12-2012

# An Investigation Of Gene Networks Influenced By Low Dose Ionizing Radiation Using Statistical And Graph Theoretical Algorithms

Sudhir Naswa
snaswa@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Sudhir Naswa entitled "An Investigation Of Gene Networks Influenced By Low Dose Ionizing Radiation Using Statistical And Graph Theoretical Algorithms." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Michael A. Langston, Major Professor

We have read this dissertation and recommend its acceptance:

Brynn H. Voy, Arnold M. Saxton, Hamparsum Bozdogan, Kurt H. Lamour

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Sudhir Naswa entitled “An investigation of gene networks influenced by low dose ionizing radiation using statistical and graph theoretical algorithms”. I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

_____
Michael A. Langston, Co-Chair

_____
Brynn H. Voy, Co-Chair

We have read this dissertation and recommend its acceptance:

Arnold M. Saxton_____

Hamparsum Bozdogan_____

Kurt H. Lamour_____

Accepted for the Council

_____

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

# An Investigation Of Gene Networks Influenced By Low Dose Ionizing Radiation Using Statistical And Graph Theoretical Algorithms

**A Dissertation Presented for the**
**Doctor of Philosophy**
**Degree**
**The University of Tennessee, Knoxville**

**Sudhir Naswa**
**December 2012**

## DEDICATIONS

This dissertation is dedicated to my parents (Mrs. Santosh Naswa and Mr. N. R. Naswa) and my wife Babita for always being there to support and encourage me. It is also dedicated to my cheerful daughter Shachi for motivating and inspiring me.

## ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Michael A. Langston and co-advisor Dr. Brynn H. Voy for guiding, encouraging and supporting me throughout my graduate education. I was fortunate to have them as my mentors since I could gain knowledge in computer science as well as biology from them that would be invaluable for me in pursuit of a scientific career in bioinformatics. I would also like to thank Dr. Arnold M. Saxton for training me in biostatistics and for always providing me valuable advice on biosatistical issues relating to my research. I am thankful to Dr. Hamparsum Bozdogan for mentoring me for the Master's degree in statistics and for teaching me mathematical principles of statistics. I would also like thank Dr. Kurt Lamour for providing me with opportunity to learn laboratory skills in his lab that have always helped me think computational biology problems from the perspective of a wet bench scientist. I would like to thank GST program at University of Tennessee for giving me an opportunity to pursue graduate studies and to Low dose radiation research program of Department of Energy for supporting my research.

## ABSTRACT

Increased application of radiation in health and security sectors has raised concerns about its deleterious effects. Ionizing radiation (IR) less than 10cGys is considered low dose ionizing radiation (LDIR) by the National Research Committee to assess health risks from exposure to low levels of IR.

It is hard to extract the effects of mild stimulus such as LDIR on gene expression profiles using simple differential expression. We hypothesized that differential correlation instead would capture the effects of LDIR on mutual relationships between genes. We tested this hypothesis on expression profiles from five inbred strains of mice treated with LDIR. Whereas ANOVA detected little effect of LDIR on gene expression, a differential correlation graph generated by a two stage statistical filter revealed gene networks enriched with genes implicated in radiation response, DNA damage repair, apoptosis, cancer and immune system.

To mimic the effects of radiation on human populations, we profiled baseline expression of recombinant inbred strains of BXD mice derived from a cross between C57BL/6J and DBA/2J standard inbred strains. To establish a threshold for extraction of gene networks from the baseline expression profiles, we compared gene enrichment in paracliques obtained at different absolute Pearson correlations (APC) using graph algorithms. Gene networks extracted at statistically significant APC ($r \approx 0.41$) exhibited even better enrichment of genes participating in common biological processes than networks extracted at higher APCs from 0.6 to 0.875.

Since immune response is influenced by LDIR, we investigated the effects of genetic background on variability of immune system in a population of BXD mice. Considering immune response as a complex trait, we identified significant QTLs explaining the ratio of CD8+ and CD4+ T-cells. Multiple regression modeling of genes neighboring statistically significant QTLs identified three candidate genes (Ptprk,Acp1 and Lamb1-1) explaining 61% variance of ratio of CD4+ and CD8+ T cells. Expression profiling of parental strains of BXD mice also revealed effects of LDIR and LDIR*strain on expression of genes related to immune response. Thus using an integrated approach involving transcriptomic, SNP and immunological data, we have developed novel methods to pinpoint candidate gene networks putatively influenced by LDIR.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1 : INTRODUCTION AND LITERATURE REVIEW

## IONIZING RADIATION AND ITS EFFECTS

There is considerable interest in the effects of radiation because of its clinical applications as well as its harmful effects. Radiation has been extensively employed for diagnostics (X- rays, CT scans) as well as therapeutic purposes (radiotherapy of cancers). Conversely, the energy in the radiation can damage bio-molecules such as proteins, DNA and lipids. Radiation capable of evicting electrons from its target atoms is called ionizing radiation. High energy alpha (positively charged helium nuclei) and beta (negatively charged electrons) particles, gamma-rays (high energy photons) and X-rays (high energy photons) are examples of ionizing radiation.

The intensity of radiation is measured and reported in different units quantifying either its energy content or its ability to damage biological tissues. The energy absorbed by a unit mass of biological tissue from the ionizing radiation is called “absorbed dose” that is measured in Gray. One Gray is equivalent to absorption of 1 Joule of energy per kilogram of biological tissue. Different types of radiations are capable of transferring energy to biological tissues at different rates along their path. Electromagnetic radiations such as X-rays and gamma rays cause comparatively less damage since they spread their deleterious effects along a longer path by moving faster. Conversely heavier alpha particles move slowly and transfer more energy to the molecules encountered on their path. Therefore, energy absorbed from alpha particles may cause more damage to biological tissues than the same amount of energy delivered by gamma radiation. The effective dose measured in Sieverts (Sv) puts all the radiations on equal footing in terms of potential biological damage caused by them. To get the

intensity of radiation in Sieverts the energy absorbed per unit of biological mass is multiplied by a radiation weighting factor. For beta, gamma and X-rays this factor is 1 and for alpha particles it is 20. Hence, same magnitude of energy absorbed from the alpha particles is estimated to cause 20 times more damage to biological tissues as compared to gamma rays [1].

### SOURCES OF IONIZING RADIATION

Ionizing radiation is emitted by both natural and man-made sources. Natural sources of radiation include cosmic rays and radon. Artificial sources include radiation emission from medical/laboratory equipments (X-rays, CAT scan), body scanners at airports, nuclear research, nuclear power, nuclear weapon programs, and nuclear disasters (Chernobyl nuclear power plant in Ukraine, 1986 and Fukishima Daiichi Nuclear Power Plant in Japan, 2011).

Cosmic radiation accounts for 15% exposure to natural radiation at sea level. The exposure increases at higher altitudes and in aircrafts. Average exposure to cosmic radiation at the cruising height of a commercial aircraft varies between 0.003 mSv and 0.008mSv per hour that is twice the amount at sea level. Another major source of natural exposure is the radon-222 gas that is produced as a decomposition product of uranium-238. Radon accounts for nearly 50% of radiation exposure due to natural sources [2]. The annual average background radiation from natural sources is estimated to be 2.4 mSv [3]. The average radiation levels at some places, for example Ramsar in

Iran, Karunagappally in Kerala, India and Yangjiang in Guangdong, China) have much higher background radiation levels than average [4, 5].

Medical exposure accounts for 98% of exposure from all the artificial sources of radiation and 20% of exposure from all the sources of ionizing radiation. Exposure to radiation from the diagnostic medical techniques is increasing with the increasing availability of these techniques over the last few years especially in developed world. The average per capita exposure due to medical diagnostic techniques stands at 0.62mSv to 1.92mSv for the period 1997 to 2007[3]. Dose of less than 10cGy (100 mSv or 100 mGy) is considered as Low dose Ionizing Radiation (LDIR) by the National Research Committee (NRC) to assess health risks from exposure to low levels of ionizing radiation [6]. The committee divided all the countries into four levels based on number of physicians per 1000 persons. Level 1 corresponded to countries with highest number of physicians and level 4 corresponded to countries with lowest number of physicians. In level 1 countries exposure due to medical diagnostic techniques was as high as 1.92 mSv per capita (Figure 1-1). In some countries the average exposure due to medical diagnostic techniques has become higher than the previously known largest source of ionizing radiation i.e. background radiation from nature [3]. The average adult effective radiation dose from some of the medical imaging procedures can be very high (Table 1-1).

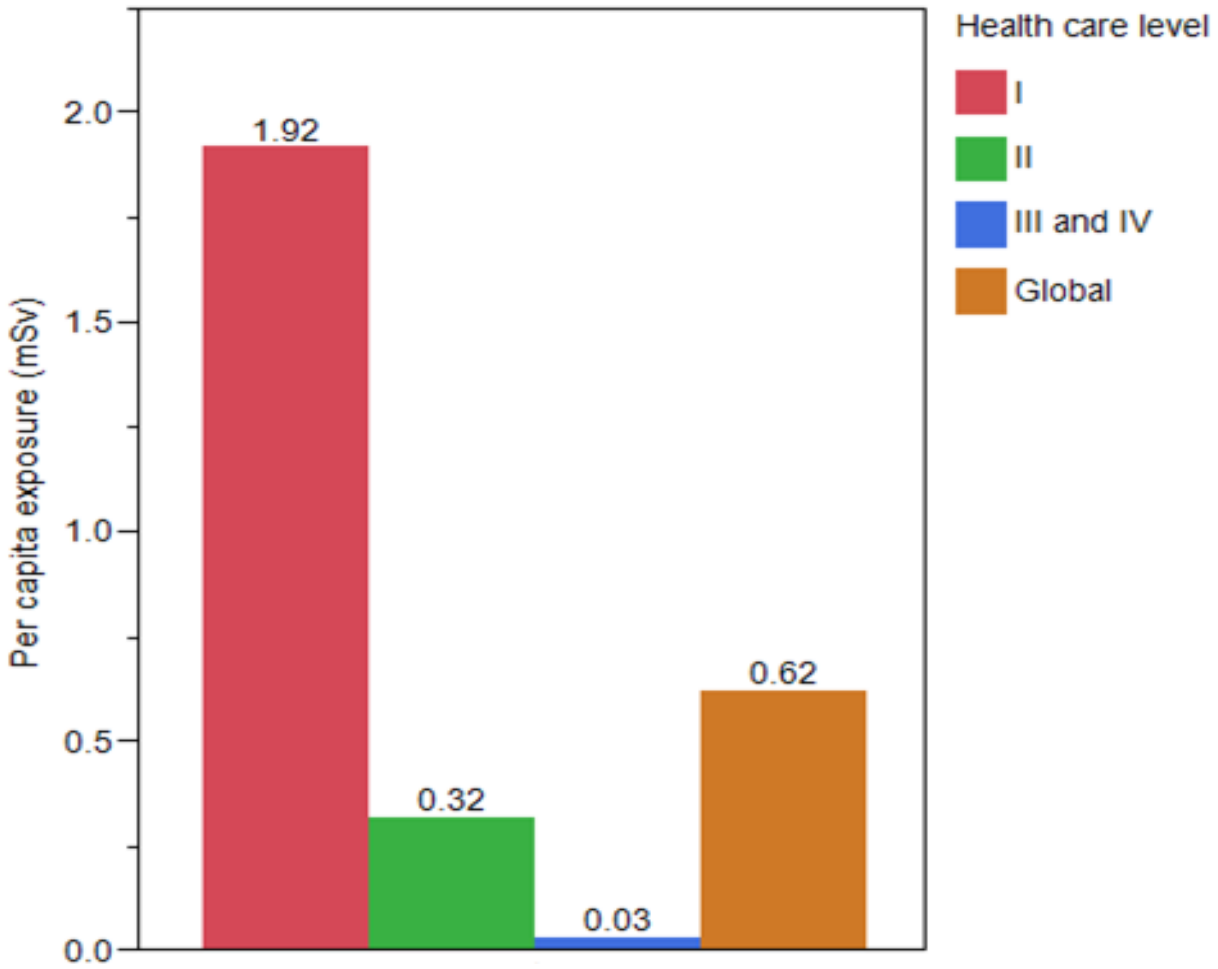



**Figure 1-1: Average per caput annual exposure to Ionizing radiation due to medical and dental examinations (1997-2007).**Data from countries with varying health care levels (Health care facilities decrease with increasing level). (Redrawn from: *United Nations Scientific Committee on the Effects of Atomic Radiation: Sources and effects of ionizing radiation Radiation: Volume I, 2008 [3]* )

**Table 1-1 Average adult effective radiation doses from various medical procedures**

| Procedure | Average Adult Effective Dose (in mSv) | Estimated Dose Equivalent (Number of Chest X-rays) |
|---|---|---|
| ***Dental X-ray*** | *0.005 to 0.016*[a] | *0.25 to 0.5* |
| ***Chest X-ray*** | *0.02* | *1* |
| ***Mammography*** | *0.4* | *20* |
| ***CT*** | *2 to 16*[b] | *100 to 800* |
| ***Nuclear Medicine*** | *0.2 to 41*[c] | *10 to 2050* |
| **Interventional Fluoroscopy** | 5 to 70[d] | 250 to 3500 |

[a] 0.005 mSv for an intraoral dental x-ray & 0.01 mSv for a panoramic dental x-ray.

[b] 2 mSv is for a CT exam of the head & 16 mSv for CT coronary angiography exam.

[c] 0.2 mSv for lung ventilation exam using $^{99m}$Tc-DTPA & 41 mSv for a cardiac stress-rest test using thallium 201 chloride.

[d] 5 mSv for a head and/or neck angiography exam & 70 mSv for a transjugular intrahepatic portsystemic shunt placement. (Source: Initiative to reduce unnecessary radiation exposure from medical imaging, February 2010, Center for devices and radiological health, U.S. Food and Drug Administration [7])

## EFFECTS OF IONIZING RADIATION

Higher doses of radiation are known to cause deterministic health hazards such as development of blisters on skin, damage in bone marrow and gastro intestinal tract, chromosomal aberrations, apoptosis, cancers etc. A dose of 5 gray received instantaneously is lethal without medical intervention. A dose above 20 gray is lethal even with medical intervention [1]. Exposure to lower doses for a longer period of time has been associated with DNA damage and development of cancers or birth defects. However most of the evidence for development of these diseases has been obtained from the epidemiological studies.

At cellular level high energy radiation can ionize and damage biological molecules by direct deposition of energy. But usually the effects of the radiation are a consequence of increase in free radicals that persist for few milliseconds and cause oxidative damage to DNA, proteins and lipids [8]. Radiation induced alterations in the genome can have adverse effects if they are passed on to the next generation of cells. Cells try to maintain the genomic integrity by initiating DNA damage responses that include cell cycle arrest and DNA repair. Irreparable cells are removed by apoptosis [9].

## LINEAR NO THRESHOLD MODEL (LNT MODEL)

The most common model used to explain the effects of lower radiation doses is the "Linear No Threshold" model. The LNT model assumes that even a small dose of radiation could potentially be harmful, and that there is no minimal threshold for deleterious effects of radiation. The harmful effects of the radiation are believed to be

directly proportional to the intensity of radiation (Figure 1- 2). Hence, according to this model the harmful effects of low doses of radiation can be estimated by extrapolating the known effects of higher doses of radiation. There is evidence for, as well as against, the applicability of the LNT model to lower doses of radiation. The LNT model is based on the fact that even a single high energy radiation particle (like a photon or alpha particle) is capable of ejecting electrons from their target molecule. A chance collision of these particles with a biological molecule like DNA could disrupt its structure.

Accumulation of damages by repetitive exposure to lower doses of radiation may culminate in harmful effects such as chromosomal aberrations [10]. Similarly a linear relationship is believed to exist between dose of radiation and breast cancer [11]. In the case of solid cancers, the LNT model provides the most suitable explanation for a relation between dose and incidence of cancer [6, 12].

### HORMESIS AND ADAPTIVE RESPONSE

The opponents of the LNT model cite hormesis (beneficial effects) and adaptive response to radiation as evidence against it. According to the hormesis model, lower doses of radiation are believed to be beneficial to an organism. Hormesis may be caused due to adaptive response to radiation i.e. decrease in vulnerability of cells to higher doses of radiation after they are pre exposed to lower doses [13-15]. The adaptive response of cultured cells is usually observed when cells are pre-exposed to lower doses of radiation between 10 to 200 mGy. Exposure between 200 mGy and

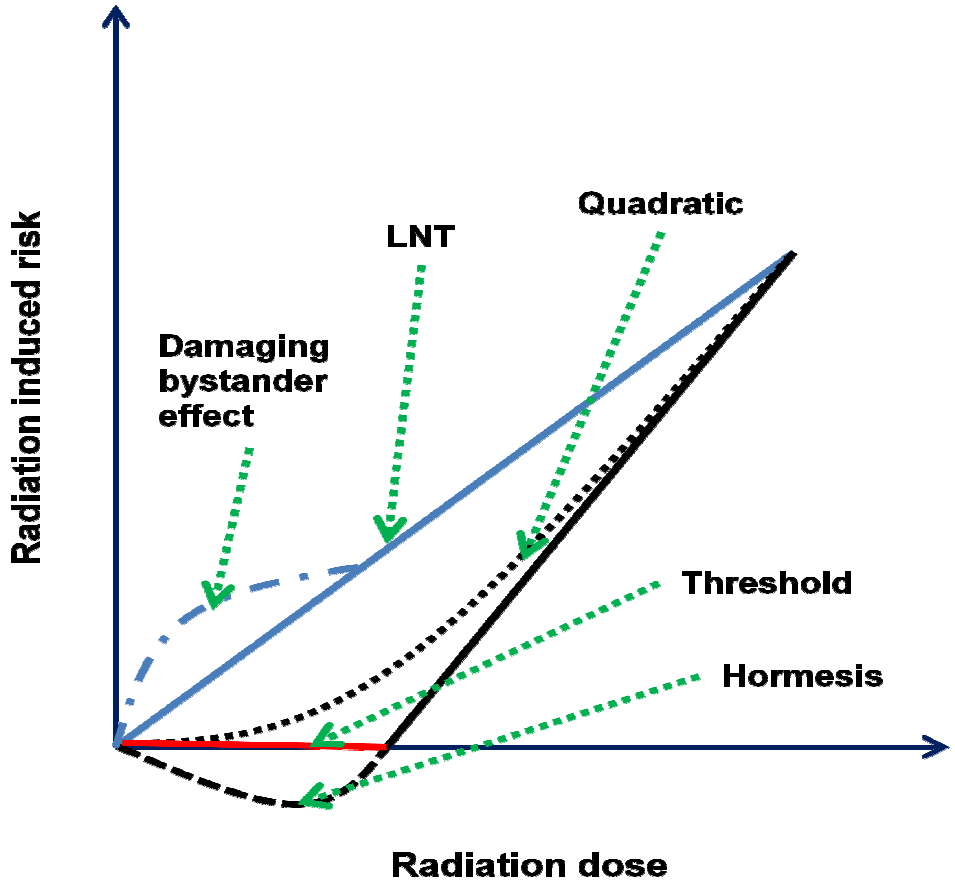



**Figure 1-2: Linear No Threshold (LNT) and other models explaining the effects of radiation**. (Adapted from Prise KM [16] and NRC [6] )

500mGY may also effectuate adaptive response but a dose greater than 500 mGy does not elicit adaptive response in the cells [17]. Adaptive response is most often used to describe decreased chromosomal damage and increase in survival rates of cells pre-exposed to LDIR when confronted with higher doses of radiation. Decrease in chromosomal damage and aberrations in response to high doses of X-rays was observed in lymphocytes when they were pretreated with lower doses of radon [13, 15]. Olivieri et al observed increase in survival of U1-Mel and Hep-2 neoplastic cells if they were primed with low dose of X-ray before being treated with higher dose [14]. Adaptive response to radiation may also enhance the capability of cells to cope with oxidative stress when confronted with higher doses of radiation. LDIR exposed lymphoblastoid cells exhibited higher activity of anti-oxidant enzymes after exposure to HDIR as compared to those that were directly treated with HDIR[18]. Similarly human colon carcinoma cells exhibited an increase in levels of SOD2 (Superoxide dismutase 2) when they were pretreated with LDIR (10cGy) before being exposed to HDIR [19]. Priming with low doses of radiation may also decrease neoplastic transformation and development of cancer caused by high doses of radiation [20]. Decrease in lung cancer in humans has been associated with increase in intensity of low dose radiation from inhalation of radon [21]. Hormesis may also be exhibited in the form of increased proliferation of cells. An increase in division of mesenchymal cells was observed when they were treated with 75mGy of ionizing radiation [22].

Low doses of radiation are also believed to have a stimulatory effect on the immune system. Mice exposed to LDIR exhibit significant change in cytokine profiles suggesting

changes favoring stimulation of innate immune system[23]. Differences in genetic background have also been shown to elicit differential immune response after exposure to low doses of radiation. The immune response was stimulated after exposure to low dose of radiation in case of C57BL/6 as exhibited by increase in proliferation of spleenocytes and decrease in expression of p53 and apoptosis. On the other hand BALB/c mice exhibited exactly the opposite effects i.e. decrease in spleenocytes and increase in expression of p53 and apoptosis [24].

Both adaptive responses and radiation damage have been reported in populations that are exposed to chronic levels of low dose of ionizing radiation. Ramsar in Iran is an area with high levels of natural background radiation. The volunteers from this area exhibited stimulated immune system (increased levels of IG-E and higher proportion of activated T-cells) as compared to volunteers from outside this area [25]. Blood mononuclear cells treated with 4 Gy of gamma-radiation exhibited lower micronuclei formation and higher apoptosis levels in case of volunteers from Ramsar as compared to those from outside this area [26]. On the other hand higher aberrations and breaks were found in chromosomes of peripheral lymphocytes from hospital workers exposed to low doses of radiation as compared to controls [27].

#### LOW DOSE HYPERSENSITIVITY

Some studies indicate that the biological systems exhibit increase in apoptosis after exposure to ionizing radiation because they are hypersensitive to it. This hypersensitivity may develop because damages caused by LDIR may not be strong

enough to activate DNA damage response pathways like ATM for repair of DNA. The exposed cells may therefore accumulate DNA damages over a period and become unviable. They are eventually eliminated by apoptosis. Cells exposed to higher doses respond by activating the DNA repair pathways preventing hypersensitive response to radiation [28-30].

### INFLUENCE OF RADIATION BEYOND THE TARGETED CELLS

The effects of radiation are not limited to its target cells alone. The influence of radiation can be spread in space to non-targeted neighboring cells by bystander effects. Radiation can also pass its effects to a different time point through genomic instability that appears in progeny of the targeted cells.

## Bystander Effect

It was initially believed that the radiation influences cells that are in its direct path through mutations, DNA damage, chromosomal aberrations, apoptosis etc. However in last few years it has been observed that cells that are not in the path of ionizing radiation may be affected by bystander effect i.e. they receive signals from the cells targeted by the radiation and get influenced by them. The bystander effect is manifested in many ways in the cells influenced by it. These effects include increased 'sister chromatid exchange' [31] , micronuclei formation [32], mutations [33, 34] and neoplastic transformation [35] of the cells under its influence. Such an effect can even be observed in un-irradiated cells grown in media from irradiated cells [35-37]. Medium from irradiated cells may also promote division in the cells grown in it due to bystander

effect [38]. Medium from cells exposed to relatively low doses of radiation (0.073, 1 and 2 Gy from silicon ion Si490) also induces bystander effects. Un-irradiated cells grown in this medium exhibited chromosomal instability [37].

Communication through gap junctions and oxidative metabolism have been used to explain the bystander effect of radiation [39]. Gap junctions are inter-cell communication channels that selectively allow transmission of some molecule over the others. There is evidence of role of gap junctions in transmission of damage signals from radiation targeted cells to the un-targeted cells [34, 40, 41] . Reactive oxygen species (ROS) are produced by the oxidative metabolic processes in cells. Disruption of balance between ROS and antioxidants is responsible for the manifestation of various pathological conditions and ageing [42]. The balance between ROS and antioxidants in a cell can be disrupted by radiation exposure since a single alpha particle traversing through a mammalian cell can generate tens of thousands of ROS [39]. The role of ROS in generating bystander effects has been shown by inhibition of ROS mediated sister chromatid exchange [31], DNA mutations and gene expression [32] by using antioxidants. Bystander effect may even be pro-mitotic for cells grown in medium of irradiated cells [38].

**Genomic instability**

Genomic instability is a characteristic feature of most cancers and ageing .It includes appearance of ‘new mutations’, ‘chromosomal aberrations’, ‘neoplastic transformations’ and ‘increased cell death’ etc. [43, 44]. The effects of radiation are not limited to direct damages to the genetic material that are passed on to the descendents of the target

cell. The genome of descendent cells may also exhibit diverse changes that are not present in the irradiated parent cells. A single cell may therefore produce a progeny of cells with diverse mutations and cytogenetic defects as in genomic instability. These changes may appear even after many cycles of cell division. New chromosomal aberrations have been observed in descendants of cells exposed to radiation [45-47]. Besides chromosomal aberrations the cells descending from the irradiated cells exhibit low platting efficiency [48], increased transformation [49, 50] and new mutations [48, 49]. Genomic instability has also been reported in descendents and bystanders of cells irradiated with low dose of ionizing radiation [37]. The mechanism of generation of genomic instability following irradiation is poorly understood. It is unlikely that the genomic instability could be due to a mutation in a gene responsible for maintenance of the genome. This is because the rate of random mutation due to radiation ($10^{-4}$ per cell per Gy) is much lower than that of radiation induced genomic instability ($10^{-1}$ per cell per Gy) [51]. The possible reasons advanced to explain the phenomenon of genomic instability include malfunction of mitochondria, epigenetic mechanisms and inflammatory response [51, 52]. The descendants of the irradiated cell may have malfunctioning respiratory processes in mitochondria that may cause oxidative stress resulting in mutations [53, 54]. Exposure to radiation also causes dose dependent epigenetic changes like hypomethylation of DNA. Hypomethylation of DNA has been associated with appearance of breaks in chromosomes and increase in expression of proteins essential for DNA maintenance [52, 55, 56].

