים של תגובות אלה. כמו כן, במחקר זה זוהו גנים המהווים 'מטרות' חדשות אפשריות של p53.

מילות מפתח: גנומיקה פונקציונלית, תגובה לנזקי דנ"א, קרינה מייננת, ביטוי גנים, מערכי דנ"א ממוזערים, מערכי אוליגונוקלאוטידים בצפיפות גבוהה, קיבוץ, אנליזה פונקציונלית, אנליזת פרומוטורים.

אוניברסיטת תל-אביב

הפקולטה לרפואה ע"ש סאקלר

החוג לגנטיקה מולקולרית של האדם ולביוכימיה

**ניתוח ברמה כלל-גנומית של תגובות תאיות לשברים דו-גדיליים
בדנ"א באמצעות מערכי דנ"א ממוזערים ואנליזה ביואינפורמטית**

עבודת הגמר מוגשת על-ידי

סיגל שביט

ת.ז. 034233205

עבודה זו בוצעה כמילוי חלקי של הדרישות לקבלת תואר מוסמך בפקולטה לרפואה ע"ש סאקלר,

אוניברסיטת תל-אביב

עבודת מחקר זו נעשתה בהנחיית:

פרופ' יוסף שילה

פרופ' רון שמיר

החוג לגנטיקה מולקולרית וביוכימיה

ביה"ס למדעי המחשב

ביה"ס לרפואה ע"ש סאקלר

הפקולטה למדעים מדויקים

ספטמבר 2007