## RADIATION AND CANCER

Various experimental as well as epidemiological studies have associated diagnostic, therapeutic as well as accidental exposures to radiation with increased risk of cancer. Elevated numbers of skin, lung, thyroid and breast cancers have been reported in the survivors of Atomic bomb [57, 58] . Thyroid is one of the most radiation sensitive organs especially in children [59]. Increased incidence of thyroid cancer has been observed in children from contaminated areas after Chernobyl accident [60]. Chernobyl cleanup workers show increased risk of leukemia and hematological malignancies [61]. Carcinogenic effects may appear after a long latency period (20 to 70 years) after exposure to ionizing radiation [62-64].

Therapeutic use of radiation has also been associated with cancers. In 1950's radiation therapy was used to treat tinea capitis, a fungal infection of the scalp. Persons treated with ionizing radiation to cure tinea capitis exhibited fourfold increase in skin cancers (mainly basal cell carcinoma) and threefold increase in benign tumors of skin later in their lives [65] . The risk of development of the basal cell carcinomas decreased with increase in age. Following two reasons could possibly explain increased susceptibility to radiation at a younger age.

i) the cells are more actively dividing at a younger age and

ii) longer period of remaining life increases the probability of synergistic effects with other carcinogens such as UV radiation [65].

Even radiation therapy for cancers can increase chances of development of secondary cancers due to exposure of healthy tissues [66, 67].

Recent years have witnessed a substantial increase in exposure to LDIR on account of increased use of radiation based therapeutics, medical imaging and increased scanning at airports [68-70] . Considering the harmful effect of radiation there is growing concern about increase in usage of radiation based diagnostic methods. There is conflicting evidence based on epidemiological studies about the effect of lower doses of radiation. Some epidemiological studies have linked cancers with LDIR used for diagnostic purposes (X-ray and CT-scan) [71-73]. Dose dependent relationship was observed between LDIR exposure from cardiac imaging and therapeutic procedures and subsequent risk for cancer [73]. Increase in malignancies have also been reported in eight counties of Sweden exposed to LDIR from fallout of Caesium-137 released in atmosphere after Chernobyl accident [74]. Higher frequency of cancers appeared in thyroid glands exposed to low average dose of 9.8 cGy in persons receiving radiation therapy for tinea capitis [64]. The carcinogenic effects of radiation may also be variable in a population. Variability in predisposition to thyroid cancer has been found in the population exposed to low doses of ionizing radiation after Chernobyl accident [75].

Conversely, many studies have not supported carcinogenic role of LDIR below 100mSv. This could be explained either by the fact that the effects of LDIR are too little to be detected statistically at a feasible sample size or it may be possible that there is some minimum threshold of dose below which cancer is not manifested. Accordingly there is a belief that extension LNT to LDIR is counterproductive. Adherence to such a belief

could even be detrimental to public health because of underutilization of available radiation based diagnostic methods [76, 77].

### OTHER EFFECTS OF RADIATION

Radiation has also been associated with many other ailments including fatigue, DNA and chromosomal damage and cardiovascular disease. Many cancer patients experience fatigue after radiotherapy [78]. Fatigue has been reported in CD-1 mice exposed to a single low dose gamma radiation (50cGy and 200cGy) [79]. Increased incidence of cataracts and cardiovascular diseases have been reported among the workers involved in cleanup work after Chernobyl accident [61]. Similarly an increase in DNA damage and chromosomal aberrations have been reported in people chronically exposed to low doses of radiation like hospital workers [80] , radiologic technologists [10] and residents of geographical areas with higher natural background radiation [81-83]. Changes in transcription profiles of genes have been reported to be induced by radiation including LDIR [84]. Radiation also induces changes in translation by influencing recruitment of RNAs to polysomes [85].

### REACTIVE OXYGEN SPECIES

Reactive oxygen species (ROS) are produced during metabolic processes in the mitochondria as well as by various oxidizing enzymes such as NADPH oxidase, xanthine oxidase, amino acid oxidases, cytochrome P450, flavoprotein dehydrogenase , glycollate oxidase etc. It is estimated that about 1% of $O_2$ taken in by the mammalian cells are converted to ROS in mitochondria [86]. Bio-molecules like proteins and DNA

can be damaged by their oxidation through reactive oxygen species (ROS) such as peroxide ($H_2O_2$, ROOH), oxygen ions ($O2^-$) and hydroxyl ions radicals ($OH^*$) [8]. The deleterious effects of ROS are neutralized by physical barriers (histone molecules surrounding DNA) as well as enzymatic and non-enzymatic molecules. Enzymes such as superoxide dismutase, glutathione peroxidase, glutathione transferase, catalase etc. and non enzymatic molecules (pyruvate, Vitamin A, E and C) keep the cellular environment in the desired reduced state [87]. The balance between oxidants and antioxidants is disturbed by external agents such as radiation by generating excess of ROS and the consequent oxidative stress. The inability of the protective methods to cope with increased levels of ROS result in increased probability of unrepaired DNA damages in the cells. The damage due to radiation could be different from the damage resulting from day to day metabolic processes. Radiation induced damage may have phosphate or phosphoglycolate at the 3' end instead of the hydroxyl group that should be repaired so that the phosphodiester bond of the backbone could be formed [88]. The difference arises from the fact that the ionizing radiation can produce hydroxyl radicals concentrated at one place. The concerted attack of these highly reactive hydroxyl radicals within a limited distance concentrates lesions close to each other on DNA producing Locally Multiply Damaged Sites (LMDS) [89]. Since the hydroxyl ions are highly reactive it has been estimated that those formed within 3 nm of DNA molecule could potentially interact with it [90].

## DNA DAMAGE REPAIR

Ionizing radiation causes damage to DNA either by ROS or by direct deposition of energy on the DNA molecule. Damages may be caused to the bases in DNA or to the sugar phosphate backbone to form Single Strand Breaks (SSBs). If two SSBs are produced closely on opposite strands within 10-20 base pairs there is a chance of breaking the DNA molecule into two pieces to produce the Double Strand Break (DSB) [91]. The double strand breaks may result in chromosomal aberrations such as deletions, inversions and insertions. If the cells containing damaged DNA are allowed to multiply and proliferate they may lead to diseases such as cancer. It is therefore imperative to maintain the genomic integrity. This is achieved either by stopping the cell cycle and repairing the damaged DNA or by initiating suicidal apoptotic response [9]. Molecular responses to radiation induced DNA damage include different but overlapping pathways such as base excision, SSB and DSB repair pathways.

### Base Excision Repair (BER)

BER pathway repairs single base lesions caused by metabolic ROS or by ionizing radiation. The pathway starts by recognition of damaged base by DNA glycosylases. The glycosylase then removes the damaged base by breaking its bond with sugar phosphate backbone to create an apurinic/apyrimidinic (AP) site. The 5' end of this abasic site is then removed by APE1 (apurinic/apyrimidinic endonuclease 1). Polymerase beta and ligase 3 are then attached to XRCC1 (a scaffold protein) at the AP site. After an appropriate nucleotide is attached to DNA at the AP site by polymerase beta the gap is sealed by ligase 3 [92, 93]. Sometimes instead of repairing a single DNA

molecule a longer patch of DNA is replaced. In this case any of polymerase- beta, polymerase-delta or polymerase-epsilon may be used for inserting appropriate nucleotides at the site. The overhanging DNA flap is then removed from the broken strand after simulation of FEN1 protein by PCNA. The strand is then ligated by ligase 1 [94, 95].

**Single Strand Break Repair (SSB repair)**

SSB repair sites are initially stabilized by a protein PARP1 (poly ADP-ribose polymerase-1). Thereafter different BER enzymes are used depending on nature of ends of SSB [96, 97]. If 3' end is hydroxyl and 5' end is phosphate or deoxyribose phosphate then polymerase beta recognizes it and removes the deoxyribose phosphate. There after the gap is filled by polymerase beta and the strand is ligated by XRCC1 and ligase-3. On the other hand if the end points of the SSB are modified then they are first converted back to usual 3' hydroxyl and 5' phosphate. The ends are processed by PNK or APE1in case of modification of 3'end to phosphate or phosphoglycolate respectively. Thereafter the SSB repair is completed by polymerase beta and XRCC1-ligase 3 [96].

**Double Strand Break Repair (DSB repair)**

DSBs are the most dangerous damages caused to the genome by genotoxic agents such as radiation and ROS produced during metabolic changes etc. The probability of generation of chromosomal aberrations (deletions, loss of heterozygosity, translocations, inversions) due to DSBs is very high because they break the DNA into two separate pieces.

The DSBs are sensed by a conserved trimer of three proteins Mre11, Rad50 and Nbs1 (MRN complex) that activate the protein serine/threonine kinase ATM (ataxia telangiectasia mutated) [98]. ATM exists as an inactive dimer in nucleus. In response to DSBs it auto-phosphorylates to an active monomer [99] that participates in phosphorylation of over 30 targets [100]. ATM helps DSB repair process in many ways that include activation of apoptosis to remove cells with irreparable DNA, activation of the cell cycle check points to provide time for DNA damage repair and recruitment of BRCA1 complex required for DSB repair [101-107].

Activated ATM phosphorylates histone protein H2Ax enabling it to bind with MDC1. The attached MDC1 is also phosphorylated by ATM to form a complex with RNF8 and UBC13 at the DSB site. MDC1-RNF8-UBC13 complex ubiquitylates H2Ax. Ubiquitylated H2AX is poly-ubiquitylated by RNF168-UBC13 dimer. Poly-ubiquitylated H2Ax is then attached with BRCA1-A complex required for DBS repair [101, 102].

Monomer ATM also activates p53 by phosphorylating it and its repressor MDM2 [103]. A hetero-dimer consisting of proteins Brca1 and Bard1 is also required for ATM induced phosphorylation of p53 [104]. Active p53 in-turn activates p21 (Cdkn1a) that suppresses Cdk2 and Cdk4 culminating in cell cycle arrest in G1 phase [105]. Similarly the phosphorylation of Nbs1 and Chk2 by ATM triggers arrest of cell cycle in S-phase [108]. Activated Chk2 also phosphorylates Cdc25. Phosphorylated Cdc25 binds with protein 14-3-3 and this complex is removed from nucleus to cytoplasm. Cdc25 is responsible for dephosphorylation of cdc2 a step required for G2 to M phase transition.

Phosphorylation of Cdc25 by Chk2 and subsequent removal from nucleus inhibits the transition from G2 to M phase [109].

ATM mediated arrest of cell cycle in one of the stages of interphase provides an opportunity to repair processes to rectify the damaged DNA. DSB repairs are carried out by two different mechanisms 1) Non Homologous end joining and 2) Homologous recombination.

Non Homologous end joining (NHEJ): In case of NHEJ the DSBs are joined together without necessarily matching the opposite strands. This process is therefore error prone and is responsible for appearance of translocations in the chromosome. Two different pathways have been identified for carrying out NHEJ. They are the classical (c-NHEJ) and alternative (A-NHEJ) pathways.

The c-NHEJ pathway is believed to be flexible in order of its occurrence. Usually after detecting a DSB, a heterodimer protein Ku (consisting of Ku-70 and Ku-86) attaches to the broken ends of DNA to protect it from exonucleases. Thereafter DNA-PKcs (DNA protein kinase catalytic subunit) joins Ku to form a trimer DNA-PK. DNA-PK activates another protein Artemis that cleans up the ends of broken DNA. This is followed by ligation of two broken ends by another trimer consisting of DNA ligase IV, XRCC4 (X-ray cross complementing 4) and XLF (XRCC4-like factor) [110, 111].

In A-NHEJ pathway the two DSB ends are first joined by overlaps of at least a few bases (microhomologies) in the opposite strands. Though the pathway is not fully known it employs a number of proteins including XRCC1, PARP1 (Poly ADP Ribose

polymerase 1, DNA ligase III, Polynucleotide kinase (PNK) , Flap endonuclease 1 (Fen1), Mre11, Rad50 and Nbs1 [111-113].

Homologous recombination: In homologous recombination the broken end of a DSB is joined to its correct partner by using the information in the sister chromatid (in G2 phase), homologous chromosome or a similar repeat in the DNA. This pathway is started by recognition of the DSB by MRN complex. The Mre11 protein of MRN complex is also a nuclease that is involved in generation of 3' single strand overhangs at the DSB site [114]. These single stranded DNA overhangs are coated with RPA [115]. Rad51 and ATP are then attached to RPA coated single strand of DNA to form right handed filament. Rad51 helps in the repair processes by directing the single strands towards their homologue. This process is aided by a number of other proteins such as Rad51B, Rad51C, Rad51D, Xrcc2 and Xrcc3 and BRCA2. Since DSB repair by homologous recombination involves matching of complimentary strands the process is more reliable as compared to NHEJ [116, 117].

The dependence of homologous recombination based repair on the availability of an existing template strand (usually provided by sister chromatids) makes it more useful during S and G2 phases of cell cycle [118]. On the other hand NHEJ is active throughout the cell cycle [119] and is more frequently employed in repair of radiation induced DSBs [118].

Radiation influences balance of proapoptotic and antiapoptotic forces

Exposure to ionizing radiation can act as trigger to cause apoptosis of the cells either through activation of p53 dependent pathway or p53 independent pathways.

**P53 dependent apoptosis**

P53 is a transcription factor that helps in maintaining the integrity of genome by either stopping the progress of cell cycle at one of cell cycle check points or by inducing apoptosis. Inhibition of progress in cell cycle provides time to rectify defects in DNA by one of the DNA repair mechanisms. This decreases the chances of passage of defective DNA to next generation of cells. Cells with severe defects are removed from the pool by p53 induced apoptosis. Under normal conditions the levels of p53 are kept in check by a feedback mechanism in combination with another protein MDM2. P53 increases the levels of mdm2 protein by stimulating its gene. An increase in p53 results in formation of MDM2-p53 heterodimer that prevents it from stimulating the transcription of Mdm2 gene. MDM2 also stimulates ubiquitylation of p53 culminating in its proteolysis. Exposure of cells to genotoxic chemicals or IR results in phosphorylation of p53 and MDM2 by ATM preventing them to join thereby causing accumulation of p53. Higher levels of p53 initiate pathways of apoptosis or cell cycle inhibition [106, 107]. P53 is also activated by phosphorylation by HIPK2 in a quaternary complex consisting of Axin/Hipk2/Daxx/p53 [120].

Activated P53 promotes the transcription of pro-apoptotic genes such as Puma, Noxa and Bax [121-123]. Pro-apoptotic protein Bax forms a heterodimer with antiapoptotic protein Bcl2 (B cell lymphoma 2) and inhibits its activity [124]. Similarly P53 transactivates PUMA that forms a complex with Bcl2 and Bclx resulting in release of

cytochrome c from mitochondria leading to cell death [120]. P53 is also known to promote the transcription of TNF receptors like Fas [125] that in combination with their adapter protein (like Fadd) activates caspases leading to cell death [126].

**P53 independent apoptosis**

In p53 independent pathway the cell membrane initiates a signaling process that leads to production of ceramide. Ceramide (a tumor suppressor lipid) is a secondary messenger that is produced by hydrolysis of sphingomyelin by sphingomylenases and causes apoptosis by stimulating a number of kinases [127-129].

**Antiapoptosis**

DNA damaging agents such as gamma radiation and oxidative stress also activate antiapoptotic proteins like Bcl2 as a protective mechanism [130-133]. Both p53 independent as well as p53 dependent pathways are countered by antiapoptotic proteins such as Bcl2 [134-136]. Proapoptotic proteins also try to neutralize antiapoptotic proteins. For example, p53 promotes the expression of Bax protein that forms a heterodimer with Bcl2 (B cell lymphoma 2) and inhibits its antiapoptotic activity [124]. P53 also transactivates other proapoptotic proteins such as puma and noxa to overcome the antiapoptotic effect of Bcl2. The fate of the cell is therefore determined by a complex interplay between proapoptotic and antiapoptotic forces [137].

## GENE EXPRESSION PROFILING USING MICROARRAYS

Microarrays are tools for analyzing the transcriptional profiles of multiple genes simultaneously. The microarray technology became popular because it enabled us to observe the changes in transcription profiles of many genes induced by differences in treatment (control and treatment), time (time course) and space (different tissues). Although various types of microarrays have been developed over last many years, the basic principle of the technology remains unchanged i.e. they consist of an array of DNA probes (either cDNA or chemically synthesized oligonucleotides) corresponding to all or a desired set of genes from the organism under study. To determine the level of expression of genes in the biological tissue under investigation its mRNA is reverse transcribed to cDNA (complimentary DNA). The cDNA is hybridized to the probes on the microarray. Microarray technology started with two color microarrays. In two color microarrays cDNA from control and the treatment groups were labeled with fluorescent dyes such as Cy3 and Cy5. The two samples were then mixed and hybridized on a single microarray. The proportion of colors emitted by these two dyes indicated the ratio of expression of control and treatment group. In recent years, mostly one color microarrays are being used.

### ONE COLOR MICROARRAYS

In one color microarrays a dye of single color is used to label all the samples. Only one labeled sample is hybridized on each array. The expression levels of genes under different conditions are then determined by comparison of intensity of their probes in the scanned image of microarrays. The use of a single color in these microarrays provides

higher flexibility in the design of experiments and obviates the need to rectify the problem of dye bias associated with the use of two colors. Their downside is the requirement of double the number of arrays than the two color arrays. One color arrays are available from commercial vendors such as Affymetrix and Illumina. In Affymetrix array each gene is represented by 10-12 different probes of 25-base oligonucleotides. The probes are synthesized in-situ by photolithography. In Affymetrix platform each probe is of two types called perfect-match (PM) and mismatch (MM) probes. PM probe is perfect complement of the gene it represents. On the other hand, the MM probe is complementary at all positions except one base in the centre. The MM probes are intended to detect the non specific hybridization. On the other hand in Illumina platform each gene is represented by probes consisting of 50 oligonucleotides attached on beads. Each bead contains many copies of same oligonucleotide. There are 30 copies of each bead. The beads are positioned randomly on the microarray. These miniature beads provide higher packing density of probes and also prevent the position effects on hybridization and scanning. Hence a decoding step based on a molecular address of each bead is needed to determine the gene represented by that bead. Unlike Affymetrix in case of illumina multiple arrays are placed on the same platform that allows their processing in parallel [138, 139]. The image obtained from both of these microarrays are processed by their proprietary imaging software that convert the intensity values associated with various probes into numerical values.

## PREPROCESSING OF MICROARRAY DATA

The data obtained from microarray imaging software cannot be directly interpreted because there are chances of introduction of variation in data due to non biological reasons that include dye bias, background fluorescence, differences in scanning, difference among technicians etc. Dye bias is defined as intensity differences between samples attributable to the differences between dyes instead of transcription differences [140]. The problem is more relevant to two color arrays where it is usually tackled by swapping the dye between samples [141].

Background noise refers to fluorescence from the places on microarray where there are no probes due to unspecific hybridization or contamination. It is believed that signal from the probes may actually be a combination of true signal and the background noise. One approach used to eliminate background noise from the true signal is to subtract a global value calculated for the complete array based on overall background signal. Another approach is to subtract different value from each probe based on background signal in the neighborhood of that probe. Both of these approaches may result in negative expression levels since sometimes the background signals are higher than the actual signal from the probes. Accordingly in some studies background correction is not applied since it leads to increase in noise levels [142].

### Transformation

The difference in intensity of fluorescence from different probes corresponds to different magnitude of expression by genes corresponding to those probes. The probes with higher expression values tend to have higher variance as compared to those with lower

expression value. Hence, it is a common practice to transform the data to mitigate the differences in variances (heteroskedacity) and to normalize it for application of various statistical methods such as ANOVA, regression etc [143, 144]. The most commonly used method is to log2 transform the data. Though log2 transformation reduces the variance at higher magnitudes of expression it increases the variance at very low values especially the values close to background signals. Another popular type of transformation is Variance Stabilizing Transformation (VST) that tries to stabilize the variance across whole range of data [144, 145].

**Normalization**

As stated above the expression levels of transcripts reported by the imaging software may contain systematic biases between arrays that are introduced by non-biological factors such as unequal quantities of starting RNAs, differences in hybridization, biases due to differences in time and space for the experiment etc. Normalization is carried out to remove these systematic biases. Most normalization methods are based on the premise that the level of expression of majority of genes remains unchanged by the treatment. Among the genes that change some are up-regulated and others down-regulated. Hence, average change in the expression level of all the genes on a microarray is expected to be very low [146]. Accordingly, the simplest method of normalization is to standardize the expression values by rescaling them on each array so as to set sum of expression values on each array to a desired common value. This is achieved by dividing expression value of each gene on the array by sum of expression

values of all the genes on that array and multiplying them with the desired common value. Other commonly used normalization methods are discussed below.

MAS5

MAS5 is both a background correction and normalization algorithm developed by Affymetrix. It uses intensity of hybridization from the perfect match as well as mismatch probes present on the Affymetrix microarrays. The algorithm divides the array into 16 regions and the dimmest 2% intensities of each region are used to detect the background noise. The background correction of all the probes is done depending upon their respective distance from centre of these 16 squares. The algorithm ensures that the expression value is positive after background correction. Tukey-biweight algorithm is then applied to the log of background corrected data [147].

Quantile Normalization

Quantile normalization assumes that all samples have the same distribution. Briefly, in this normalization the probe intensities are ordered in each array. The values in each row are then replaced by the average of that row. These average values are then arranged according to original order of the probes [148]. The simplicity and computational efficiency has not only made it one of popular algorithm for normalization of microarray data but also a basis for development of a variety of other normalization methods such as RMA (Robust Multichip Average) [149], GCRMA[150] and RSN (Robust Spline Normalization) [151]. The disadvantage of forcing same distribution on

all the samples and replacing the values with averages is that small differences in expression may be lost.

RMA (Robust Multichip Average)

The RMA normalization is commonly employed for normalization of data from Affymetrix arrays. The RMA algorithm includes three steps including i) background correction ii) quantile normalization and median polishing. For background correction RMA uses the perfect match probes and ignores the mismatch probes of Affymetrix arrays. The background noise is believed to be proportional to intensity of the perfect match probes. The algorithm assumes that the true signal is exponentially distributed and the background noise is normally distributed. Quantile normalization is done on the background corrected expression values. There after median polishing is performed so that the medians of expression values for each row and column is close to each other [149].

GCRMA

GCRMA is a variant of RMA algorithm where sequence information (differential bond energy between GC pair and AT pair) of the probes is used to estimate the affinity of the probes to nonspecific binding. The background correction is done according to this affinity. After background correction quantile normalization and median polishing methods of RMA are used [150].

RSN (Robust Spline Normalization)

Robust spline normalization combines of quantile normalization with ability of loess normalization to fit the data continuously [151]. The quantile normalization is simple and computationally efficient algorithm that preserves the rank order of genes. But it forces all samples to same distribution and also results in loss of ability to distinguish between small changes in expression. Loess or spline normalization on the ether hand does not preserve the rank of genes but fits the data continuously [152].

### EXTRACTION OF GENE CO-EXPRESSION NETWORKS

A variety of methods have been employed for extraction of gene co-expression networks using transcriptomics data obtained by microarrays. These include Boolean networks, Bayesian networks, clustering and relevance networks.

#### Boolean networks

Boolean networks assume that genes are present in two discrete states i.e., on and off. Due to this assumption the Boolean models are computationally very efficient. Lang et al. developed a mutual information based algorithm to extract gene regulatory networks from gene expression data represented in a binary state of on and off [153]. The main drawback of Boolean networks is the lack of intermediary stages of gene expression. Binary representation of gene expression is a biologically untenable assumption.

## Bayesian networks

Bayesian networks use directed acyclic graphs to establish probabilistic relations among the genes. They establish dependence structure between genes based on their expression levels. Thus the edges between genes represent conditional dependencies among them. The ability of Bayesian methods to determine the causality of interaction is a major advantage of these methods [154]. On the other hand learning Bayesian networks is computationally very expensive. Hence they either employ heuristic algorithms or prior knowledge of networks to make them efficient[155].

## Clustering Microarray Data

Cluster analysis is a set of data mining methods used to aggregate observations of the data into groups based on some criterion for evaluating similarity or dissimilarity between the observations. For analysis of gene expression profiles clustering is a commonly employed method to group genes and/or samples. The genes / samples that are similar in their expression profiles are placed in same or closer clusters. The dissimilar genes/ samples are placed in different clusters. The distance between clusters is proportional to the dissimilarity between members of the clusters. The measures used to determine the distance between clusters include Euclidean distance, Manhattan distance, correlation (Pearson and Spearman), Kullback–Leibler divergence [156] etc.

Euclidean distance is the shortest straight line Pythagorean distance between two points in an n-dimensional space. Manhattan distance is the walking distance between

two points. Whereas Euclidean distance between two points in an n-dimensional space can be tracked by simultaneous movement among all dimensions the Manhattan distance can be traced by movement along one dimension at a time.

Correlation measures the similarity or dissimilarity in trends in expression of genes. The most commonly used measure of correlation is the Pearson's correlation. Another measure of correlation is the non-parametric Spearman's rank correlation. Like Pearson's correlation it varies between -1 and 1 but instead of using the values of observations the relative ranks of the observations are used for calculating it.

## SUPERVISED AND UNSUPERVISED CLUSTERING

Supervised clustering methods require prior knowledge of number of clusters in the data. The clustering algorithms under this category then allocate the genes to one of the clusters. Unsupervised clustering does not require prior knowledge of number of clusters. The number of clusters is derived from the data itself [157].

Popular clustering algorithms that have been applied to clustering of microarray gene expression profiles include K-means [158], hierarchical clustering [159] and Fuzzy clustering [160, 161]. .

## K MEANS CLUSTERING

In K-means clustering the number of clusters 'K', is predefined by the user. Each of the K clusters is initiated by one observation as its mean. Thereafter following steps are followed [162].

a. Each observation is assigned to the cluster whose mean is closest to that observation.

b. After assignment of all the observations the means of each cluster are recalculated from the current members of that cluster.

c. The observations are then reassigned to the clusters based on their distance from the new means.

Steps ‘b’ and ‘c’ are iterated till distance of observations from the means of the clusters is minimized.

K-means clustering is prone to effects of outliers since mean can be influenced by even a few outlier observations. To overcome the influence of outliers, K-medoids clustering uses the observation closest to mean as a reference point for aggregation. The distance of other observations from this reference point is used as a criterion for membership of this cluster [163]. Though simple in their implementation the main drawback of the K-means and K-medoid clustering is the requirement of an appropriate ‘K’ [162]. Especially in the case of gene expression profiles it is difficult to prejudge the number of groups amongst thousands of genes.

FUZZY CLUSTERING

Most clustering algorithms divide the data into distinct non-overlapping clusters. On the other hand gene networks are all interconnected as one gene may participate in more than one biological function (pleiotropy). Fuzzy clustering tries to resolve this by allowing observations (genes) to be associated with more than one cluster. Fuzzy-C means clustering is the most popular of these clustering algorithms. It is similar to K-

means clustering except that it assigns probabilities (varying from 0 to1) that determine the membership of the objects to various clusters [160, 161]. The standard K-means clustering assigns every gene to a single cluster. On the other hand fuzzy-C means can assign a gene to multiple clusters with varying degrees of membership [161]. To determine the extent of overlaps this algorithm requires specification of a fuzziness parameter (>1) in addition to parameter K for the required number of clusters. Values of fuzziness parameter closer to one imply distinct and isolated clusters. Higher values result in higher overlap amongst the clusters. The algorithm also requires a parameter for convergence of clustering. Pre-specification of all these parameters and their tuning is a downside of this algorithm.

HIERARCHICAL CLUSTERING

Hierarchical clustering arranges all the objects in a tree like structure called dendrogram without pre specification of number of clusters. After creation of dendrogram the number of clusters can be decided by choosing an appropriate level to cut the tree [159, 164]. The main advantage of this approach is the ability to visualize complete dataset and then determine the number of clusters based on both statistical methods as well as the domain knowledge. For example, in microarray data both the genes and samples can be represented as dendrogram that can be cut at specific levels to place them in appropriate clusters. Hierarchical clustering of data can be achieved by following both top down and bottom up approach. In top down approach (dissociative) all the objects are initially placed in a single cluster. Thereafter the clusters are successively divided till the number of clusters is equal to number of objects being clustered. In bottom up

approach (agglomerative) each object is initially placed in a separate cluster. These clusters are then repeatedly merged and subsumed into bigger clusters until a single cluster encompassing all the objects is achieved. Various mathematical measures such as correlation (1-r), Euclidean distance, Manhattan distance, mutual information etc. could be used as criteria for evaluating the distance among objects/clusters. Based on the reference point in the clusters from where the distance is measured various algorithms of hierarchical clustering are used. These include average linkage (average of distances between all points in one cluster with all points in the other cluster), centroid linkage (Euclidean distance between centroids of two clusters), complete linkage (distance between farthest points in two clusters) and single linkage (distance between closest points in two clusters). Hierarchical clustering has often been used to produce two-way clustering of expression data i.e. simultaneous clustering of genes and conditions that enables visualization of gene clusters under different experimental conditions [165, 166].

**Relevance networks**

Relevance networks start by employing measures of similarity such as correlation and mutual information amongst all the gene pairs in a transcription profile. This pair-wise similarity network is filtered by employing a suitable threshold to retain only strong associations [167-169]. Extraction of biologically meaningful information from a large network representing pair-wise association among thousands of genes requires efficient computational algorithms. Graph algorithms provide a means to extract dense and highly connected regions from these networks[168]. The main advantage of relevance

networks is their ability to allow multiple relations among genes. They can handle positive as well as negative correlations between genes and easily combine information from data of diverse types [167, 170, 171].

**Graph algorithms**

Biological networks such as metabolic pathways, protein interactions and gene regulation networks can be easily be represented as graphs where biological entities like genes, metabolites and proteins etc. are represented as vertices and the relationships between them is represented as edges. Accordingly, graph theory has been employed in various fields of computational biology such as prediction and comparison of protein structure [172-174], prediction of reactions in metabolic pathways [175], representation of functional relationships between genes as in gene ontology (GO) database[176-178], creation of bioinformatics tools [179-181] and for extraction of dense and interconnected portions of biological networks such as cliques, paracliques, hubs, bipartite networks, network motifs. Our group has also developed [182-186] and applied [169, 179, 181, 187, 188] graph theory algorithms for extraction of putative gene networks from gene expression data.

CLIQUES AND PARACLIQUES

A Clique is a subgraph such that every pair of its vertices is connected by an edge between them. In case of gene expression networks a clique models a group of tightly co-regulated genes that may be participating in a common biological pathway or are influenced by a common stimulus in a case control study. For extracting cliques from microarray data, first the gene expression profile is represented as a complete graph.

Each gene in the graph is represented as a vertex and is connected with every other gene on the graph by an edge. The edges between the vertices represent the strength of relation between the genes [169]. As in the case of clustering methods, the strength of relationship between genes can be evaluated in terms of distance measures such as correlation (Pearson or Spearman), Euclidean distance, Mutual information etc. A threshold is applied on the complete graph to retain only highly correlated genes and the edges connecting them. Various methods [169, 183, 185] have been proposed to threshold the graph but most commonly a high pass correlation threshold [169] typically in the range of 0.8 to 0.875 is used to remove edges of low correlation. The filtered, un-weighted and undirected graph is used to extract maximal cliques using fixed parameter tractability (FPT) algorithms. To imitate the natural biological networks cliques can be relaxed by allowing it to miss a few edges. These relaxed cliques are termed paracliques. Paracliques try to compensate the inherent noise in the microarray data and stochastic nature of biological processes [169].

HUBS AND BETWEENNESS CENTRALITY

Hubs are nodes of high degree found in a graph. In biological networks they reflect genes, proteins or metabolites that interact with and influence many other genes or proteins. Common examples of hubs include transcription factors, kinases, transferases etc. that are known to have many targets. Hub genes or proteins are critical for function of biological networks [169, 189, 190]. Though biological networks are robust, the disruption of hubs is more likely to be lethal to organism, often termed as central lethality [191].

Another set of biologically significant vertices in a graph are those with high 'betweenness centrality'. A vertex that is in the shortest path of a large number of other vertex pairs is said to have high betweenness centrality. Proteins with high betweenness are believed to be functionally and evolutionarily significant [192].

NETWORK MOTIFS

Network motifs are patterns of interconnections that are present in biological networks at a frequency higher than that possible by random chance. Directed graphs were used to identify such motifs in gene regulation networks [193].

BIPARTITE GRAPHS

A bipartite graph is an undirected graph consisting of two subsets of vertices U and V. Each edge E of this graph connects a vertex in U with a vertex in V. Bipartite graphs have been employed for extracting relationships between genes and phenotypes [181, 187], assembling and identifying proteins from peptides generated by mass spectrometry [194], matching regions of brain in different brain atlases [195], identifying drug and target interaction [196], associating protein domains with proteins [197], integrating drugs with genetic information for predicting drug and target interaction [198] and for determining evolutionary significance of metabolites and metabolic reactions [199].

## Multiple test correction

Large scale transcriptomics experiments typically test the affects of change in condition, strain or time on the expression levels of thousands of genes simultaneously. The

change in expression levels of genes are statistically tested by repetitive application of algorithms such as ANOVA (analysis of variance). The repetitive hypothesis testing increases the probability of rejecting a null hypothesis by random chance resulting in increase in false positives. Increase in false positives due to multiple testing is rectified by application of multiple test correction methods that either control Family Wise Error Rate (FWER) or False Discovery Rate (FDR).

FWER

FWER methods control the probability of false positives among all hypotheses tested. Bonferroni correction controls FWER by dividing the desired alpha by number of tests[200]. In case of microarray and other high throughput methods it is too restrictive since the number of tests is usually in thousands. This makes the desired alpha too small and increases the number false negatives.

BENJAMINI AND HOCHBERG FDR

Benjamini and Hochberg introduced the concept of false discovery rate (FDR) for controlling false positives due to multiple hypothesis testing. The FDR method controls the probability of false positives among the tests declared significant. Benjamini and Hochberg method takes into consideration the number of hypothesis tested and relative rank of pvalues amongst the population of pvalues obtained from all the tests[201].

Q-VALUE

Storey introduced another FDR based method for multiple test correction that introduces q-value (a Bayesian posterior of p-value)[200]. Q-value estimates the FDR by taking into consideration proportion of true null hypotheses. The proportion of true null hypotheses is estimated from the distribution of pvalues by assuming that the pvalues are uniformly distributed [202].

**Gene Ontology Database**

Gene ontology (GO) is a tripartite database consisting of controlled vocabulary that defines the functions of genes. Three aspects of gene function that are included under the GO are:

i) cellular component i.e. physical space in the cell where the gene is expressed,
ii) biological process in which the gene is known to participate and the
iii) molecular function of the gene [176].

A directed acyclic graph (DAG) is used to represent each aspect of gene function. Terms relating to gene function are arranged in a hierarchy in each of DAG. The root node (Cellular component, Biological process or Molecular function) is the highest term in hierarchy for that DAG. The terms become more and more specific as we move away from the root node [176]. Each term in DAG may be a descendent of multiple parent terms. Likewise, each term may have multiple child terms. Similarly, a gene may be associated with multiple terms and a term may be associated with multiple genes [176]. GO not only provides uniformity of terminology for description of gene function but

also enables automation of functional analysis of groups of genes that forms the basis of many studies of gene enrichment.

Usually gene ontology database is employed to detect GO categories over represented in a list of genes derived by one of the clustering or gene network extraction methods. Statistical methods such as hypergeometric test, Fisher's exact test and Chi-square test are employed to detect the GO categories [203-205] that are over represented in the queried gene set. Another method that is frequently employed to interpret functional similarity among a group of genes is to determine the semantic similarity of the GO terms corresponding to the genes in the cluster. A high correlation between gene expression and semantic similarity of GO terms annotating those genes indicates functional similarity of the genes [206].

# CHAPTER 2 : REVEALING PUTATIVE GENE NETWORKS PERTURBED BY LOW DOSE IONIZING RADIATION USING DIFFERENTIALLY CORRELATED GRAPH GENERATED BY A TWO STAGE STATISTICAL FILTER

(This manuscript will be submitted for publication in PLoS ONE with following authors Sudhir Naswa, M.A. Langston, Rachel M. Lynch, Suchita Das, Arnold Saxton, B.H. Voy)

## ABSTRACT

Strong stimuli elicit changes in gene expression that can be detected with statistical methods for differential expression such as ANOVA. In contrast, more modest stimuli induce changes in gene expression that may confer significant biological impact but are difficult to detect by ANOVA, especially in the presence of background genetic variation. We hypothesized that capturing mutual relationships of genes by their differential co-expression (correlation) would uncover effects of modest biological stimuli that might be missed by traditional differential expression approaches. To test this hypothesis we used a two stage statistical filter to generate a differential correlation graph (DCG) that identifies differences in gene networks between control and treatment groups. We applied this approach to microarray data produced from five strains of inbred mice exposed to a single low dose of ionizing radiation. The exposure level (10 cGy) is consistent with exposures obtained during increasingly common diagnostic CT scans, which may increase cancer risk. RNA was extracted from spleen of control and irradiated mice 24 hours after exposure and profiled using the Illumina microarray platform. After producing the DCG, random selection tests were used to identify statistically significant network hub genes with higher connectivity than would be expected by chance. Whereas differential expression methods identified few differences between control and irradiated mice, differential co-expression revealed gene networks highly enriched with genes implicated in radiation response, DNA damage repair, apoptosis and cancer with hub membership enriched in members of the

BRCA complex. These findings illustrate the value of differential correlation for extracting the biological response to subtle environmental stimuli.

## INTRODUCTION

Technologies for profiling the expression of genes are getting faster, cheaper and efficient enabling us to generate huge amount of data in a very short time. Availability of high throughput methods has facilitated development of strategies for extraction of information about inter-relationships of genes in biological networks. Genetic networks maintain a homeostatic relationship among genes. Maintenance of this steady state relationship among genes is critical for the survival of an organism. A powerful external stimulus that disrupts vital parts of a gene network or causes its total breakdown may lead to disease or death. Such strong external stimuli also effectuate relatively higher changes in magnitude of expression that can easily be detected with statistical methods for differential expression such as ANOVA. On the other hand biological organisms survive most sub-lethal external stimuli. Relatively lower changes in gene expression by such stimuli make it difficult to detect by differential expression methods especially because of lower sample sizes and hence statistical power. This difficulty is often compounded by differences in genetic background amongst the subjects of treatment.

We believe that genes readjust their mutual relationships after exposure to external stimuli and these readjustments can be captured by differential co-expression (differential correlation) among genes. Differential co-expression methods used for extraction of gene networks from transcription profiles have been classified by Tesson et al. [207] into two types: targeted [208-211] and untargeted [169, 207, 212-215]. In targeted approach the emphasis is on finding differential co-expression among

predefined clusters of genes whereas untargeted methods cluster genes based on differential co-expression between them.

Our group has developed [182-186] and applied [169, 181, 187, 216] graph theory based methods for extraction of co-expressed as well as differentially expressed putative gene networks from gene expression profiles. Here we use an untargeted differential correlation method based on graph algorithms for extraction of differentially coexpressed gene networks. The previously proposed untargeted method by Tesson et al. [207] uses a variable tuning parameter for determining differential correlation between genes under different conditions. They recommend choosing a tuning parameter that either makes the gene networks scale free or maximizes the difference in correlation between the genes under different conditions. Similarly, Van Nas et al. use a tuning parameter that makes the gene correlation network scale free [214]. In another untargeted method Southworth et al. use a fixed value (0.75) of hierarchical clustering height parameter to cluster the genes based on differential correlation. Cho et al.[208] highlight inability of untargeted method to detect differential correlations to justify the targeted method developed by them. In their untargeted method they used a single cutoff based on Bonferroni multiple test correction to determine the significance of correlation pvalues for control, treatment and difference in correlation between control and treatment. They use too strict a cutoff because of two reasons i) Bonferroni correction is known to be very conservative and ii) the number of tests for detecting difference in correlations should be much less than the number of tests for determining the significance of correlation under each condition since comparison of insignificant

correlations is unnecessary. Moreover, just comparing the gene pairs that are significant under both the conditions misses a lot of information. The knowledge derived from comparison of gene pairs that are uncorrelated under one condition and correlated under the other is equally vital. Altay et al.[213] used a differential co-expression method that needs specification of user defined parameter based on rank value of mutual information of each gene with its neighbors. Yu et al. [215] use q value to filter the nonsignificant correlations and there after use maximum or log correlation values to determine differentially correlated genes. Here we use a two stage statistical filter by employing multiple test corrections at each stage to generate a differentially correlated graph since the possibility of false positives due to multiple testing arises at two stages. One at the stage of deciding significance level of correlation under two conditions and second at the stage of comparison of correlation of gene pairs under different conditions. We use q-value for multiple test correction at each of these stages. Q value controls the false discovery rate but is not too restrictive like family wise error rate correction method by Bonferroni [202].

We begin by generating a graph that has edges between pairs of genes that exhibit statistically significant multiple test corrected correlation under at least one of two experimental conditions and multiple test corrected differential correlation under different experimental conditions. Thus our method not only ensures that the correlation coefficients are significant but also the difference between them is significant and FDR corrected. We use graph algorithms to extract dense regions of the graph such as connected components, cliques and hubs from this differentially correlated graph.

Graph theoretical methods for mining gene networks are advantageous over other clustering methods because they do not impose restriction on topology of network and can simulate biological reality of overlaps and interconnections between biological pathways. They can easily be used for targeted as well as untargeted methods and can be adapted to integrate biological data of diverse types such as gene expression, SNP, proteomics and phenotype [169, 187, 217, 218].

We employ aforementioned differential co-expression method to determine the effects of Low dose ionizing radiation (LDIR) on the gene networks. Ionizing radiation up to 10cGY (centi Gray) is classified as low dose radiation by US department of energy. Sources of exposure to LDIR include medical imaging techniques like CT-scan, certain occupations (nuclear waste clean-up, nuclear power plant) and non lethal nuclear accidents. There is evidence of harmful [219] as well as beneficial effects [220] of LDIR. Annotation of the extracted networks with Gene Ontology, Kegg pathways and literature search show that the gene networks extracted by our method are highly enriched with radiation sensitive genes and processes. Whereas differential expression method using ANOVA detected very little effect of LDIR, differential co-expression revealed gene networks highly enriched with genes implicated in radiation response, DNA damage repair, apoptosis and cancer.

# RESULTS AND DISCUSSION

## GENERATION OF DIFFERENTIALLY CORRELATED GRAPH

We used gene expression data obtained from spleen cells of five different strains (129S1/SvImJ, NOD/LtJ, CBA/J, CAST/EiJ and WSB/EiJ) of genetically diverse mice as a part of our ongoing effort to elucidate gene networks influenced by LDIR following a systems genetic approach. Illumina microarrays were used to measure the expression profiles of control (sham-irradiated) and radiated (treatment) mice. The mice were irradiated with single low dose of ionizing radiation (10 cGy of gamma radiation) and the spleen cells were harvested for expression profiling on illumina microarrays. The microarray data was preprocessed using variance stabilized transformation followed by robust spline normalization. Following the belief that stimulus induced perturbations in gene networks are reflected in readjustment of correlations amongst the genes we employed graph algorithms to investigate the differences in inter-gene correlations between control and radiated mice. We focused on changes in pair wise linear correlations between genes and used Pearson correlations to extract gene networks that correlate differentially after exposure to radiation. The gene networks were initiated by creating a gene to gene correlation matrix for both the control and radiated mice. After creation of correlation matrices, we used a two stage statistical filtering process to generate a differential correlation graph (DCG) as depicted in Figure 2-1. First we obtained a multiple test corrected two tailed p-value for significance of correlation coefficients of gene pairs for both control and radiated data. Multiple test correction was done by controlling false discovery rate by using Q-value [202]. Q-value cutoff of 0.05

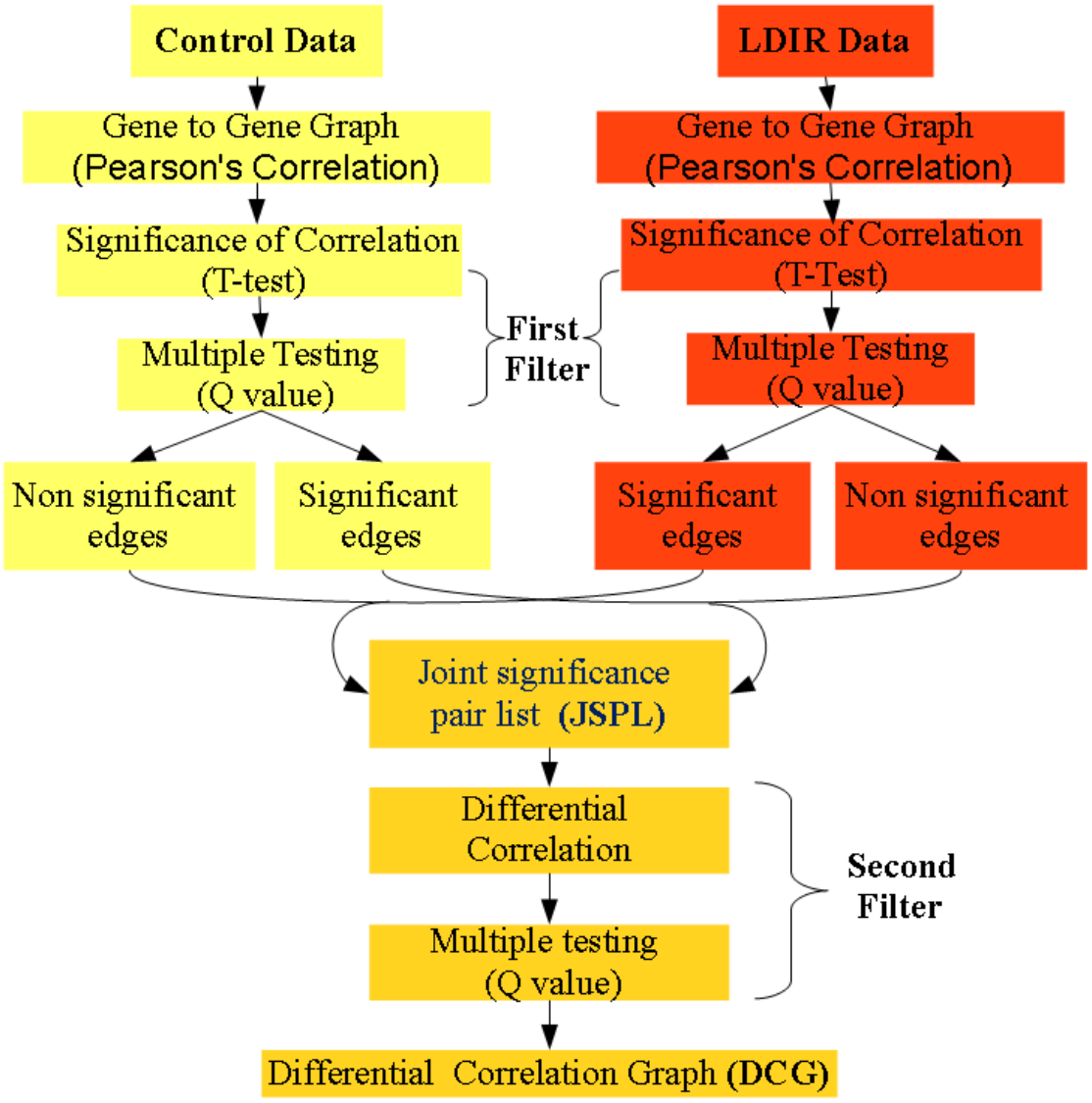



**Figure 2-1: Generation of differential correlation graph (DCG) using two-stage statistical filtration process.**

was used to get multiple test corrected p-values of 0.00276 and 0.00667 for control and radiation data respectively. A joint significance pair list (JSPL) was then created by keeping all the gene pairs that were significantly correlated in at least one of two conditions. A second filter was applied to retain only those gene pairs of JSPL that had statistically different correlations for control and radiated mice. Difference between control and radiation correlation coefficients of a gene pair was determined by first transforming them to Z-scores using Fisher's z transformation and then testing the significance of difference between the two Z scores against a null hypothesis of no difference. Q value was used to get a multiple test corrected p-value for differences in correlation coefficients. Only those pairs of JSPL were retained that had a multiple test corrected p-value of less than 0.00051 corresponding to a q-value of 0.05. The gene pairs of JSPL that survived the double filtration process were converted into a differentially correlated graph (DCG) by representing each gene with unique vertex. An edge was placed between genes of each pair in the double-filtered JSPL. Effectively, the DCG consisted only of vertices (genes) that survived double filtration process. An edge was placed between two vertices if and only if they met the following two conditions:

i) Correlation between the two vertices was statistically significant in control and/or radiated mice.

ii) The difference between the correlations of the two vertices under control and radiation was statistically significant.

This differentially correlated graph had 20438 edges and 5707 vertices with mean vertex degree of 7.162. We plotted distribution of edge weights (absolute difference in gene pair correlation between control and radiated mice) in the DCG (Figure 2-2). The mean and median edge weights were 1.18 and 1.21 respectively. Most (99.5%) of these edge weights were greater than 0.57. A shift of 0.57 and above in the absolute value of correlation coefficient suggests readjustment of mutual relationships among genes of DCG after exposure to radiation.

As depicted in Figure 2-2 the exposure to radiation causes statistically significant shifts in correlation coefficient between gene pairs suggesting emergence, disruption and inversion of relationships between them. After exposure to LDIR 11587 gene pairs exhibited 'correlation emergence' (non significant correlation in control becomes significant) and 8634 gene pairs showed 'correlation disruption' (significant correlation in control becomes non significant) events. LDIR also resulted in inversion of correlation relation between genes in 217 pairs (significantly positive correlation becomes significantly negative after exposure to LDIR or vice versa). 113 of these LDIR induced inversions were positively directed (negative correlation of control becomes positive) and 104 were towards negative direction (positive correlation of control becomes negative). The distribution of edge weights for both the correlation emergence as well as correlation disruption events closely follows the overall distribution of edge weights in the DCG. As expected the mean and medians of edge weights for the more drastic inversion events are higher (Table 2-1).

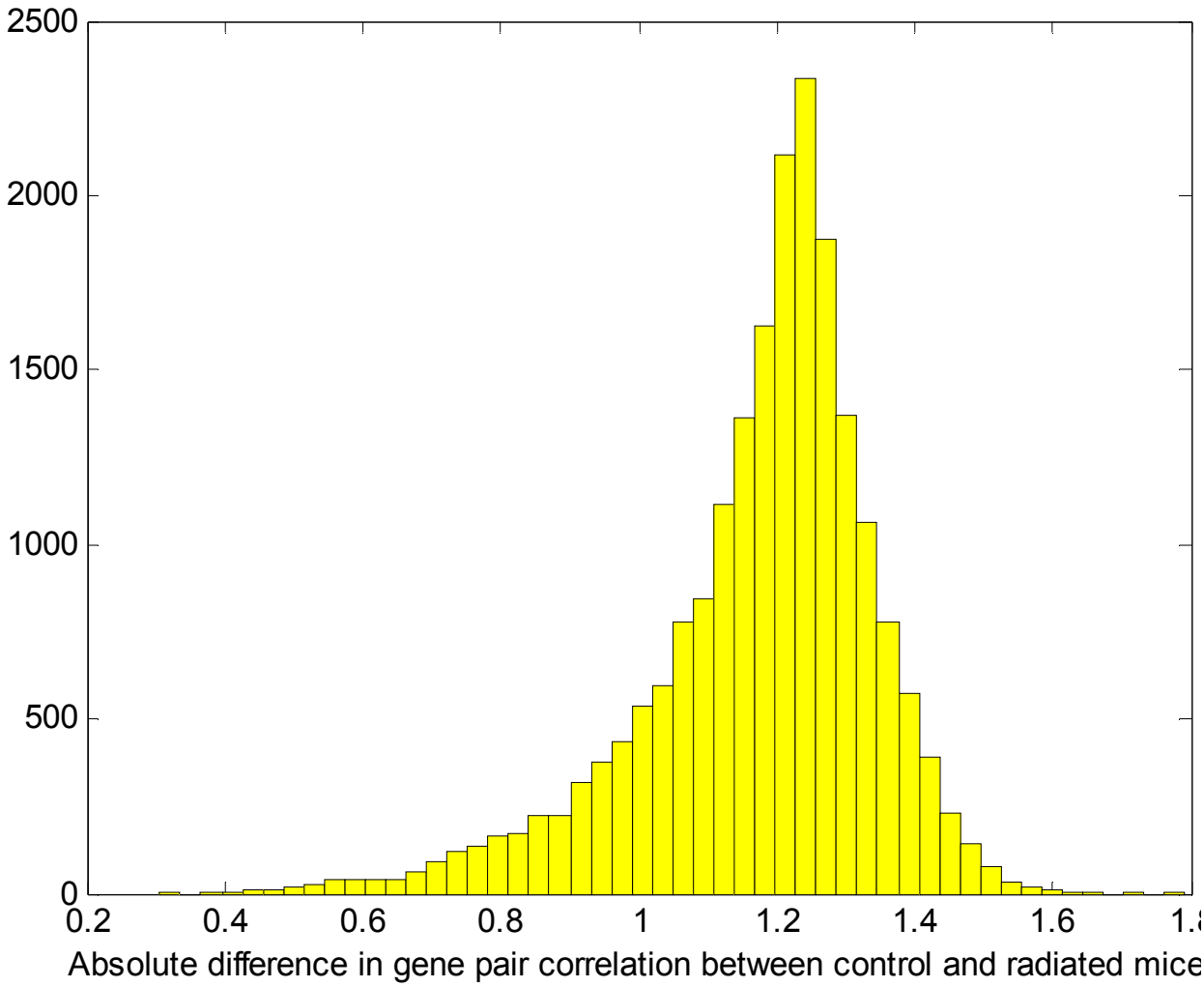



**Figure 2-2: Frequency distribution of edge weights in differentially correlated graph.**

**Table 2-1 Classification of edges and distribution of their edge weights in the differentially correlated graph.**

| Edge type | Number of | | Percentiles of Edge Weights | | | Mean Edge weights |
|---|---|---|---|---|---|---|
| | Vertices | Edges | 25th | 50th | 75th | |
| All edges | 5707 | 20438 | 1.1 | 1.21 | 1.28 | 1.18 |
| Correlation emergence | 4508 | 11587 | 1.12 | 1.22 | 1.29 | 1.19 |
| Correlation disruption | 3802 | 8634 | 1.08 | 1.19 | 1.26 | 1.16 |
| Correlation inversion | 296 | 217 | 1.45 | 1.48 | 1.52 | 1.49 |

### ENRICHMENT ANALYSIS OF LARGEST CONNECTED COMPONENT OF DCG

As a first glance into DCG we extracted its connected components. A connected component is a subgraph in which each vertex can be connected to every other vertex through a path consisting of edges and other vertices in the graph. The largest connected component of the DCG consisted of almost all the edges (20389/20438= 99.76%) and vertices (5618/5707 = 98.44%) in the DCG. The interconnectedness of almost all the genes underscores that at systems level changes in correlation of one pair may lead to changes elsewhere in the network. GO analysis of this connected component indicated its extreme richness in genes participating in various metabolic processes (FDR corrected p-value <E-100). Surprisingly it was also highly enriched in radiation sensitive genes with an FDR corrected p-value (fPval) of 2.37 E-9. 64 genes (Tables A2-1 & A2-2) in this connected component belonged to the GO category "response to radiation". 26 out of these 64 genes were annotated with GO category "response to ionizing radiation" with an fPval of 8.63 E-6. 31 genes were annotated with the GO category response to UV radiation with an fPval of 1.6728E-8. Enrichment of radiation sensitive genes in the largest connected component of the DCG suggests a shift in relationships of these genes with their neighbors consequent upon exposure to LDIR. This largest connected component was also enriched in other GO categories reflecting processes relevant to radiation response such as apoptosis, response to biotic and abiotic stimuli, response to DNA damage stimulus, response to oxidative stress, signal transduction, immune system process, lymphocyte mediated immunity, cellular homeostasis, protein synthesis and transport, DNA metabolism and Cell cycle.

Enrichment of genes pertaining to a large number of biological processes suggests that LDIR effects change in relationships of many genes. To delve into the influence of LDIR on gene networks, we decomposed the DCG to extract hubs and maximal cliques.

### EXTRACTION OF HUB NODES FROM THE DCG

A hub in a graph is a vertex with high degree (many neighboring vertices). Extraction of hubs in DCG enabled us to concentrate on radiation sensitive genes that changed their correlation with a large number of other genes after exposure to LDIR. We performed 1000 random selection tests to determine the size of hubs that had a significant probability of formation in the DCG. In each test 5000 genes were randomly selected from DCG and probability of vertex degree of a randomly selected gene was determined. A randomly selected gene never had a vertex degree greater than 29 at a p-value of 0.05 in the 1000 runs. Therefore, a vertex degree of 30 was used as a cutoff for selection of hub genes. We focused on hubs formed by radiation sensitive genes. Two of the 64 radiation sensitive genes viz. Bcl2 (B-cell leukemia) and Rnf168 had vertex degrees of 47 and 36 respectively.

### Bcl2 (B cell lymphoma 2) Hub

Bcl2 is a radiation sensitive anti apoptotic gene implicated for its role in cancers of breast, lungs, thyroid, oropharynx, ovaries and prostrate [221-226]. It protects the cells from the effects of ionizing radiation [131] , oxidative stress [130] and facilitates DNA damage repair [132]. Many genes of this hub are involved in processes relevant to radiation response such as nucleic acid metabolism, DNA replication, B and T cell

lineage commitment and apoptosis. These include Il33 (interleukin 33), Fkbp5 (Fk506 binding protein 5), Rpa1 (replication protein A1), Tubb2c (tubulin beta 2C), Tsc22d1 (Tsc22 domain family member 1) and Mfng (Mfng O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase).

Il33 has been associated with lung [227] and skin [228] cancers. UV radiation is known to induce it to higher levels in skin cancers [228]. It has also been reported to increase the expression of Bcl2 in liver of BALB/c mice [229] and neonatal cardiomyocytes of rat [230]. Our data indicates that LDIR changes the correlation of Bcl2 and Il33 from negative ($r = -0.83$) to slightly positive (but statistically not significant with $r = 0.27$) (Figure 2-3, a and b). It could be speculated that LDIR tutors these two genes towards a better cooperation in expectation of a higher dose of radiation.

Fkbp51 aka Fkbp5, another gene of this hub is a member of group of genes synthesizing highly conserved proteins immunophilins. Fkbps' are known to play role in protein folding/transportation, receptor signaling, T-cell activation, apoptosis and modulation of oxidative stress [231] . Fkbp5 is also implicated in providing resistance to apoptosis in melanoma cells by activating Nf-κB (nuclear factor of kappa B) in response to ionizing radiation [232]. The correlation between Fkbp5 and Nf-κB (Nf-κB1) changed ($p\text{-value}<0.004$) from slightly negative ($r=-0.413$, $p\text{-value} =0.12$) to positive ($r=0.635$, $p\text{-value} =0.01$) to confirm their cooperative behavior in response to LDIR. It has been reported that activation of Nf-κB induces expression of Bcl2 [233].

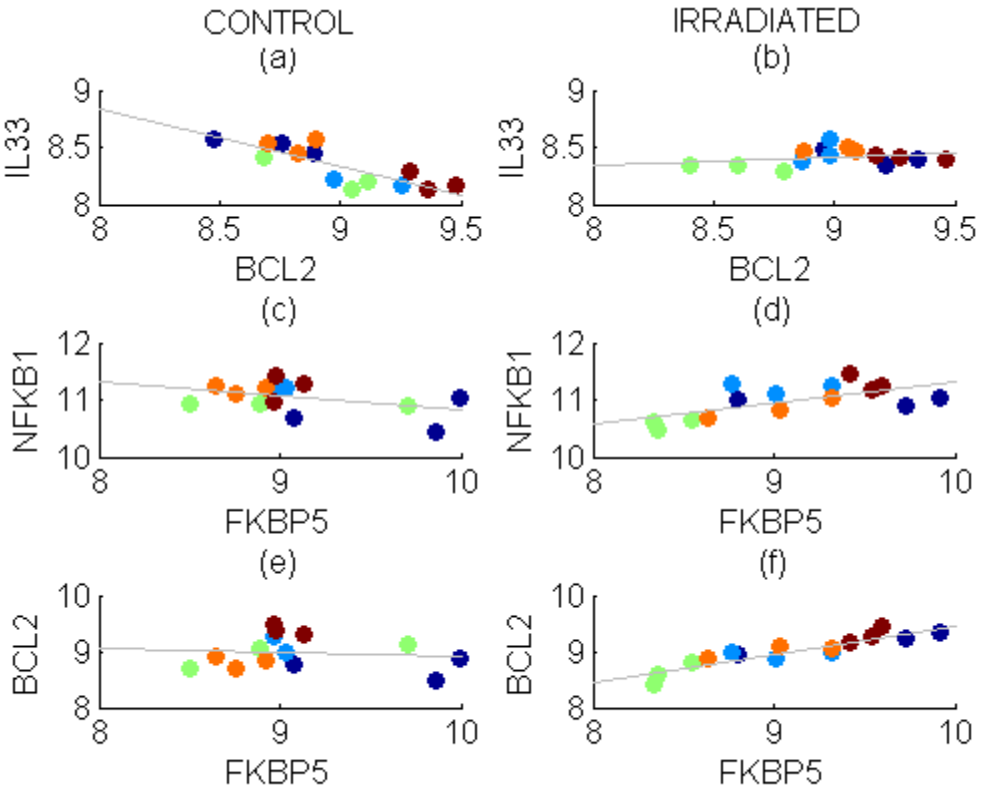



**Figure 2-3: Change in correlation of genes after irradiation in Bcl2 Hub** (Each colored dot represents mice of a particular strain).

The correlation between Bcl2 and NF-κB1 increased slightly from 0.54 (p-value<0.05) to 0.66 (p-value<0.008). Since Nf-κB is a transcription factor that is activated and inhibited by many factors and in turn regulates a large number of genes the slight change in correlation with one of its target is not surprising. Increased correlation of Fkbp5 with Nf-κB and emergence of very high positive correlation between Bcl2 and Fkbp5 (r=0.904) after LDIR treatment (Figure 2-3 c-f) may therefore suggest their increased cooperation to counter the apoptotic effects of LDIR. Bcl2 is known to interact with another immunophilin FKBP38, and provide resistance to apoptosis [234].

Another gene of Bcl2 hub, Rpa1 is a conserved protein that is active during DNA replication, repair and recombination processes [235] that are quite common in post radiation damage to genetic materials. It is also involved in DNA damage double strand repair and prevention of apoptosis following exposure to IR [236]. Polymorphisms in this gene are associated with cancer [237]. Mice with heterozygous mutations of Rpa1 are reported to develop lymphoid tumors and defects in repair of DNA double strand breaks whereas its homozygous mutations result in embryonic lethality [238]. Besides its presence in this hub Rpa1 was also a part of a maximal clique of four genes that is discussed below in a separate section.

Bcl2 hub also includes other cancer associated genes such as Tubb2c, Tsc22d1 and Mfng. Expression levels of Tubb2c (tubulin beta 2C) have been reported to vary in tumor cells [239]. Tsc22d1 (Tsc22 domain family member 1) is a transcription factor believed to induce apoptosis in cancer cells [240, 241]. High doses of ionizing radiation are known to up regulate it in human keratinocytes [241]. It has also been associated

with formation of spontaneous pulmonary adenoma in female mice [242]. Mfng (Mfng O-fucosylpeptide 3-beta-N-acetyl glucosaminyl transferase) is expressed in human tumor derived cell lines and its over expression in 3T3 cells made them tumorigenic [243]. Changes in correlation of antiapoptotic and proto-oncogenic BCL2 gene with other genes involved in apoptosis, cancer and DNA damage responses indicates their concerted response to face the challenge posed by LDIR.

**Rnf168 (ring finger protein 168) hub**

Rnf168 is a radiation sensitive 'ubiquitin ligase' that participates in DNA damage repair by facilitating Rnf8 mediated histone ubiquitylation [101, 244-246]. Radiation induced double strand break foci were salvaged by its ectopic expression in lymphoblastoid cells with homozygous nonsense mutation in Rnf168 gene [247]. Loss of Rnf168 in mice results in increased sensitivity to radiation, immunodeficiency and genomic instability. Rnf168 deficiency along with p53 inactivation enhanced tumor formation in mice [248]. A member of this hub, Cbx5 (chromobox homolog 5) also known as HP1a (Heterochromatin protein 1-alpha), is recruited to DNA damage site after UV and IR exposure and its deficiency is associated with increase in DNA damage and genomic instability [249-251]. The emergence of correlation between Rnf168 and Cbx5 ($r = -0.378$ for control and $r = 0.785$ after LDIR) confirms similarity in their response to radiation (Figure 2-4). Two other members of this hub, Ccng1 (Cyclin G1) and Gltscr2 (Glioma tumor suppressor candidate region gene 2) participate in p53 response to radiation.

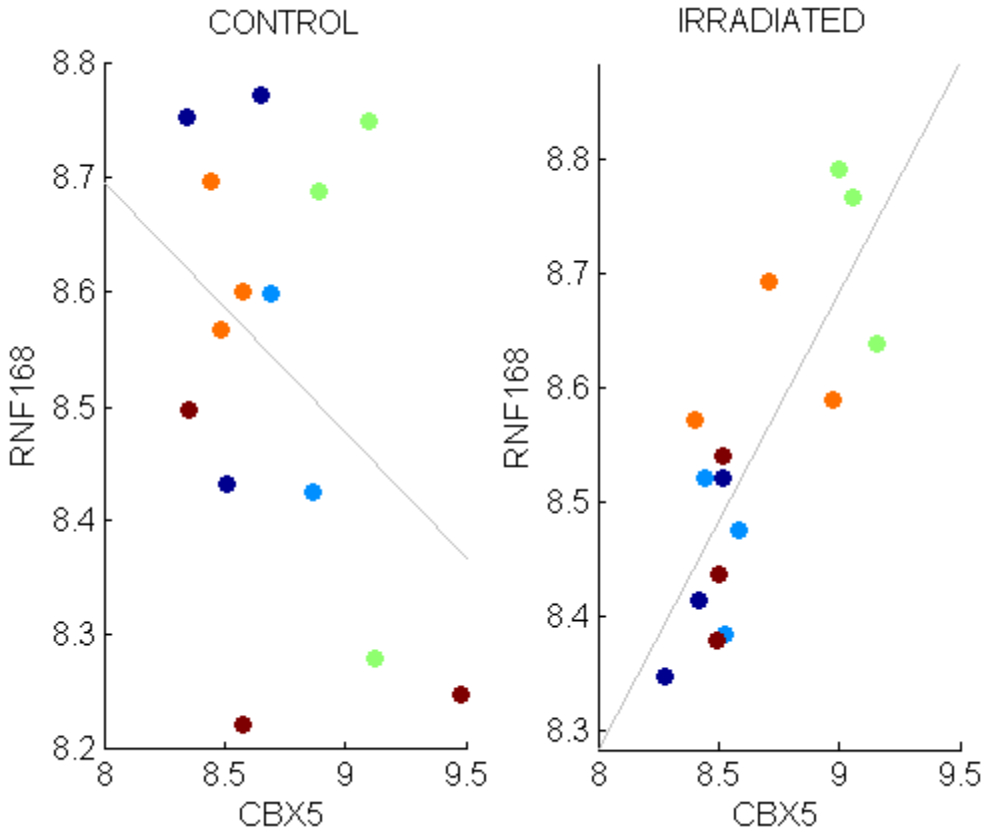



**Figure 2-4: Change in correlation of CBX5 and Rnf168 after irradiation** (Each colored dot represents mice of a particular strain).

Ccng1 is a target of tumor suppressor transcription factor p53 [252] and participates in p53 related processes in response to DNA damage such as apoptosis and check point regulation. It interacts with p53 ubiquitinylation protein Mdm2 and regulates accumulation of p53 at DNA damage site after gamma irradiation [253]. Gltscr2 (also known as Pict1) is also involved in regulation of Mdm2-p53 pathway and tumor growth [254].

Among other genes of this hub are Map1lc3 (Microtubule associated protein 1 light chain 3) that is involved in formation and activity of autophagic vesicles [255, 256] and Tlk2 (a homologue of tousled gene in Arabidopsis thaliana) that has been linked to cell cycle and DNA replication [257]. Accordingly genes of this hub are enriched in radiation response related GO processes such as 'DNA damage response', 'nucleic acid metabolism', 'DNA replication', 'cellular response to stress', 'cell cycle', 'signal transduction' and apoptosis. The differential expression of genes in these hubs highlights the impact of LDIR on DNA damage repair processes.

#### INTERCONNECTED RADIATION SENSITIVE HUBS

To understand the connections between radiation genes we extracted a network consisting of interconnected radiation sensitive hub genes (IRSH). Initiating IRSH with hubs Bcl2 and Rnf168 we iteratively merged other radiation sensitive hub genes that shared their peripheral genes with one of the current hubs of IRSH. The resultant IRSH (Figure 2-5) had 311 vertices and 330 edges.

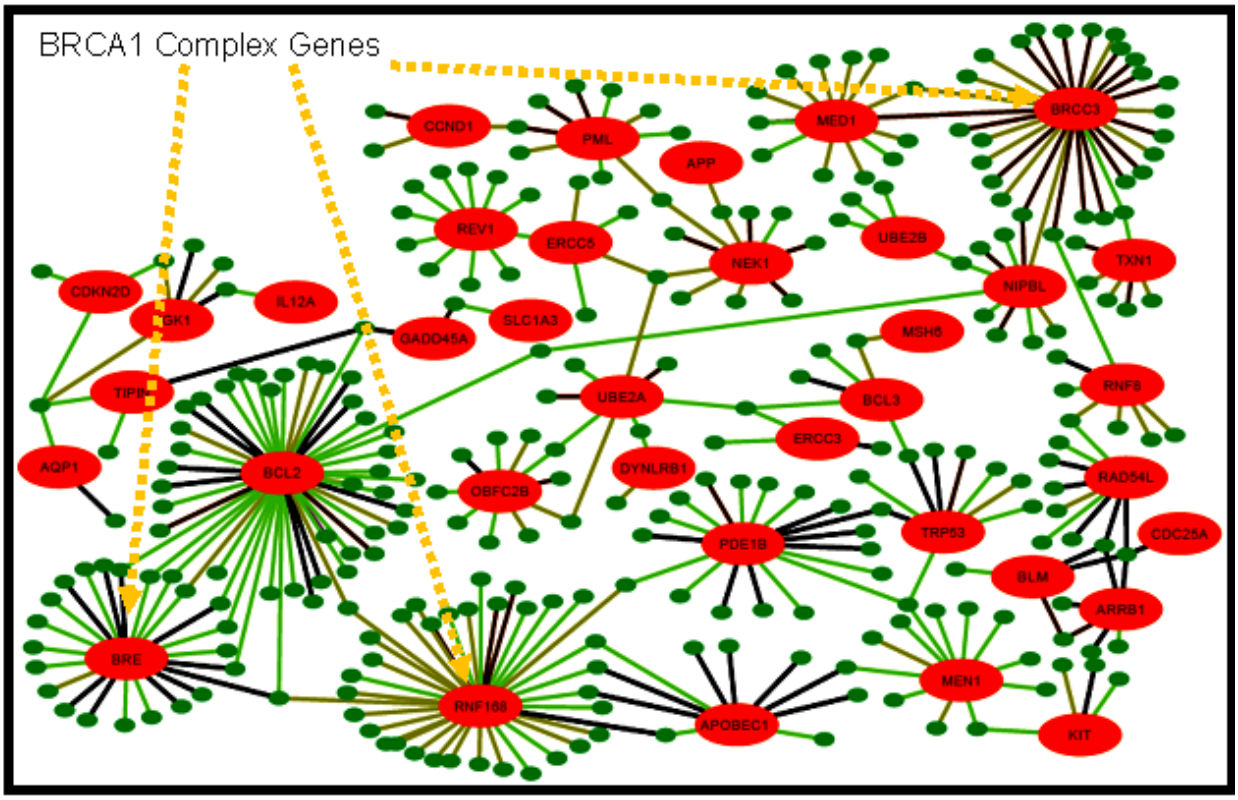



**Figure 2-5: Interconnected radiation sensitive hub genes.** Radiation sensitive genes are shown in red color. Green ellipses are other genes. Black edges connect gene pairs with very low APC(< 0.25) in one condition and significant APC in other.

IRSH was enriched with nearly 58% (37 out of 64) of GO annotated radiation genes present in DCG. Expectedly, the GO category “response to radiation” had a very low fPval of 2.69 E-36 since nearly 12% (37/311) of genes in IRSH are radiation sensitive. Other GO categories enriched in this network indicated presence of genes related to processes which are usually expected in a radiation response. These include response to stress (51 genes, fPval =1.94E-16), DNA damage(34 genes, fPval =4.011E-23), cell cycle (34 genes , fPval= 6.78E-14), regulation of apoptosis (33 genes, fPval =1.2E-10), chromosome organization (21 genes, 2.46E-8), histone modification (9 genes, fPval =3.89E-8), protein phosphorylation (20 genes, fPval =1.84E-5), oxidative stress (8 genes, fPval = 4.97E-4), protein ubiquitination (12 genes, fPval = 2.04E-5) , hemopoises (12 genes, fPval = 1.19E-4), immune system processes (20 genes, fPval = 4.6E-4) including B and T cell activation & regulation of interleukin production , intracellular signal transduction(23 genes, fPval = 2.36E-05), signal transduction in response to DNA damage (6 genes fPval = 5.18E-5).

Interestingly three of the radiation sensitive genes belonging to IRSH are related to well known breast cancer gene Brca1 that is also involved in initiation of DNA damage response following exposure to ionizing radiation [258]. Two of these genes, Brcc3 (with 29 neighbors) and Bre/Brcc45 (with 23 neighbors) are part of Brca1 complex [259]. The third gene, Rnf168 (with 36 neighbors) is required for recruitment of Brca1 complex at DNA damage site [245]. GO enrichment of these genes and their neighbors indicated that they are involved in radiation response related biological processes such as “stress response”, “response to ionizing radiation”, “apoptosis” ,”nucleic acid metabolism”, “cell

cycle", "response to DNA damage stimulus" ,"double strand break repair", "DNA integrity checkpoint" and " chromatin modification". For independent confirmation of relationship between genes from these three hubs we used GeneMania prediction server [260] to ascertain interactions among them. Out of 90 genes belonging to these hubs Genemania database had annotation for 56 genes connected together by 151 edges (58.6% coexpressed, 4.1 % interacted physically, 15.1 % colocalized and 8.5% predicted interaction). Human orthologs of 67 out of 90 genes of these 3 hubs were connected together with 344 edges (70% coexpressed, 19.2 % interacted physically, 5.2 % colocalized and 5.5% predicted interaction) i.e. on an average, each of these 67 genes had evidence for connection with over five other genes of this network. 27 out of 67 (40.1%) genes formed a connected graph with previous evidence of physical interaction.

Like Rnf168 (described above), Brcc3 is a radiation sensitive gene involved in DNA damage repair. Genes connected to Brcc3 hub are enriched in GO categories such as "nucleic acid metabolic process", "response to ionizing radiation", "cell cycle", "response to stress" and "chromosome organization". Brcc3 is known to interact with Bre, a novel stress response, anti-apoptotic gene that is up-regulated in hepatocellular and esophageal carcinomas [261, 262] and tumors. Both of them also potentiate the ubiquitin ligase activity of Brca complex and their deficiency increases the sensitivity to ionizing radiation [259]. Neighbors of Bre including Ssrp1, Nkiras2, Hrmt1l2/ Prmt1, Rhobtb2 and Lyl1 are also involved in processes relevant to radiation response such as cancer, DNA damage repair and apoptosis. Ssrp1(structure-specific recognition protein

1) may have a role in DNA damage prevention as cancer cells deficient in Ssrp1 become more sensitive to cisplatin, a DNA damaging anti-cancer chemotherapy drug [263, 264]. Similarly, Prmt1 (protein arginine N-methyltransferase 1) has a role in maintenance of integrity of DNA and its loss has been reported to be associated with increase in DNA damage, checkpoint defects and chromosomal aberrations [265]. It also cooperates in transcriptional activation by p53 [266]. Rhobtb2 (Rho-related BTB domain containing 2) is a pro apoptotic tumor suppressor gene of breast cancer [267, 268]. Nkiras2 (Nf-κB inhibitor interacting Ras-like protein 2) aka κB-Ras2 is known to suppress NF- κB1 (NF- κB). The correlation of Nkiras2 and NF- κB1 changed from 0.664 to -0.365 after LDIR exposure. As stated previously the correlation between NF- κB and its activator Fkbp5 changed in the opposite direction from -0.413 to 0.635. As expected the correlation of the activator and the repressor of NF- κB changed in opposite direction. The smaller magnitudes of the correlation coefficients can be explained by the fact that NF- κB is a very busy protein interacting with over 150 inducers and 150 targets [269].

One of the neighbors of Bre is an oncoprotein Lyl1 (lymphoblastic leukemia gene), that is ubiquitously expressed transcription factor [270] . The correlation of Lyl1 and Bre changes forms almost zero (0.04) in control to highly positive (0.91) (Figure 2-6). Bre and Lyl1 are both cancer genes hence the change in their correlation following exposure to radiation is worth noticing. Over-expression of this pro-leukemic gene is known to increase the number of T-cells and hematopoietic progenitors in bone marrow of mice [271] and also causes B and T-cell lymphomas [272].

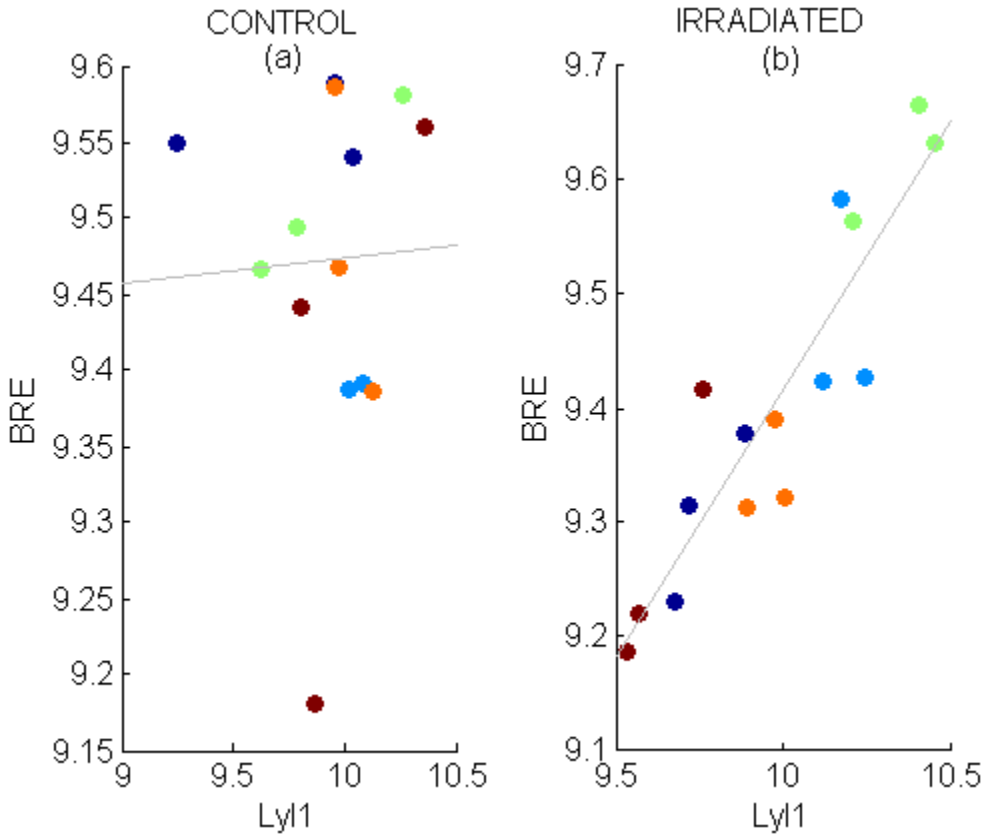



**Figure 2-6: Change in correlation of Lyl1 with BRE before (left) and after (right) irradiation** (Each colored dot represents mice of a particular strain).

Lyl1 and its 61 neighbors are enriched in genes belonging to GO category leukocyte differentiation (fPval= 0.0452). Since radiation is believed to have effect on immune system and population of leukocytes like T and B cells we further explored the neighbors of this gene by lowering the threshold of differentiation. From the genes in DCG we extracted those that had minimum absolute correlation of 0.6 with Lyl1 in either control or radiation and also had a difference of correlation of 0.6 with Lyl1 between radiation and control. The 737 neighbors of this relaxed hub of Lyl1 are highly enriched in immune system genes and genes involved in differentiation of lymphocytes as is evident from their GO enrichment. This hub is enriched with genes related to immune processes (42 genes, fPval= 1.21 E-6), regulation of apoptosis (53 genes, fPval=7.48 E-11), regulation of B cell activation ( 8 genes, fPval= 9.4 E-4), B cell homeostasis ( 5 genes, fPval= 1.1 E-4), B cell apoptosis ( 3 genes, fPval= 7 E-3), regulation of T cell differentiation ( 6 genes, fPval=9 E-3), T cell activation ( 8 genes, fPval= 2.67 E-2) and T cell homeostasis ( 3 genes, fPval= 4.45 E-2). Out of the 42 genes relating to immune system in this hub 17 are hematopoietic. Following exposure to radiation, 15 out of these 17 genes get negatively correlated with Lyl1 and two are positively correlated (Figure 2-7). The shift in correlation among so many hematopoietic genes following exposure to LDIR indicates that it effectuates changes in genes pertaining to this system in spleen.

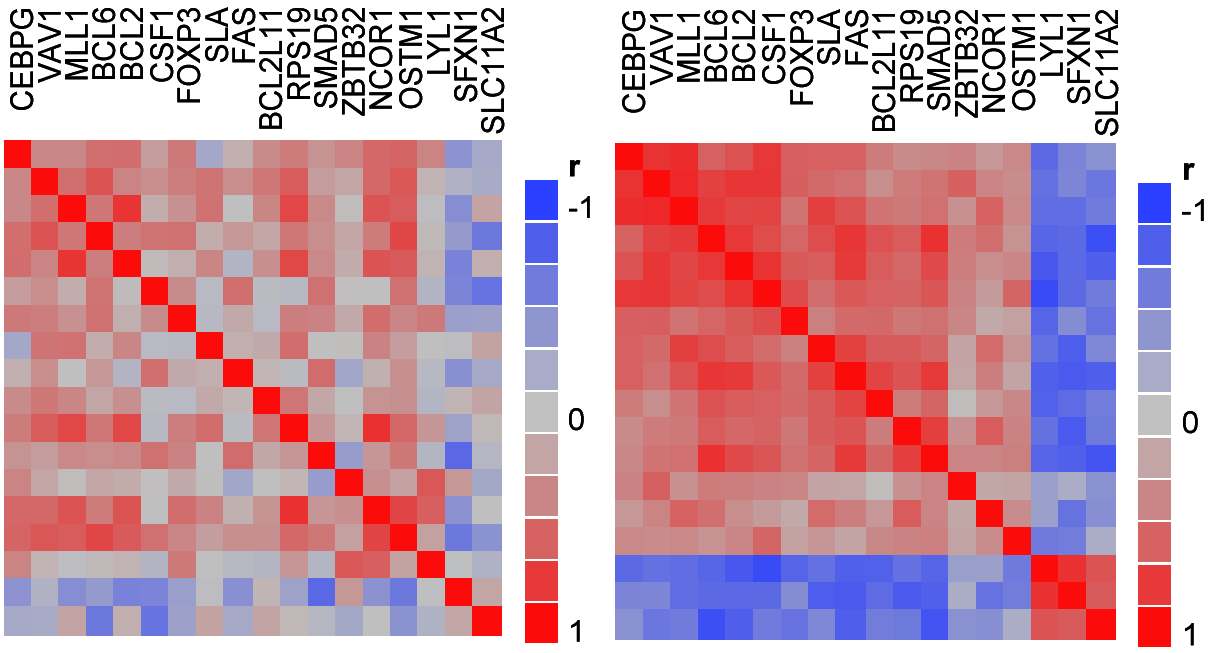



**Figure 2-7: Change in correlation of Lyl1 and other hematopoietic genes before (left) and after (right) irradiation.** Lyl1 gets differentially correlated with 17 genes after irradiation.

To further explore DCG we examined the types of its edges. Only a slight difference in percentage of edge types existed between DCG and IRSH. LDIR effectuated 60.84% correlation emergence and 38.86% correlation disruption events in IRSH as against 56.69% and 42.24% respectively in DCG. Since two of the hub genes (Rev1 AND Apobec1) of IRSH exhibited correlation emergence with all of their neighbors we investigated them further.

**Rev1 Hub**

Rev1, a Y family DNA polymerase, participates in DNA translesion synthesis by error prone incorporation of deoxycytidine at DNA damage sites in eukaryotes and prokaryotes [273, 274]. Rev1 also plays a role in introducing somatic hyper mutations in the variable region of immunoglobulins [275]. It is also implicated for its role in DNA damage repair. Introduction of mutations in Rev1 or its depletion caused defects in ionizing radiation induced mutagenesis, increase in chromosomal aberrations, residual DSBs and sites of homologous recombination repair [276, 277]. Inactivation of Rev1 in DT40 chicken cells led to increase in apoptosis and sensitivity to DNA damaging agents [278]. Caspase-2, a neighbor of Rev1, is a conserved gene involved in radiation induced apoptosis [279, 280]. The correlation between Rev1 and Caspase-2 changed from non significant (r=0.377) in control mice to significant negative (r= -0.801) after irradiation. Similarly the correlation between Rev1 and Ercc5 (excision repair cross-complementing rodent repair deficiency, complementation group 5) also changed from 0.15 to -0.833. Ercc5 is DNA excision repair gene required for induction of LDIR induced adaptive response to higher doses of gamma radiation [281]. Negative

correlation between Caspase-2 and Rev1 can be explained by the fact that Rev1 is anti-apoptotic and Caspase-2 is pro-apoptotic. In irradiated mice caspase2 got positively correlated (r = -0.63266314 for control and r = 0.63266314 for radiated) with another pro-apoptotic proteolytic gene Srgn (Serine Glycine). Srgn is known to play a role in granule mediated apoptosis by cytotoxic T lymphocytes [282]. Srgn remained negatively correlated with Rev1 under both the conditions (r=-0.78296878 for control and r = -0.68 for radiation). The changing correlations of these genes with Rev1 suggest readjustment of balance among pro-apoptotic and anti-apoptotic forces in response to LDIR.

**Apobec1 Hub**

Apobec1 (apolipoprotein B mRNA editing catalytic subunit 1), a radiosensitive gene, is a cytidine deaminase capable of introducing somatic mutations in mRNA and DNA [283, 284]. It provides protection against radiation as demonstrated by decrease in survival of gamma-irradiated Apobec1 (-/-) intestinal stem cells [285]. Over-expression of Apobec1 results in development of tumors in transgenic mice [286]. Apobec1 exhibited correlation emergence with genes implicated for their role in cancer, apoptosis and DNA damage repair like Rras-2 (related RAS viral oncogene homolog 2), Mcm7 (minichromosome maintenance deficient 7), and Cbx5 (chromobox homolog 5). Rras-2 (also known as TC21), a member of ras family of genes, has a high degree of sequence identity with N-terminal catalytic domain of Ras proteins [287, 288] . Mutations have been observed in this gene in cancer cells in human. Mutant forms of Rras-2 induce cellular transformation and promote cell survival. [287]. Mcm7 is needed for recruitment

of cell cycle check point and DNA damage repair protein ATR (ataxia telangiectasia and rad3 related) to DNA damage sites [289, 290]. Decrease in level of Mcm7 increases DNA damage [289].Cbx5 also known as HP1a (Heterochromatin protein 1-alpha) is recruited to DNA damage site after UV and IR exposure and its deficiency is associated with increase in DNA damage and genomic instability [249-251]. Apobec1 had nearly zero correlation with Mcm7 (r=-0.06) and Cbx5 (r=0.07) in control. On the other hand, in LDIR exposed mice Apobec1 (Figure 2-8) had a very significant negative correlation with Mcm7 (r=-0.94) and Cbx5 (r=-0.93). Since Apobec1 is potentially mutagenic, emergence of strong negative correlation of this gene with two of DNA damage repair genes after LDIR exposure is interesting.

#### PATHWAYS INFLUENCED BY RADIATION

We used annotations from all the pathways in Kegg database to determine the effect of LDIR on relationships of genes in biological pathways. Comparison of genes in the DCG with KEGG pathways revealed that 15 genes (11 gene pairs) relating to cancer pathways and 11 genes (9 gene pairs) pertaining to Mapk pathway (mitogen activated protein kinase pathway) were present in the DCG. Mapk pathway is closely related to cancer pathways. Nine out of 11 genes pertaining to Map kinase pathway formed a connected graph (Figure 2-9). Ten out of 15 genes of the cancer pathway were also connected to the nine genes of Map Kinase pathway forming a connected graph of 18 genes (Mapk1 is present in both the pathways).

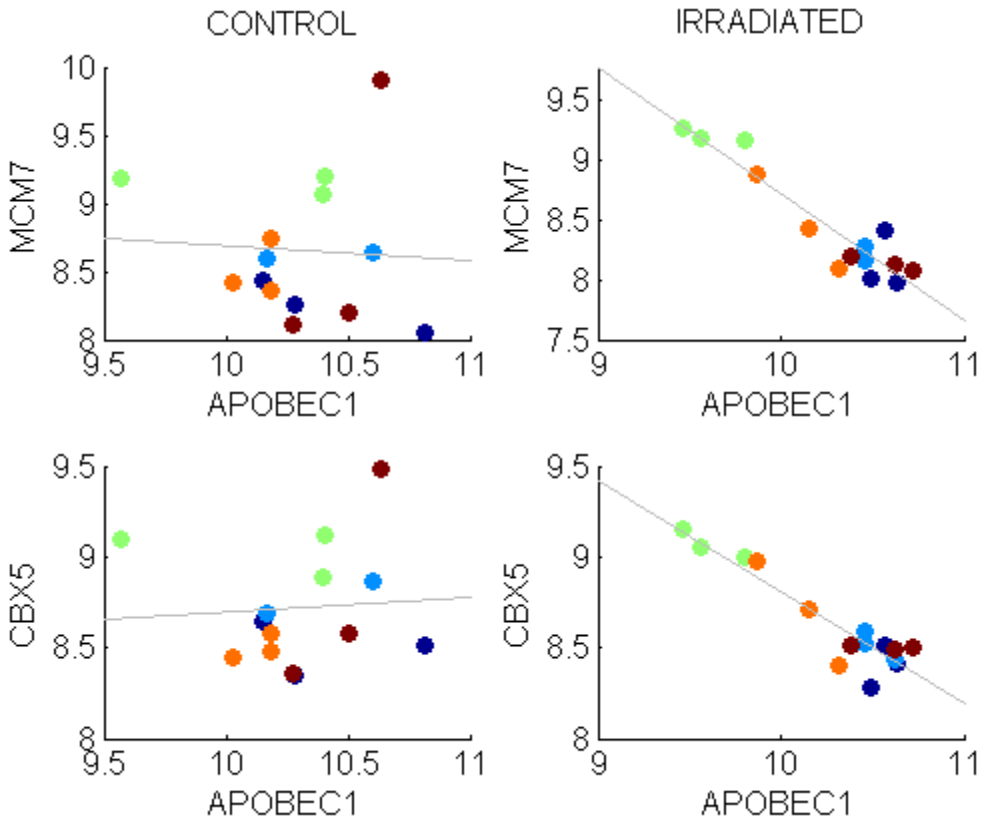



**Figure 2-8: Change in correlations of APOBEC1 with CBX5 and MCM7 after irradiation** (Each colored dot represents mice of a particular strain).

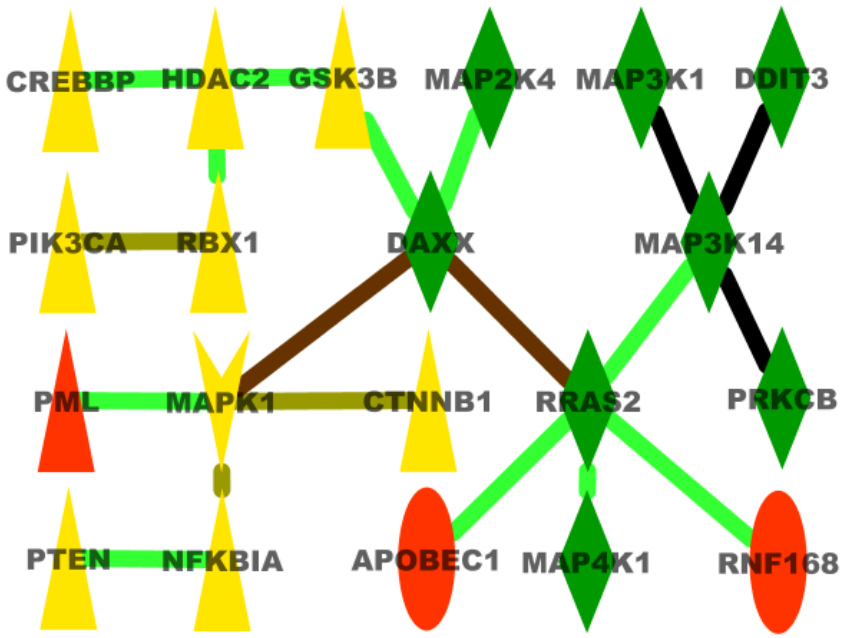



**Figure 2-9: Cancer (triangles) and MAP Kinase (diamonds) genes change their relationship after exposure to radiation.** Mapk1 shown as V belongs to both pathways. Radiation sensitive genes connected to this network are shown in red color.

We did random sampling of DCG to determine the probability of finding connection among genes by chance alone. We selected nine genes from the DCG ten million times. None of these tests returned a connected graph of 9 randomly selected genes (p-value <1.0E-7). The connection between genes belonging to these two pathways indicates that low dose radiation might be instrumental in readjustment of relations between genes belonging to Mapk signaling pathway and cancer pathways. Three genes (Apobec1, Rnf168 and Pml) connected to this network are annotated as "sensitive to radiation" in GO. Apobec1 and Rnf168 were attached to Rras2 gene of Mapk pathway. Pml is a member of cancer pathway. Three genes of cancer pathway, viz. Pml (promyelocytic leukemia), Daxx (Fas death domain-associated protein) and Mapk1 (Erk2), cooperate to promote apoptosis. The exact mechanism of promotion of apoptosis is not fully understood, but phosphorylation of Pml helps formation of apoptosis promoting Daxx-Pml complex. Mapk1 is known to phosphorylate Pml in response to arsenic tri-oxide treatment. Phosphorylated Pml forms a complex with Daxx which promotes apoptosis [291, 292]. Pml is also phosphorylated after exposure to gamma radiation and plays a key role in gamma radiation induced apoptosis [293]. The change in correlation of these three genes after exposure to LDIR was intriguing. The Daxx-Pml correlation was reduced from 0.482 to almost zero (-0.1). Daxx-Mapk1 is reduced from 0.87 to almost zero (-0.17) and Pml-Mapk1 correlation changes from positive (0.58) to negative (-0.68). The relatively low positive correlations of Pml with Daxx and Mapk1 under control can be explained by absence of a strong apoptosis inducing stimuli. Moreover, Pml is regulated at various transcriptional, post transcription

and translational levels depending on cellular context and external stimuli [294]. Hence, it would be difficult for it to become highly correlated to a single gene. The change of Mapk1 and Pml correlation from positive to negative after exposure to LDIR suggests a possible feedback inhibition which may also lead to disarray in correlations of Pml with the other two genes. The overall LDIR effect on the relationship between these three needs to be investigated further.

Two other genes of cancer pathway, HDAC2 (histone deacetylase 2) and GSK-3β (glycogen synthetase kinase 3 beta), changed their correlation from -0.7 to +0.7 after exposure to LDIR. Both of these genes are involved in radiation response. Cells depleted of HDAC2 are known to be sensitive to radiation and exhibit defective DNA damage response [295]. It has been shown that inhibition of GSK-3β reduces radiation induced apoptosis in endothelial cells. GSK-3β is inactivated by phosphorylation of its serine9 under the influence of AKT after exposure to IR [296, 297]. HDAC2 is also known to be involved in phosphorylation of GSK-3β via Inpp5f and AKT. In H9c2 rat ventricular myocytes decrease in HDAC2 led to increased Inpp5f and decrease in phosphorylation of GSK3β [298]. Close coordination of these two genes for regulation of apoptosis may explain development of positive correlation between HDAC2 and GSK3β (Figure 2-10) following LDIR.

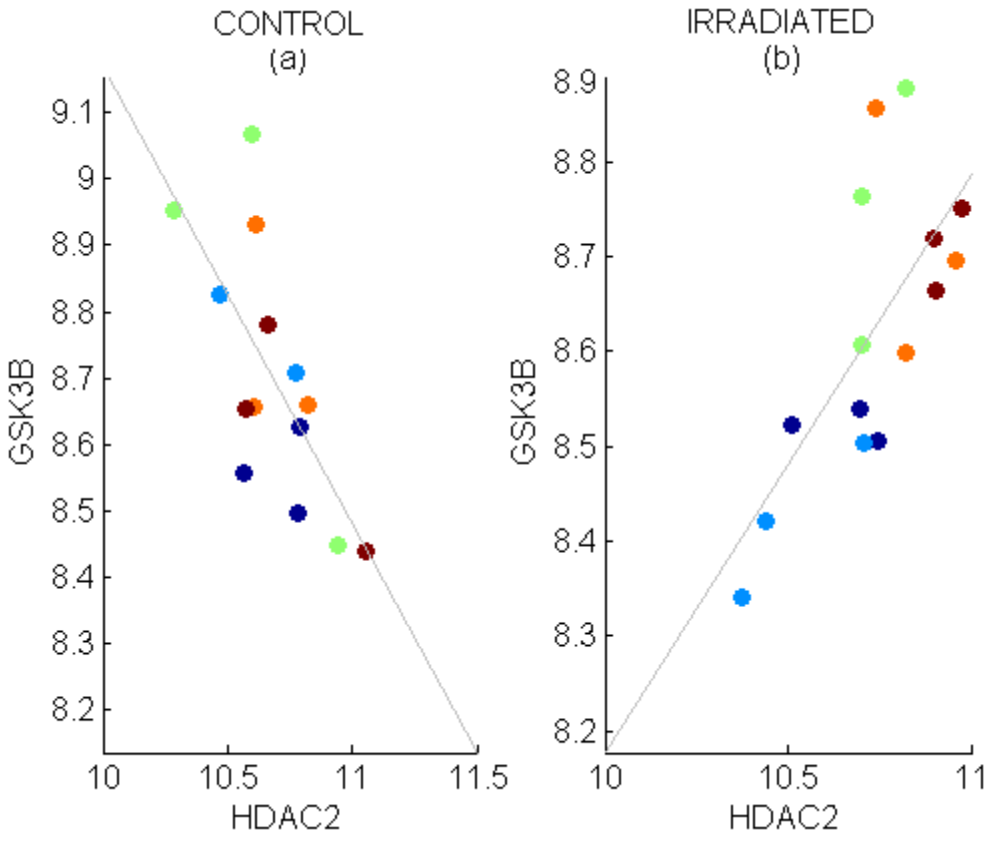



**Figure 2-10: Change in correlation of HDAC2 and GSK3B after irradiation**. (Each colored dot represents mice of a particular strain).

### MAXIMAL CLIQUES

A clique is a subset of vertices in a graph that are all interconnected by edges. A maximal clique is a clique of largest size which is not subset of any other clique in the graph. We extracted all the maximal cliques of size four and above. Due to sparse connections in DCG only 51 maximal cliques of size four were present. One of interesting maximal clique consisting of Rpa1, H2afz (H2A histone family, member Z), Fkbp2 (FK506 binding protein 2) and Npm3-Ps1 (nucleoplasmin 3, pseudogene 1) exhibited correlation emergence amongst all its genes (Figure 2-11). Correlation among its genes changed from non significant to significant (towards positive direction) under LDIR (Figure 2-12 and 2-13). Rpa1, one of the genes in this maximal clique, also belonged to Bcl2's hub. Rpa1 is known to be active during DNA replication, repair and recombination [235], DNA damage double strand repair and prevention of apoptosis following exposure to IR [236]. Heterozygous mutations of Rpa1 resulted in development of lymphoid tumors and defective repair of DNA double strand breaks in mice. Homozygous mutations of Rpa1 caused embryonic lethality in mice [238]. Fkbp2 is an immunophilin that interacts with C1q protein of C1 complement suggesting its role in the complement system [299]. It has been demonstrated that IR elevates the classical complement system in blood of rat [300]. H2afz is a variant of H2a, a nucleosomal protein involved in transcriptional control. It has been established that double strand break induced phosphorylation of H2av, a member of H2az family, prevents radiation induced apoptosis in imaginal disc cells in Drosophila larvae [301]. Npm3-ps1 is designated as a putative pseudogene of nucleoplasmin 3 [302].

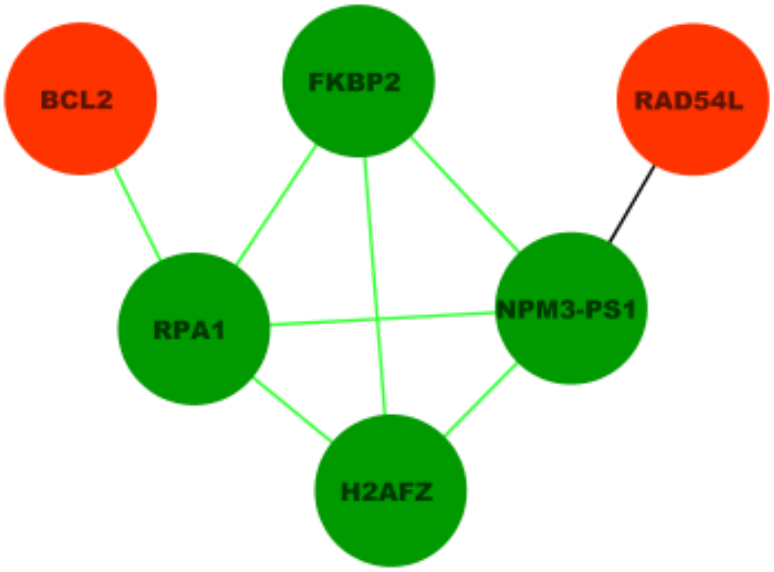



**Figure 2-11: A maximal clique (in green) with 4 genes exhibiting emergence of correlation among its members.** Red circles represent radiation sensitive genes.

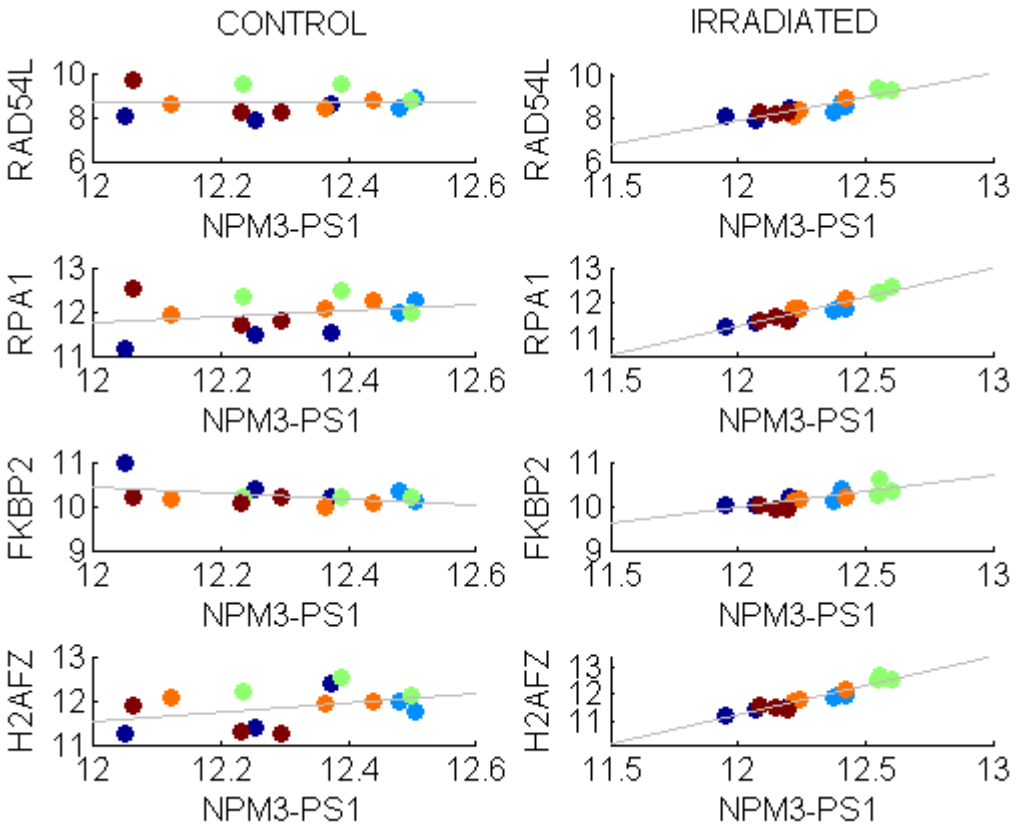



**Figure 2-12: Emergence of correlation between NPM3-PS1 and four other genes.**

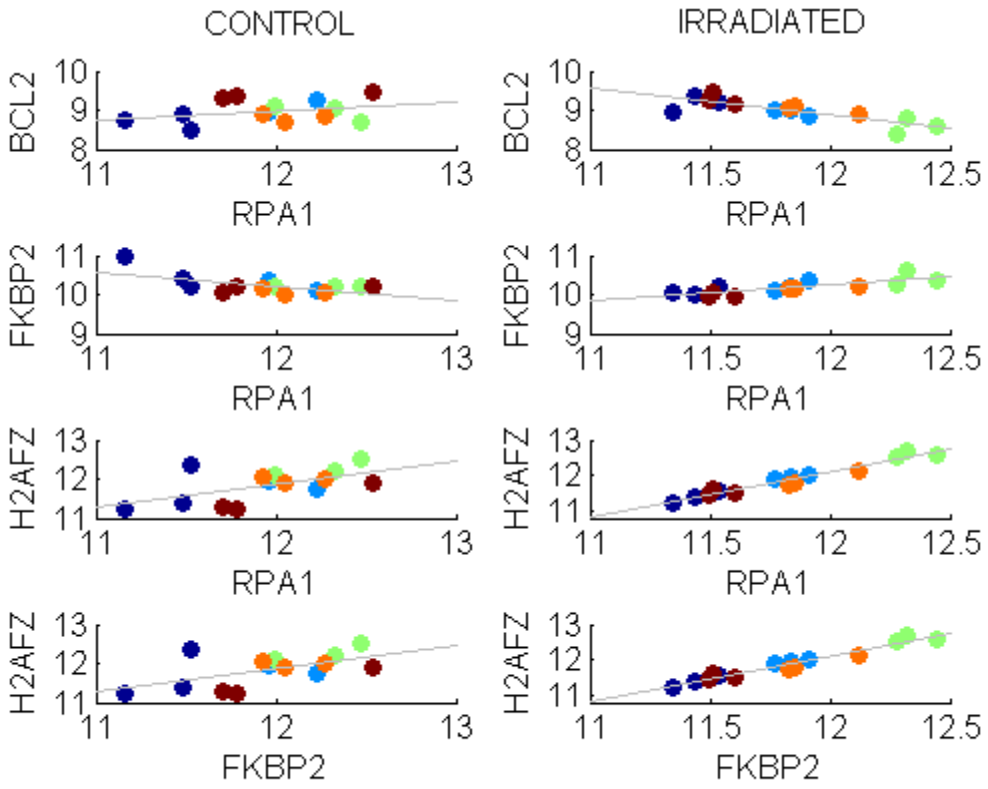



**Figure 2-13: Change in correlation of RPA1 and FKBP2 with their neighboring genes.**

Nucleoplasmins are known to improve survival of cells exposed to ionizing radiation [303]. Besides exhibiting correlation emergence with other members of its maximal clique, Npm3-ps1 exhibited emergence of very high correlation with Rad54l (Rad54-like) gene after exposure to LDIR (r= 0.01505143 for control and r= 0.90344726 after exposure to LDIR). Rad54 is an evolutionarily conserved member of Snf2/Sw12 family involved in various stages of homologous recombination [304, 305] and DNA repair including double strand break repair [304, 306]. Cells with disruptions in Rad54 gene had lower survival rate than those with functional Rad54 [305]. Correlations of Npm3-Ps1 with other radiation response genes may therefore suggest its role in sensitivity towards LDIR. The shift towards positive correlation of all the four genes suggests a concerted protective response of these four genes to LDIR.

In view of the LDIR affiliated response of the genes in this maximal clique we investigated its neighboring genes. The resultant gene network consisting of 299 genes and 487 edges was highly enriched in GO categories such as response to stress (44 genes , fPval=6.3 E-12), cell cycle(42 genes, fPval =1.54E -20), apoptosis (20 genes, fPval=1.65 E-6), DNA metabolism (29 genes, fPval= 2.1E-15), response to DNA damage (22 genes, 4.36E-11), response to oxidative stress (9 genes, fPval=1.12E-4), protein transport, response to radiation (6 genes, fPval=3.6E-2) and cellular homeostasis. This further suggests that the genes belonging to this clique are relevant to LDIR response. Six (Sgk1, Cdkn2d, Bcl2, Tipin, Aqp1 and Rad54l) out of 64 GO annotated radiation response genes exhibited emergence of correlation with members of this maximal clique after exposure to LDIR. Out of these six genes 4 of them

changed their relationship with Fkbp2. Fkbp2 is also the largest hub gene of the DCG with 267 neighbors that makes it a suitable candidate for investigation of its role in LDIR. The other two radiation response genes Bcl2 and Rad54L change their relationship with Rpa1 and Npm3-Ps1 respectively. The LDIR induced changes in interrelationships of genes in this network need to be investigated further.

To conclude, we have demonstrated that the two stage statistical filtration method used here detected gene networks that are differentially coexpressed (correlated) after exposure to a stimulus. We used graph theory to extract relatively dense portions of the differentially correlated graph. Differential correlation revealed gene networks highly enriched with genes implicated in radiation response, DNA damage repair, apoptosis and cancer with hub membership enriched in members of the BRCA complex. Though we used microarray data this method can easily be adapted to gene expression data obtained from any high throughput method including next generation sequencing (NGS) and exon arrays. Our two stage filtration method generates a network consisting only of statistically significant differential correlations among genes. The generation of this graph does not require a predefined threshold for determining the inter-relationship of genes. The threshold is determined statistically in a data dependent manner. A relatively sparse network created with rigorous statistical filter makes it attractive and suitable for NGS and exon array methods that generate huge data and enable exploration of the correlation between not only genes but also the variants of genes. After extraction of differentially correlated portions of the gene network we can always

explore the rest of gene network to comprehend the pathways and signaling networks impacted by them.

## METHODS

### EXPRESSION PROFILING

Five strains of mice (129S1/SvImJ, NOD/LtJ, CBA/J, CAST/EiJ & WSB/EiJ) were used for gene expression profiling using illumina microarrays. 3 mice of each strain irradiated with 10cGy of γ-radiation using a $^{137}$Cs source. Control mice (3 mice for each strain except 129S1/SvImJ.2 mice for 129S1/SvImJ.) were sham irradiated. Mice were sacrificed 24 hours following exposure and spleen cells of all the mice were stabilized using in RNAlater (Sigma-Aldrich, St. Louis, MO) until RNA was extracted. Illumina (San Diego, CA) Mouse WG-6 v1.1 BeadChips were used for expression profiling. Expression profiling was performed by Genome Quebec (Montreal, Canada) as previously described [187].

### GRAPH THEORY AND STATISTICAL TESTING

Expression data from illumina microarray were preprocessed by Variance Stabilizing Transformation (VST) followed by Robust Spline Normalization (RSN) using lumi package [151] of Bioconductor [307]. Raw and Normalized data will be uploaded to GEO [308] database of NCBI. Only expressed probes with detection p-value less than 0.05 were considered for analysis. Matlab scripts were used to calculate pair wise gene to gene Pearson correlations and their significance. Statistical differences in correlations were calculated after Fisher's Z transformation of correlations using matlab. Multiple test

correction was done by q-value [202] package of Bioconductor. Each gene in the DCG was represented by a unique vertex by retaining maximally connected probes. Edges were placed between two genes if they had statistically significant difference in correlation for sham irradiated (control) and LDIR treated mice. 9 statistically different edges with low magnitude in difference (<0.3) because of their near perfect correlation in both conditions were removed. Perl scripts were used for extraction of hubs and random selection tests for connected components and hubs from DCG. Maximal cliques were extracted using Grappa, a tool developed by our lab. Gene enrichment analysis was done using Cytoscape [180] package Bingo [204] and perl scripts using Gene Ontology [176] annotation for biological processes. Benjamini-Hochberg [201] false discovery rate-corrected p-values were used for enrichment analysis. All source codes are available with the authors.

## CONCLUSIONS

We have shown that the differential co-expression method is able to detect subtle changes in gene expression that could not be detected by differential expression method of ANOVA. Differential co-expression method extracted putative gene networks perturbed in response to LDIR. The extracted networks were enriched with genes implicated in radiation response, DNA damage repair, apoptosis and cancer with hub membership enriched in members of the BRCA complex.

# APPENDIX

**Table A2-1: Enrichment of gene ontology IDS (GOIDS) associated with biological processes in largest connected component of differentially correlated graph.**

| GO-ID | Benjamini-Hochberg FDR corrected pvalue | Number of genes | Description |
|---|---|---|---|
| 42981 | 1.24E-63 | 345 | Regulation of apoptosis |
| 6950 | 1.45E-60 | 457 | Response to stress |
| 2376 | 8.98E-59 | 321 | Immune system process |
| 22402 | 2.53E-43 | 209 | Cell cycle process |
| 6259 | 1.38E-42 | 188 | DNA metabolic process |
| 6974 | 4.66E-34 | 152 | Response to DNA damage stimulus |
| 6281 | 2.35E-21 | 104 | DNA repair |
| 6955 | 8.59E-19 | 119 | Immune response |
| 2764 | 1.56E-16 | 63 | Immune response-regulating signaling pathway |
| 6260 | 1.65E-16 | 64 | DNA replication |
| 2253 | 2.93E-13 | 65 | Activation of immune response |
| 6979 | 6.86E-12 | 59 | Response to oxidative stress |
| 75 | 1.82E-10 | 40 | Cell cycle checkpoint |
| 7050 | 1.93E-09 | 34 | Cell cycle arrest |
| 9314 | 2.37E-09 | 64 | Response to radiation |
| 2366 | 2.85E-07 | 26 | Leukocyte activation involved in immune response |
| 2285 | 1.90E-06 | 18 | Lymphocyte activation involved in immune response |
| 77 | 4.78E-06 | 21 | DNA damage checkpoint |
| 42770 | 5.00E-06 | 22 | Signal transduction in response to DNA damage |
| 31570 | 2.04E-05 | 21 | DNA integrity checkpoint |
| 718 | 1.10E-04 | 11 | Nucleotide-excision repair, DNA damage removal |
| 2429 | 1.25E-04 | 21 | Immune response-activating cell surface receptor signaling pathway |
| 2703 | 1.48E-04 | 31 | Regulation of leukocyte mediated immunity |
| 2313 | 3.13E-04 | 5 | Mature B cell differentiation involved in immune response |
| 6873 | 4.02E-04 | 88 | Cellular ion homeostasis |

**Table A2-1 (Continued from previous page)**

| GO-ID | Benjamini-Hochberg FDR corrected pvalue | Number of genes | Description |
|---|---|---|---|
| 71158 | 9.73E-04 | 8 | Positive regulation of cell cycle arrest |
| 30330 | 1.24E-03 | 13 | DNA damage response, signal transduction by p53 class mediator |
| 42771 | 4.93E-03 | 9 | DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis |
| 42772 | 7.34E-03 | 6 | DNA damage response, signal transduction resulting in transcription |
| 6297 | 9.69E-03 | 8 | Nucleotide-excision repair, DNA gap filling |
| 6977 | 1.64E-02 | 4 | DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest |

**Table A2-2: 64 Radiation sensitive genes in largest connected component of differentially correlated graph and their neighbors.**

| Radiation sensitive gene | Degree | Neighbors |
|---|---|---|
| HMGN1 | 13 | CCDC50, CCL19, CCM2, CXX1C, D11WSU47E, D14ERTD668E, GARNL3, GJA4, LOC100043918, LTB, SAR1B, TBCA, USP37 |
| APOBEC1 | 10 | 9430065L19RIK, ABI1, CBX5, IL16, MCM7, MRPS34, PDAP1, RRAS2, STARD3, UPF1 |
| XRCC6 | 2 | ACOT7, ADK |
| INTS3 | 1 | DSCR3 |
| AQP1 | 2 | 9330180L10RIK, FKBP2 |
| MEN1 | 12 | 2310014G06RIK, ACAT3, ATOX1, CAB39L, CCDC65, CDKN1B, D10WSU52E, GAB1, IER3, LOC666621, QTRT1, UPF1 |
| APP | 1 | NOLA2 |
| EAR2 | 2 | DHX58, POR |
| SLC1A3 | 1 | 4833426J09RIK |
| CDKN2D | 3 | FKBP2, GTF3C5, MACROD1 |
| NR2F6 | 1 | LOC223653 |
| EGR1 | 2 | IQCB1, NDUFC1 |
| UBE2A | 8 | 4833438C02RIK, AP2B1, C030048B08RIK, GOLM1, LOC100048613, MICAL1, NR1H2, SUHW4 |
| SGK1 | 5 | FKBP2, KHK, MACROD1, P2RX1, TRIM41 |
| REV1 | 11 | CASP2, CIAPIN1, ERCC5, HNRNPH3, LLGL2, LMAN2L, LOC100042405, LOC100047827, MKNK1, MLL1, YKT6 |
| BRCC3 | 29 | 1110055N21RIK, 2310079P12RIK, 2900010M23RIK, ANKMY2, BC085271, BRD7, BZW1, C130045I22RIK, DHX15, DMTF1, DNAJC5, EG382843, EIF3K, ETFA, EXOSC6, LOC277856, LOC380707, LOC432554, MRPL52, MYD88, NIPBL, PRKCBP1, PSMG2, PTPRE, ROBLD3, SBDS, SETD3, SLC12A6, VAMP8 |
| USP1 | 4 | LOC100041864, LOC381448, TUBA1A, TUFM |
| MECP2 | 2 | ARNTL, ID2 |
| UBE2B | 4 | PJA1, SACM1L, SH3RF1, TAGAP |
| XPA | 4 | FARSB, PARN, RPL26, TIMM44 |
| USP28 | 5 | 1110020P15RIK, CTSZ, FBXL5, GANAB, GGTA1 |
| CCND1 | 3 | CD22, NFKB1, PTPLA |
| ARRB1 | 6 | 5033414D02RIK, AP1S1, ARPC1B, ATP5O, FRRS1, MTF2 |
| IL12A | 1 | KHK |
| DDB2 | 11 | 2610015J01RIK, ATP5L, B230386D16RIK, ENDOGL1, LOC100048508, LOC677551, NDUFS4, PPP1R9B, RNF213, RPL23, TMEM2 |

**Table A2-2 (continued from previous page)**

| Radiation sensitive gene | Degree | Neighbors |
|---|---|---|
| MAPK9 | 2 | SLC35D2, STK24 |
| DYNLRB1 | 2 | C030048B08RIK, UPP1 |
| GADD45A | 2 | 4833426J09RIK, CTPS |
| MED1 | 12 | 0610007P14RIK, A530082C11RIK, CGEF2-PENDING, CHMP2A, GMIP, LOC100041725, LOC277856, LOC432554, LOC671641, NDUFB10, PANK4, PTP4A2 |
| BLM | 4 | AP1S1, ASNS, ATP5O, MTF2 |
| NEK1 | 10 | 1810008A18RIK, BC057552, FNBP1, HDGFRP2, MICAL1, MUM1, NDUFA1, NOLA2, PHKG2, PPA1 |
| TIPIN | 3 | CTPS, FKBP2, NDUFS7 |
| PML | 9 | DGKA, IFT140, LUC7L, MAPK1, PHKG2, PTPLA, RALY, SNRP70, ZFP654 |
| TXN1 | 6 | BC057627, BC085271, GMPS, RNUXA, TMCC1, TPP2 |
| CHEK2 | 1 | ZFP623 |
| KIT | 4 | ACAT3, BCAP29, CAPNS1, SCL0004175.1_57 |
| 5430437P03RIK | 2 | LOC630242, MKKS |
| ERCC5 | 5 | BXDC5, GOLPH3L, MICAL1, NDUFA2, REV1 |
| PDE6D | 3 | ANAPC5, POLR2D, TMEM9B |
| NIPBL | 10 | 1110055N21RIK, AAAS, BRCC3, CUL1, EG433865, EPS15L1, IFI47, LRMP, PJA1, SQLE |
| PDE1B | 18 | 1200014J11RIK, 1810037I17RIK, 2310044H10RIK, 2600005C20RIK, B3GALT4, C230091E03RIK, D10WSU52E, DNAJB6, EG625917, EVL, GORASP2, GOSR2, GRCC10, HDAC5, LOC635086, RRP1B, TMEM147, ZBTB7A |
| BCL2 | 47 | 2700029M09RIK, 4930422G04RIK, ALDH2, BZW2, CAD, CAND1, CCDC53, COX7A2L, CTPS, EG432721, EG433865, EG633692, FKBP5, GPR89, HRMT1L2, IL10RB, IL33, INTS10, INTS7, LOC100046793, LOC100047749, LOC381649, MBNL2, MFNG, MTERF, N4WBP5-PENDING, NKIRAS2, NOL5, NUP133, ORC5L, PLOD3, PRICKLE3, PRPS2, RAB32, RPA1, SAPS1, SCHIP1, SEPP1, SERBP1, SRGN, SSRP1, STAU2, TMEM176A, TSC22D1, TUBB2C, WIPI2, ZFAND3 |
| BCL3 | 5 | BAT2, BC038822, SF1, SUHW4, TMEM177 |
| NFATC4 | 2 | BCL9L, NCOA6 |

**Table A2-2 (continued from previous page):**

| Radiation sensitive gene | Degree | Neighbors |
|---|---|---|
| RNF168 | 36 | ABI1, AK2, ARL8A, ATP6V1C1, B930006L02RIK, CBX5, CCDC53, CCNG1, CCNT1, CNOT1, EEF1B2, EG384525, EG432721, EG625917, GLTSCR2, H13, IGK-V38, ITM2B, KIF1B, LIMA1, LOC100039786, LOC100046343, LOC545396, LOC545487, MAP1LC3B, NDUFS3, OS9, PAPOLA, PDAP1, RHOT2, RRAS2, TCEB3, TLK2, VKORC1, YWHAB, ZCCHC6 |
| ERCC3 | 3 | 0610012G03RIK, PPIB, SUHW4 |
| ERCC4 | 1 | AHCYL1 |
| PIK3R1 | 2 | 0610038F07RIK, SLFN2 |
| TRP53 | 9 | 0610007P22RIK, 4933404K08RIK, BAT2, CPSF1, CRELD1, D10WSU52E, GRCC10, UCHL3, YIPF3 |
| MSH6 | 1 | TMEM177 |
| UBE4B | 1 | ARL2BP |
| BRCA2 | 1 | EIF2AK2 |
| RAD54L | 9 | 2210016F16RIK, 6720463L11RIK, AP1S1, ATP5O, LOC383897, MAP3K8, NPM3-PS1, SEC61B, ZC3H7A |
| CDC25A | 1 | AP1S1 |
| SOD2 | 3 | DOLPP1, SENP1, SSBP1 |
| OBFC2B | 10 | AP2B1, BRD2, BSCL2, FAM113B, HGS, LOC100048613, MED25, NDUFC2, RNF145, ZRANB1 |
| RNF8 | 7 | CNO, ETFA, GRIPAP1, MRPS28, RAB1, RNPEP, SEC61B |
| UACA | 1 | 2610036D13RIK |
| SFRP1 | 2 | NDUFA8, NIT2 |
| MAPK14 | 1 | 2310011J03RIK |
| BAX | 2 | C030046I01RIK, UFC1 |
| USP47 | 1 | PUF60 |
| BRE | 23 | 2310036O22RIK, B930004C15RIK, CCDC53, CHMP4B, D10ERTD641E, DGAT1, HRMT1L2, HS3ST3B1, LOC676724, LYL1, MBNL2, MRTO4, MTIF3, NKIRAS2, PRPS2, PSMC5, RHOBTB2, RNASEN, RPS2, SSRP1, STAU2, UBE2Q1, ZFAND3 |
| CIRBP | 1 | LOC381774 |

# CHAPTER 3 : CORRELATION THRESHOLD FOR EXTRACTING GENE NETWORKS FROM BASELINE GENE EXPRESSION PROFILES FROM MICROARRAY DATA

(This Manuscript will be submitted for publication with following authors: Sudhir Naswa, Dr Arnold Saxton, Dr Brynn H Voy, Charles Phillips, Dr. Michael A. Langston)

## ABSTRACT

Pearson correlation is often used by clustering and graph theoretical algorithms to extract putative genetic networks from gene to gene correlation matrices derived from high throughput gene expression data. Higher correlation coefficients (r=0.875) are frequently employed by these algorithms to establish relationship among genes. Higher number of samples in recently popular system genetics and genetical genomics methods enable us to test the feasibility of employing lower correlation coefficients for extraction of gene networks from expression data. Following this hypothesis we gradually reduced the absolute Pearson correlation (APC) threshold from conventionally used high value to a low but statistically significant (pvalue <0.01) value to investigate gene enrichment in microarray data from liver and spleen tissues of BXD mice. Graph algorithms were used to extract paracliques from the thresholded graph.

The graphs generated at significant APCs of 0.413 for liver data and 0.41 for spleen data had higher number of paracliques with bigger size as compared to graphs generated at higher APCs of 0.6, 0.75 and 0.875. Paracliques extracted at significant APCs of 0.41 (liver) and 0.413 (spleen) were more enriched with biological processes as compared to paracliques at higher thresholds of 0.6, 0.75 or 0.875.

## BACKGROUND

High throughput gene expression data generated through large scale 'omics' technologies such as microarray, next generation sequencing techniques and SAGE (serial analysis of gene expression) have enabled us to extract information about biological pathways and gene regulatory networks using powerful statistical and computational algorithms. High quality putative gene regulatory networks can be extracted from expression profiles because correlated expression patterns are believed to be co-regulated and involved in common biological pathways ( guilt by association). The gene networks extracted from transcription profiles by employing measures of similarity such as correlation and mutual information amongst all the gene pairs are called relevance networks [167]. The relevance networks allow multiple relations among genes, can handle positive as well as negative correlations between genes and easily combine information from data of diverse types [167, 170, 171]. Extraction of biologically meaningful information from a large network representing pair-wise association between thousands of genes requires efficient computational algorithms. Graph algorithms provide a means to extract dense and highly connected regions from these networks[168]. We have developed [182-186] and applied [169, 179, 181, 187, 216] graph algorithms for extraction of putative gene networks from gene expression data. Gene networks can be represented as a graph where each gene is placed on a single vertex and the correlation between these genes are depicted as edges. Edges above a suitable threshold are retained in the graph. After removing the edges below

threshold, the dense regions of the graph such as cliques and hubs are extracted to model putative gene networks. A clique is well known graph theory conception where all vertices are connected to one another by edges. In biological context a clique models a group of tightly co-regulated and hence correlated genes that may be participating in a common biological pathway or may be influenced by a common stimulus in a case control study. To further imitate the natural biological networks we relax the clique by allowing it to miss a few edges. We call these relaxed cliques paracliques [169]. This relaxation tries to compensate for the inherent noise in the microarray data and stochastic nature of biological processes.

Various methods have been used to derive a correlation threshold for filtering the gene networks. These include use of an arbitrary high absolute correlation [309], retention of top 1% correlates of each gene [310], selection by spectral graph theory [185], doubling of number of paracliques [183], values exceeding the correlations of buffer spots with genes [169] and permutation testing [170, 171]. The thresholds obtained by these methods have been applied to derive co-expression networks from gene expression data [169, 183, 185] as well as from a combination of expression profiles with drug response data [170, 171]. These methods usually suggest use of a high absolute Pearson correlation (APC) threshold varying between 0.7 and 0.9. An approach employing a high APC coefficient filter [169] followed by clique centric and other clustering methods has served well for many case control studies where a perturbed biological system causes stimulus-sensitive genes to be either up or down regulated simultaneously resulting in very high correlation among them. The choice of higher APC

coefficient in such studies also safeguards against false positives, because case control studies usually contain fewer microarrays so that lower correlations may be neither statistically significant nor biologically relevant to the stimulus. Due to recent popularity of system genetics and genetical genomics [311-313] the experiments profiling gene expression of panel of unperturbed organisms are becoming popular. Recombinant Inbred (RI) lines of mice are used commonly for such studies. RI lines are derived by mating inbred lines of mice. BXD mice belong to RI lines derived by brother sister mating of F2 generation of C57BL/6J and DBA/2J strains of mice. The fact that they are inbred makes them a suitable model for studies across time and space. Since the mice in a panel of RI lines are inbred progenies of genetically different ancestral strains, they exhibit genetic diversity as a result of random repetitive recombination of genomes of these strains. This genetic diversity makes RI lines model genetic reference populations for unearthing metabolic pathways and gene interaction networks that influence complex traits relating the central nervous system[313], the hematopoietic system[312] , the immune system [187] and so forth. On the other hand the diversity may also reduce the correlation of genes among the organisms of a panel of RI strains. Lack of a stimulus deriving these expression profiles may also be a reason for relatively lower correlations among genes involved in common genetic networks and metabolic pathways. Availability of expression profiles from a higher number of organisms associated with genetical genomics studies provides us with better statistical power to extract genetic networks and pathways even at a lower APC. Here we compared the enrichment of gene networks obtained at high thresholds with networks obtained at a

low but statistically significant APC (p-value < 0.01). The decrease in APC threshold resulted in increase in enrichment of biological processes in gene networks in liver and spleen tissues of BXD RI mice.

## METHODS

Gene expression profiles of spleen (EPS) and liver (EPL) tissue of BXD mice were used for this study. A detailed description of the experimental procedures relating to acquisition of EPS from 38 BXD mice and its preprocessing is available in [187]. Briefly, total RNA from spleens of BXD mice was isolated and expression profiling was done on Mouse WG-6 v1.1 Beadchips from Illumina Inc. Expression data were preprocessed using Variance Stabilizing Transformation (VST) followed by Robust Spline Normalization (RSN) using the lumi package [151] in R/Bioconductor [307]. Raw and normalized microarray data are available in NCBI's GEO database (http://www.ncbi.nlm.nih.gov/projects/geo; Accession GSE19935). Lowess normalized EPL pertaining to 39 BXD male mice was downloaded from GEO (GSE17522) [314]. The data for mouse gene to gene interaction pairs were downloaded from Biogrid [315], Amadeus compendium [316], Integrated Transcription Factor Platform [317] and LymphTF database [318]. Biological pathways data and gene regulatory network data were obtained from KEGG [319], MGI MouseCyc [320, 321], T cell Gene Regulatory networks[322] and Transcriptional Regulatory Element Database[323]. Data from various sources were integrated using mySQL database. APC coefficients were calculated for the gene pairs obtained from these databases. Mean gene correlation

was calculated for each of the pathway and gene regulatory networks. Matlab scripts were used for enumerating the vertex degrees of hub genes, their percentile distributions and for generating edge lists for graphs.

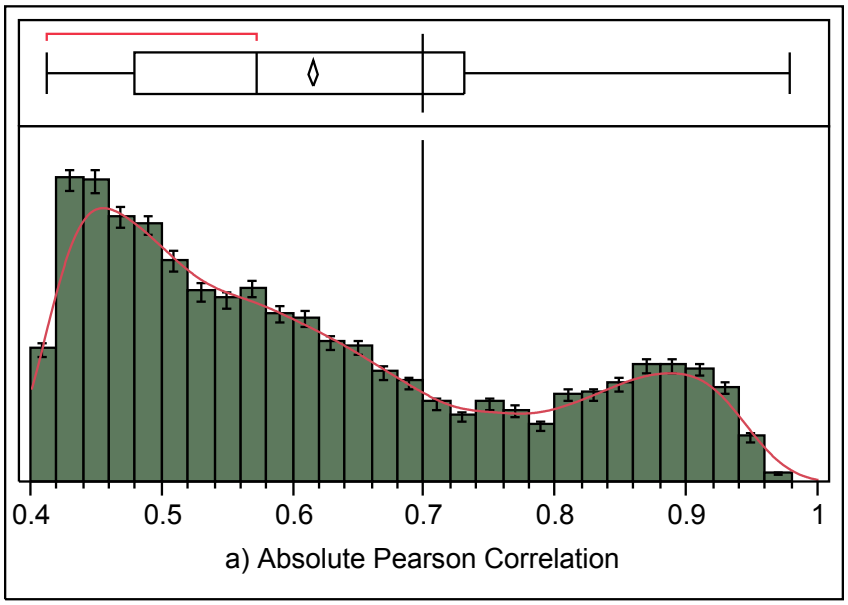
To extract paracliques we started with a graph obtained from Pearson's correlation matrix amongst all the genes and filtered it by using four different thresholds. APC coefficient thresholds corresponding to p-value of 0.01 were used for EPS (r=0.413) and EPL (r=0.41) for generating one graph (G0) each from spleen and liver data. APC coefficient thresholds of 0.6, 0.75 and 0.875 were used to generate three more graphs G1, G2 and G3 respectively for EPS as well as EPL. For each gene with degree greater than 500 in G0 we calculated APC corresponding to its 90th percentile. Edges above the $90^{th}$ percentile of a gene were retained if the $90^{th}$ percentile of that gene was greater than 0.6. These additional edges were merged with edge lists for G2 and G3 to obtain two enhanced graphs (EG2 and EG3).

Paracliques were extracted from these graphs using Grappa, a graph algorithms toolkit developed at University of Tennessee. Grappa uses FPT based algorithms for extraction of paracliques as described previously [169]. Gene enrichment analysis was done using Cytoscape's[324] package Bingo[204] and perl scripts using Gene Ontology (GO)[325] annotation for biological processes. Percentiles of APCs were regressed against degree of genes using the statistical software JMP [326]. Benjamini-Hochberg false discovery rate [201] corrected p-values were used for enrichment analysis. All source codes are available from the authors.

## RESULTS

We used t-test to determine the statistical significance of Pearson's correlation between gene pairs at a p-value of 0.01 for both the expression profiles (EPS and EPL). This corresponded to APC (Absolute Pearson Correlation) of 0.413 and 0.41 for EPS and EPL, respectively. Using these thresholds we generated one graph each for EPS and EPL (G0). From G0 we extracted the edges having evidence of interaction in literature (Transcription factor-Target pairs and other gene-gene interaction pairs obtained from Biogrid and Amadeus). The distribution of APCs of these gene pairs varied from 0.413 to 0.98 for EPS. The mean and median of this distribution were 0.61 and 0.57 respectively (Figure 3-1 a). Similarly, for EPL the APC varied between 0.41 to 0.97 with a mean and median of 0.53 and 0.51 respectively (Figure 3-1 b). Only 10.16% and 0.2% of these gene pairs had an APC greater than the traditionally used threshold of 0.875 in EPS and EPL data respectively. In both the tissues the distribution of APCs was skewed towards lower value indicating higher number of gene pairs with lower correlation.

Further we investigated the distribution of APCs of genes associated with known pathways and gene regulatory networks (GRNs). We selected pathways from KEGG and known gene regulatory networks for which we had data for at least five genes and ten edges.



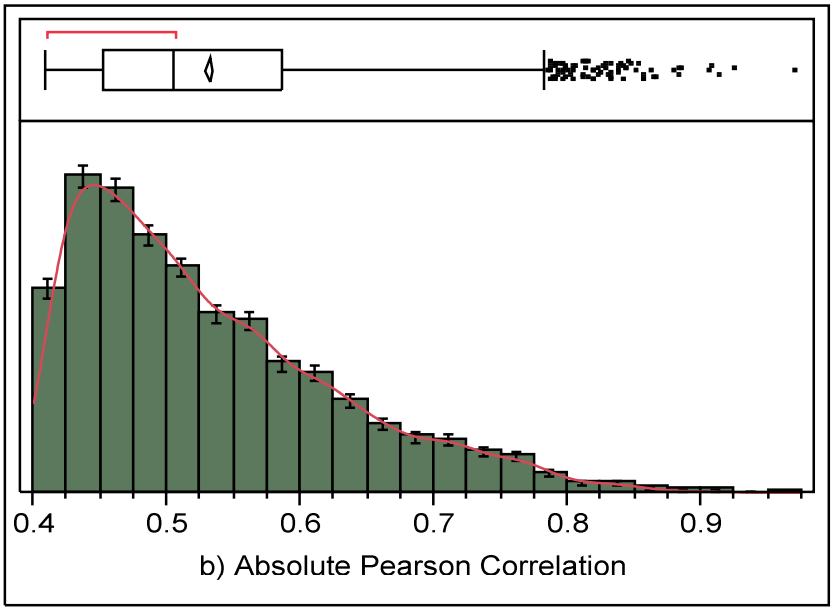




**Figure 3-1: Distribution of Absolute Pearson Correlations of gene pairs in EPS (Expression Profile of Spleen) and EPL (Expression Profile of Liver).** For EPS (a) the mean and median of the distribution were 0.61 and 0.57 respectively. The mean and median of APCs for gene pairs from EPL (b) are 0.53 and 0.51 respectively.

In EPS 189 pathways/GRNs met this criterion. Similarly, 248 pathways/GRNs were selected for EPL. Distributions of means of APCs among pathways/GRNs in EPS and EPL were plotted. For EPS this distribution varied between 0.47 and 0.74. The mean and median of this distribution was 0.52 and 0.53 respectively (Figure 3-2 a). The range of mean APCs for 248 pathways/GRNs relating to EPL varied from 0.47 to 0.65. Both mean and median of this distribution equaled 0.52 (Figure 3-2 b). The distribution of APCS for genes belonging to known pathways and GRNs suggests existence of important biological information at low APCs.

Gene networks are believed to be enriched with genes of high degree called hubs. We evaluated the relationship between degree (number of neighbors) of a gene and APC coefficients of that gene with its neighbors in G0 to investigate the influence of hub genes . The degree of genes was regressed against the $50^{th}$, $70^{th}$, $75^{th}$, $80^{th}$, $85^{th}$ and $90^{th}$ percentiles of APCs of that gene with its neighbors. Regression analysis revealed a significant positive relationship between degree of genes and APCs at an alpha of 0.01 for both the tissues. The vertex degree explained 61 to 82 percent of variance in the percentiles of APCs in EPS (Figure 3-3) and 47 to 76 percent of variance in the percentiles of APCs in EPL (Figure A3-1). To rule out the possibility of random relationship between degree of genes and percentiles of APCs we extracted 400 random samples of edges varying from size 1 to 5000 from G0. None of $50^{th}$, $60^{th}$, $70^{th}$, $80^{th}$, $85^{th}$ or $90^{th}$ percentiles of the random samples exhibited an increase in correlation value with increasing size of sample in any of the two datasets (Figures A3-2 and A3-3).

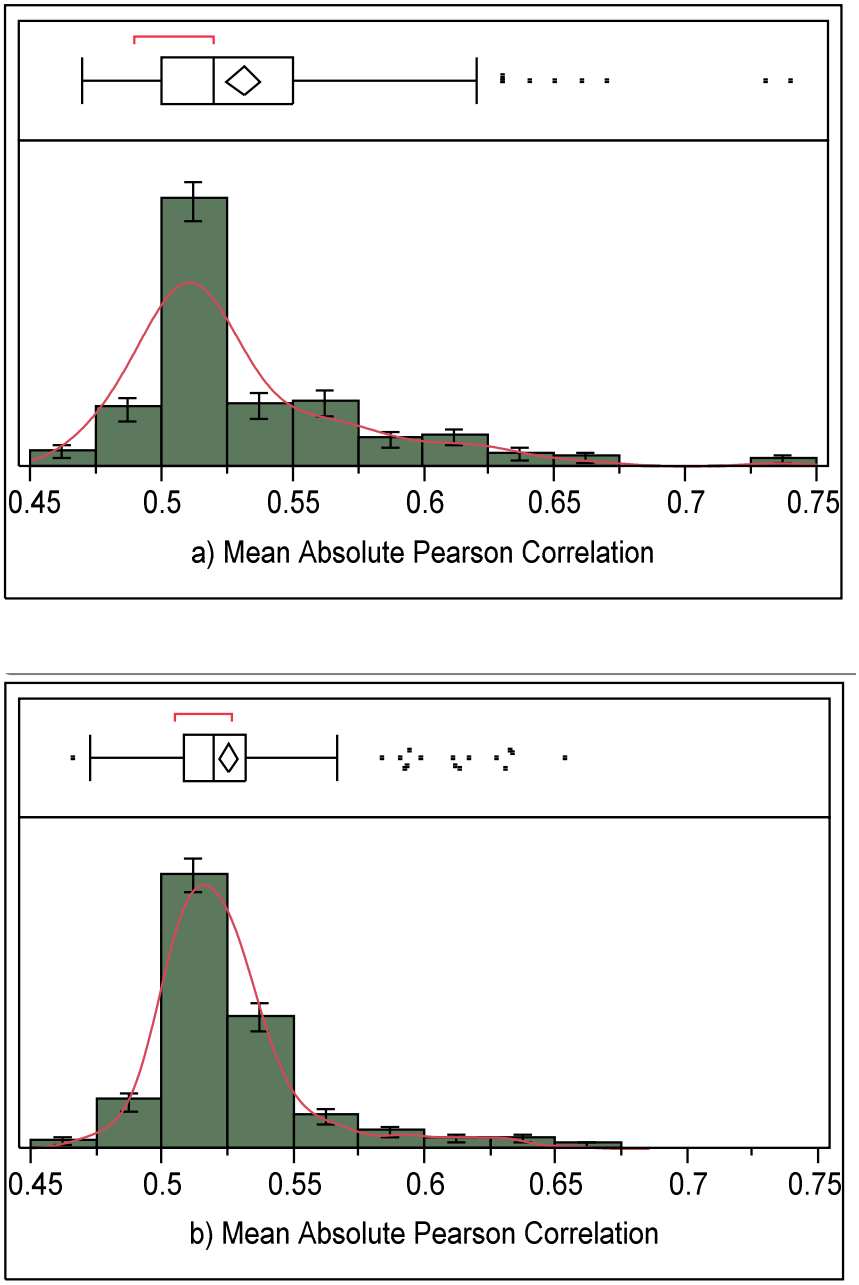



**Figure 3-2: Distribution of mean absolute Pearson correlations among genes of biological pathways and gene regulatory networks.** The mean of the distribution was 0.52 in EPS (a) and EPL (b). The median of the distributions were 0.53 and 0.52 for EPS and EPL respectively.

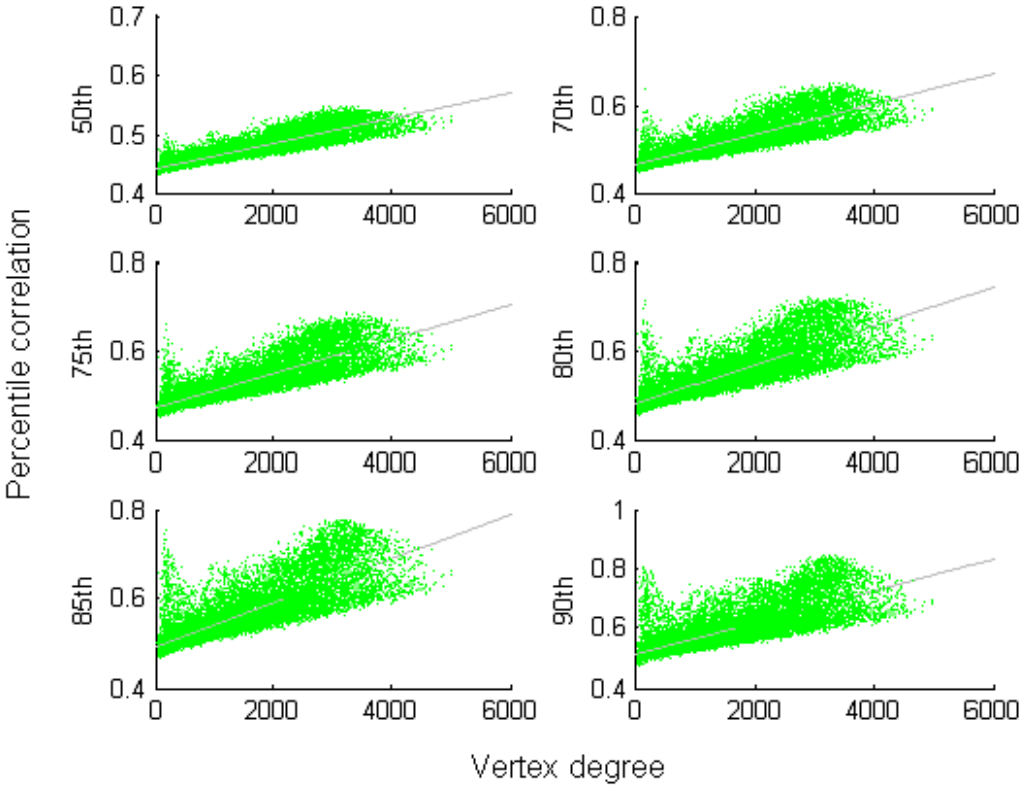



**Figure 3-3: Relationship between vertex degree of each gene and percentiles of its absolute Pearson correlations with its neighbors in expression profile of spleen.** A best fit least squares regression line is shown in blue.

This suggests that the increase in APC percentiles with increasing degree of genes is not an artifact of the data and may indicate biological information in the hubs. In view of the positive relationship between degree of genes and their APCs with their neighbors we investigated the effect of inclusion of hubs on enrichment of genes pertaining to common biological processes. We selected hub genes with high degree (>500) and high APCs (90th percentile of APC of the hub gene with its neighbors>0.6). For each hub the edges with APC greater than 90th percentile of that hub were retained. These additional edges were merged with edge lists for the graph with threshold 0.75 and 0.875. We called these two graphs as hub enhanced graphs. In effect two hub enhanced graphs were constructed by merging edges with top 10 percent APCs from the selected hubs with graphs obtained at a threshold of 0.75 (EG2) and 0.875 (EG3).

Graph algorithms were used to extract the paracliques from each of six graphs (G0, G1, G2, G3, EG2 and EG3). GO IDs were used to determine enrichment of biological processes in the extracted paracliques. As expected, the enhanced graphs EG2 and EG3 had higher number as well as higher size of paracliques as compared to G2 and G3 respectively (Figure 3-4 and Figure A3-4). Paracliques from EG2 were enriched with higher number of GOIDs pertaining to biological processes (4979 in EPS and 9075 in EPL) than paracliques from G2 (2110 in EPS and 3804 in EPL) in both the datasets. Similarly EG3 was enriched with higher numbers of GOIDs relating to biological processes (4716 in EPS and 9085 in EPL) as compared to G3 (547 in EPS and 845 in EPL) (Figure 3-4 and Figure A3-4).

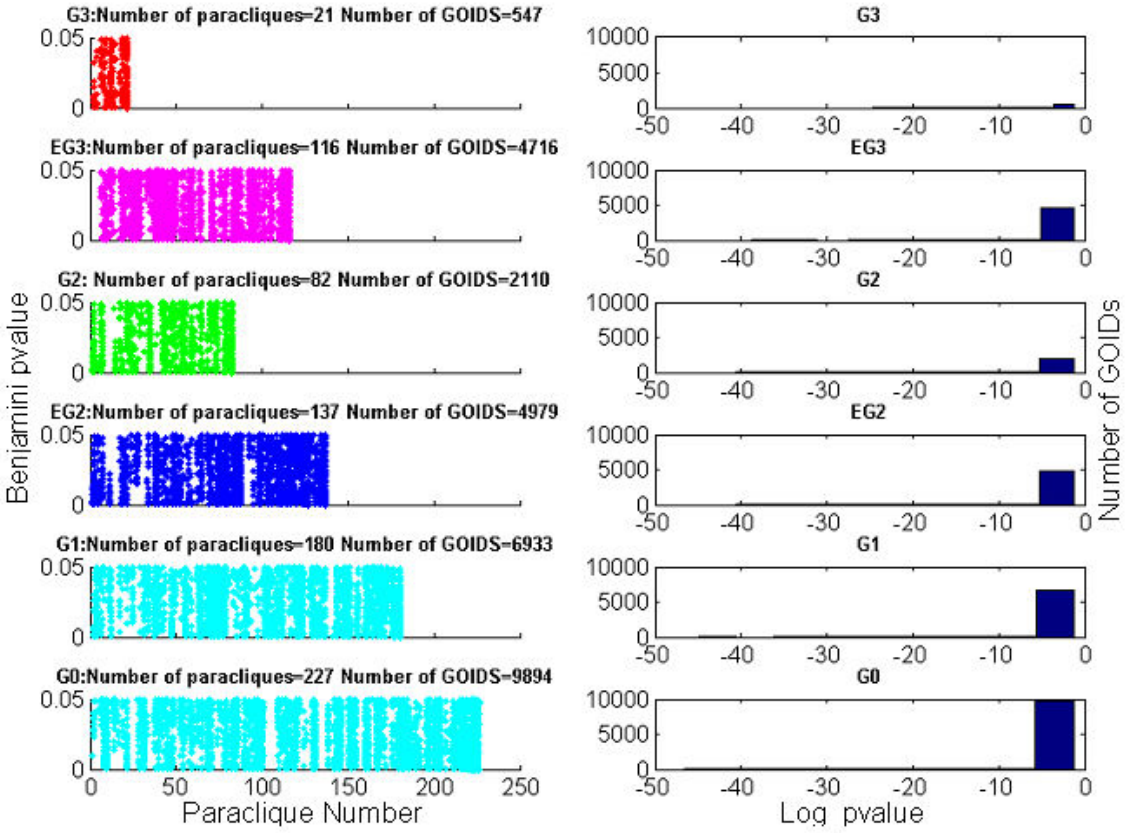



**Figure 3-4: Distribution of FDR corrected p-values associated with GOIDS corresponding to biological processes enriching the paracliques at different thresholds in expression profile of spleen.** Left panel: Paraclique wise distribution of GOIDS relating to biological processes in EPS. The paracliques were extracted from the graphs G0, G1, EG2, G2, EG3 and G3 (shown serially from bottom to top).The paracliques are enumerated along x-axis for each graph. GOIDs associated with each paraclique are represented by a dot above that paraclique's number. The position of each GOID along the Y-axis indicates FDR corrected p-value for the enriched biological process corresponding to that GOID. Right panel: Overall distribution of log of FDR corrected p-values associated with GOIDs corresponding to biological processes enriching all the paracliques obtained from the graphs G0, G1, EG2, G2, EG3 and G3 (shown serially from bottom to top).

Association of higher number of GOIDs with enhanced graphs EG3 and EG2 as compared to G3 and G2 respectively indicates enrichment of higher number of corresponding biological processes in enhanced graphs. The increased enrichment of genes belonging to related biological processes in EG2 and EG3 reconfirms the importance of hub genes in the network. The paracliques extracted from the graphs obtained at a threshold of 0.413 (G0) were significantly associated with highest number (9894) of GOIDS corresponding to biological processes in EPS (Figure 3-4). Similarly in case of EPL paracliques obtained at a threshold of 0.41 were associated with maximum number (14985) of GOIDS (Figure A3-4). This implies that amongst these six graphs, the one with the lowest threshold (G0) produces paracliques enriched with maximum number of biological processes.

We also compared maximally enriched paraclique (those with maximum number of GOIDs associated with them) obtained from G0 (Pg0) with paracliques generated from G1, G2 and G3 that contain genes belonging to Pg0. Genes belonging to more than one paraclique from G1(Pg1s), G2 (Pg2s) or G3 (Pg3s) were present in Pg0 in both the expression profiles (EPL and EPS). Log of FDR corrected p-values associated with GOIDs corresponding to biological processes enriching both Pg0 and Pg1s were compared and plotted for EPS as well as EPL (Figure 3-5). Similarly log of FDR corrected p-values for GOIDS corresponding to biological processes were compared between Pg0 and Pg2s as well as Pg0 and Pg3s.

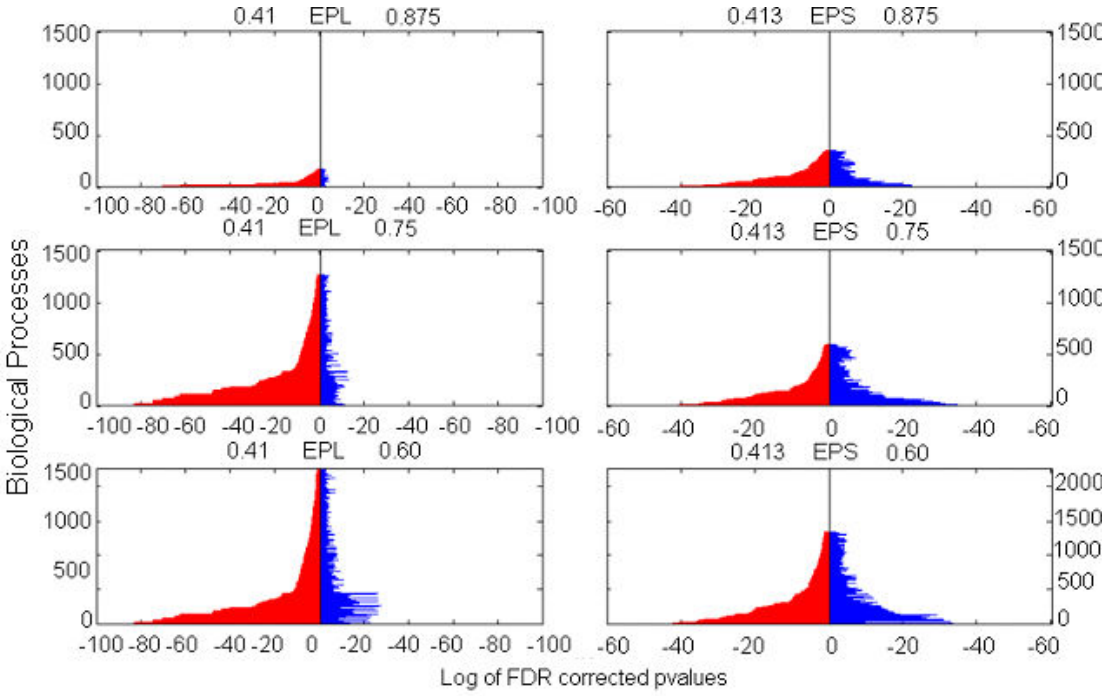



**Figure 3-5: Comparison of significance of enrichment of biological processes in paracliques obtained at different threshold from expression profiles of liver (EPL) and spleen (EPS).** Each figure compares log FDR corrected p-values associated with GOIDS common between maximally enriched paraclique in G0 (shown in blue, APC=0.41 in EPL and 0.413 in EPS) and paracliques in G1, G2 and G3 (shown in red, APC=0.6, 0.75 and 0.875 respectively) that contain subset of genes from G0. Figures on left correspond to EPL and those on right relate to EPS.

It is apparent from Figure 3-5 that the FDR corrected p-values for biological processes for Pg0 was either better or comparable to Pg1s, Pg2s and Pg3s in EPS as well EPL. Comparison of Pg0 with Pg3s revealed that in case of EPS nearly 83% of GOIDs associated with biological processes had more significant pValue in Pg0 than in Pg3s. Among biological processes with less significant pValues in Pg0, 65% had higher number of genes in Pg0 as compared to Pg3s. Similarly, in case of EPL, over 83% of GOIDs associated with biological processes had more significant pValue in Pg0 as compared to Pg3s. Among the GOIDS with less significant pValue in Pg0, over 96% had higher number of genes in Pg0 as compared to Pg3s. Thus lower or equivalent p-values in larger sized paraclique Pg0 in both EPS and EPL implies clustering of larger number of genes participating in common biological processes in Pg0 as compared to Pg1s, Pg2s and Pg3s.

## DISCUSSION

We investigated the gene expression profiles from liver and spleen tissues of BXD RI mice by extracting paracliques from six graphs G0, G1, G2, G3, EG2 and EG3. We have demonstrated that paracliques extracted from the graph G0 generated at statistically significant (p value < 0.01) correlation thresholds of 0.41 (in EPL) or .413 (in EPS) were enriched with highest number of GOIDs corresponding to biological processes. The lower threshold not only increases the size of paracliques but may also increase the number of biologically relevant genes in those paracliques. This may be due to coalescence of biologically relevant parts of gene networks into larger

paracliques at the lower threshold. Both technical and biological reasons can explain relatively lower APC among the genes participating in common gene networks. Microarrays take a snapshot of gene expression profile at a given time and not across a time continuum. The gene to gene correlations on the other hand can be affected by time lag [327, 328]. Even a small time lag can result in decreased correlation. Relationship among the coexpressed genes may also be reflected as low APC because of complex interaction of genes in biological networks like presence of feedback loops. Some of the interactions are affected by inhibitors while some others may be controlled by silencing mechanisms such as micro-RNAs [329]. Genetic diversity of BXD population and absence of external stimulus in baseline gene expression profiles may also be factors contributing to lower APCs between genes in microarray data from genetical genomic studies. The availability of large number of arrays associated with the genetical genomic studies provides us an opportunity to extract biological signal by decreasing the threshold of APCs among the genes.

## CONCLUSIONS

We chose statistically significant APC (pvalue<=0.01, r=0.413 for EPS and r=0.41 for EPL) threshold for investigating gene enrichment in two different datasets of BXD RI mice and compared them with higher APC thresholds of 0.6, 0.75 and 0.875. Graph generated at threshold of 0.413 (or 0.41 in EPL) resulted in higher number and bigger sizes of paracliques as compared to paracliques extracted from graphs thresholded at APCs 0.6, 0.75 or 0.875 in EPL (or EPS). The lower threshold improved the enrichment

of extracted gene networks than the higher threshold as indicated by increase in number and significance of biological processes associated with the paracliques. Enhancement of graphs with hub genes also enriches the paracliques with genes belonging to biologically related processes. Thus we have shown that a low statistically significant APC threshold can be used for extracting gene networks from baseline gene expression profiles obtained from a population of genetically different strains of BXD mice.

# APPENDIX

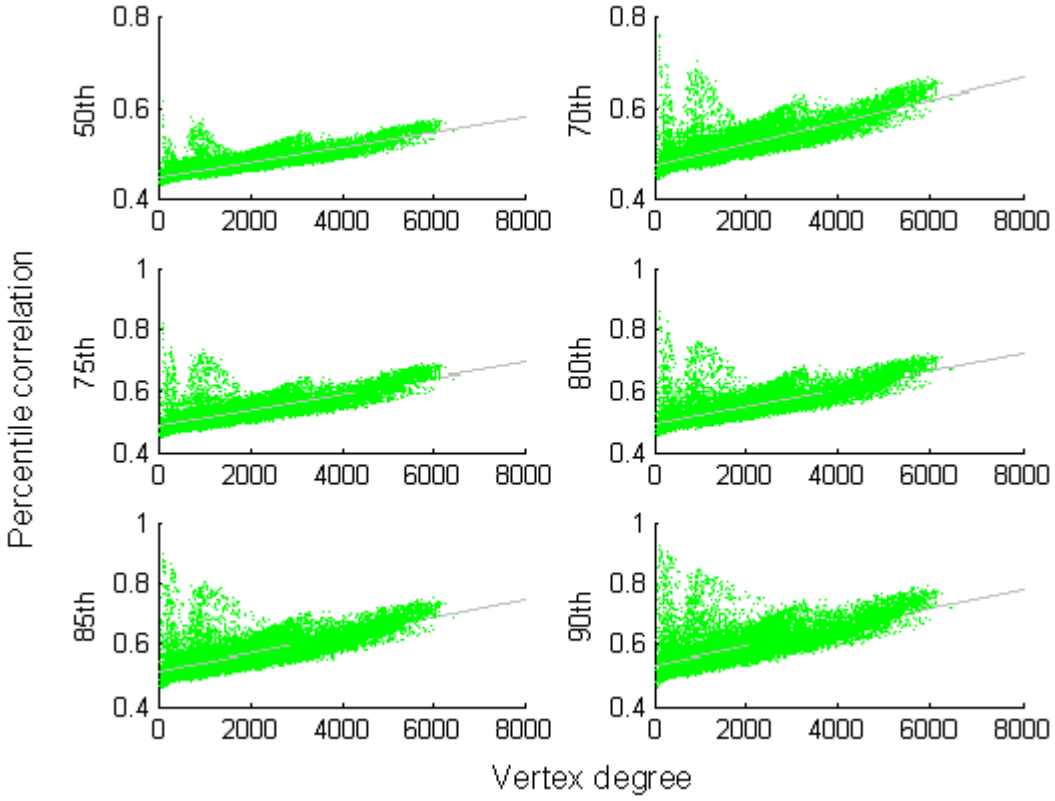



**Figure A3-1: Relationship between vertex degree of each gene and percentiles of its absolute Pearson correlations with its neighbors in expression profile of liver.** A best fit least squares regression line is shown in blue.

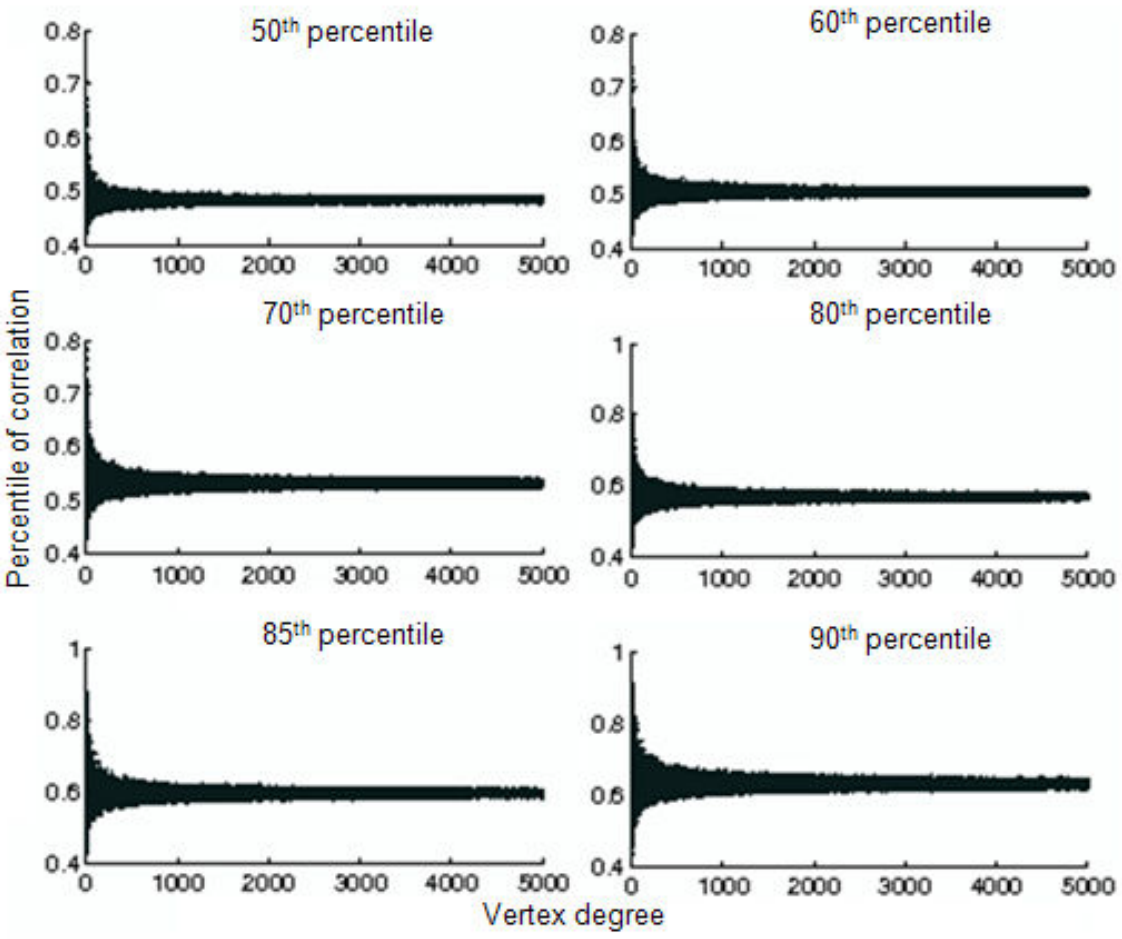



**Figure A3-2: Percentiles vs. Pearson Correlations of random samples of different sizes from EPS (expression profile of spleen).** Percentiles do not increase with increasing number of randomly drawn edges.

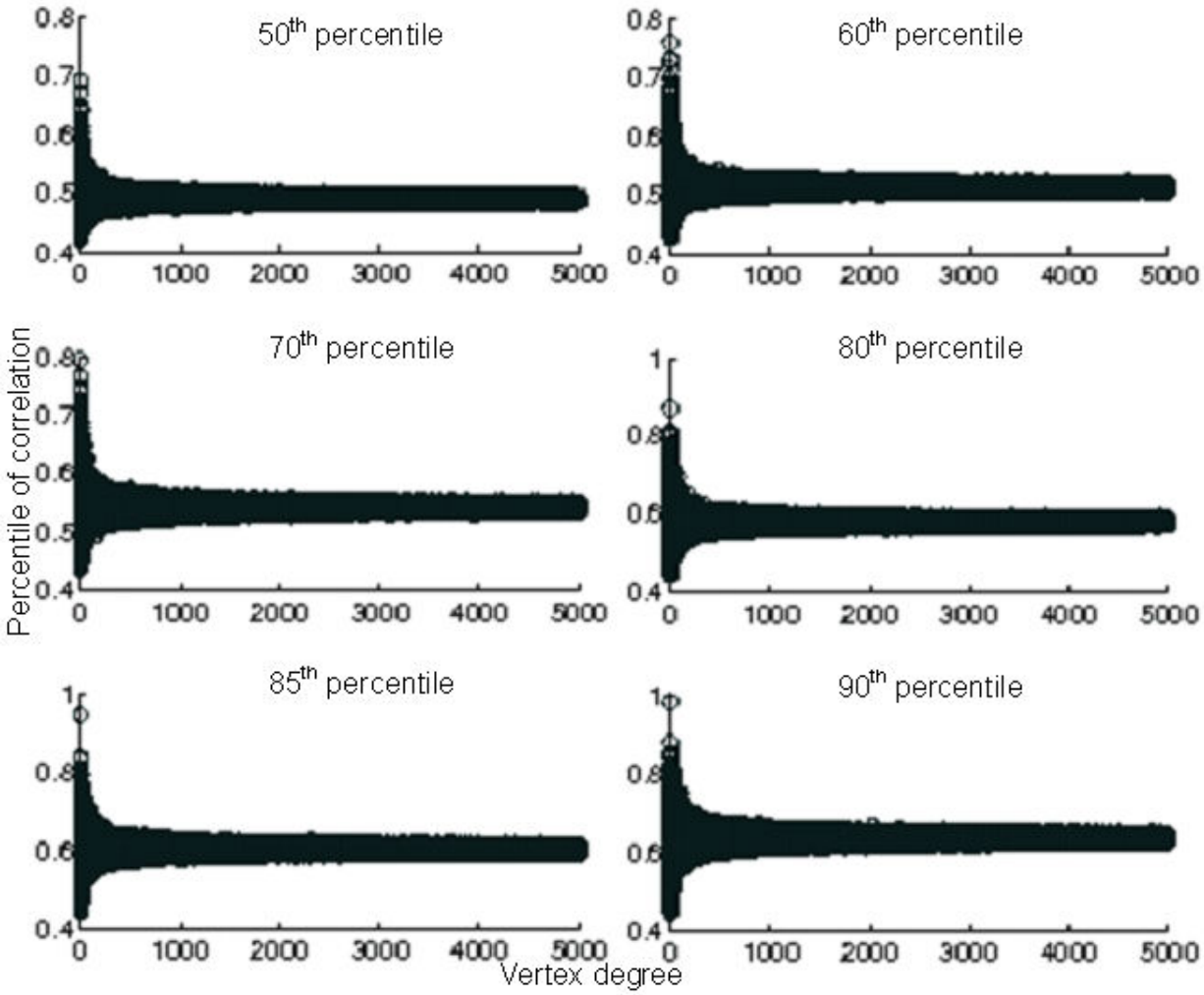



**Figure A3-3: Percentiles vs. Pearson Correlations of random samples of different sizes from EPL (expression profile of liver).**Percentiles do not increase with increasing number of randomly drawn edges.

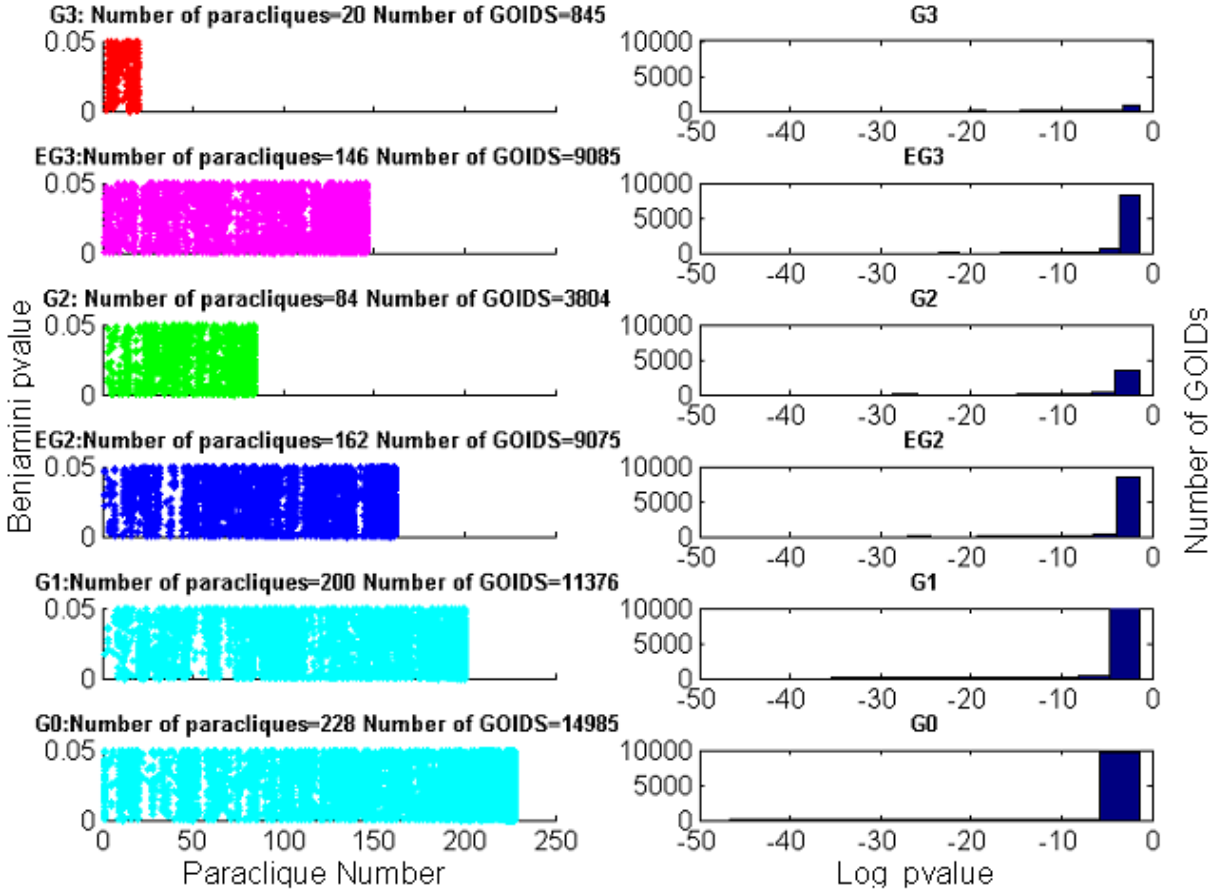



**Figure A3-4: Distribution of FDR corrected p-values associated with GOIDS corresponding to biological processes enriching the paracliques at different thresholds in expression profile of liver (EPL).**Left panel: Paraclique wise distribution of GOIDS relating to biological processes in EPL. The paracliques were extracted from the graphs G0, G1, EG2, G2, EG3 and G3 (shown serially from bottom to top) The paracliques are enumerated along the x-axis for each graph. GOIDs associated with each paraclique are represented by a dot above that paraclique's number. The position of each GOID along the Y-axis indicates FDR corrected p-value for the enriched biological process corresponding to that GOID. Right panel: Overall distribution of log of FDR corrected p-values associated with GOIDs corresponding to biological processes enriching all the paracliques obtained from the graphs G0, G1, EG2, G2, EG3 and G3 (shown serially from bottom to top).

# CHAPTER 4 : SYSTEMS GENETICS APPROACH TO UNCOVER THE EFFECTS OF LOW DOSE IONIZING RADIATION

(This chapter gives a brief overview of two low dose ionizing radiation papers. The lead author of these papers is Rachel M. Lynch (Appendix following chapter 4 and [187]). Work related to regression analysis, eQTL analysis and identification of trans and cis QTL bands, microarray data analysis and differential expression was done by S. Naswa)

Radiation response effects large numbers of biological processes and genes related to immune system, apoptosis, cell proliferation, DNA damage, cancer, etc. There are conflicting beliefs about the biological effects of LDIR that vary from hormesis and adaptive response to harmful effects. The low magnitude of LDIR, presence of background natural radiation and the inherent genetic variation at population level make LDIR response a suitable candidate for investigation using systems genetics approach. Here we investigated the effects of LDIR at a population level using systems genetic approach.

Since the immune system is considered to be sensitive to ionizing radiation we began by investigating genetic signatures responsible for variability in immunophenotypes in a genetic reference population. We used recombinant inbred strains of BXD mice derived from a cross between C57BL/6J X DBA/2J mice to investigate the variability in various parameters of immune system. Variability in percentages of T cells and their subtypes (CD4+, CD3+, CD8+), B cells and the ratios of these cells (LN T: B and LN CD4: CD8) was collated with QTL analysis and gene expression profiles from BXD mice. Multiple regression modeling of the correlates of genes neighboring statistically significant QTLs revealed three candidate genes (Ptprk, Acp1 and Lamb1-1) explaining 61% variance of ratio of helper (CD4+) and cytotoxic (CD8+) T cells.

After baseline expression profiling in BXD population, we investigated differences in response of the inbred parental strains (C57BL/6J & DBA/2J) to LDIR. Expression profiles in spleen tissue of irradiated (10cGy and 1Gy) and sham irradiated mice were obtained using Illumina microarrays. Analysis of variance (ANOVA) and post ANOVA

contrast tests were used to test effects of dose, strain and their interaction on gene expression. GO enrichment of genes differentially expressed in response to LDIR revealed that immune system processes exhibited radiation effect in DBA/2J. Genes related to neutrophil function were differentially expressed after exposure to LDIR in both the strains but in opposite direction. The parental strains also exhibited the effects of radiation on immune system at cellular level. LDIR significantly enhanced the percentage and activity of neutrophils in peripheral blood.

# APPENDIX

## Systems genetics approach to low dose radiation sensitivity in BXD recombinant inbred mice

(May be published in radiation research with following authorship)

Lynch RM, Naswa S, Rogers GL, Jr., Kania SA, Das S, Chesler EJ, Saxton AM, Langston MA, Bogard JS, and Voy BH.

## ABSTRACT

Radiation protection guidelines are designed to protect the population from exposure to harmful doses of radiation. Defining a harmful dose is clear for exposures for which the effects permit epidemiological detection. However, defining the lower level of exposure that continues to have adverse effects is complicated. Low dose radiation (LDR) engages many pathways that mediate normal cellular functions, making it challenging to detect radiation-specific effects even in simplistic models such as tissue culture. In a population, assessing risk is further challenged by the complexity of in vivo exposures and by genetic variation inherent to individuals. We present initial results from modeling radiation sensitivity with a panel of recombinant inbred mouse strains and using systems genetics to extract mechanisms of heritable radiation sensitivity. Emphasis is on the immune system because of its inherent radiosensitivity and its potential to impact other processes including malignancy. We demonstrate that exposure to 10 cGy ionizing radiation significantly enhanced neutrophil phagocytosis across strains. In contrast, genetic background impacted LDR-induced changes in spleen superoxide dismutase activity. Transcriptome data from spleens of the BXD parental strains highlighted the impact of genetic background on LDR responses. These data highlight the need to consider genetic variation when assessing LDR outcomes.

## INTRODUCTION

Average levels of radiation exposure over the past thirty years are estimated to have doubled, largely due to the widespread use of diagnostic imaging procedures such as

computed tomography (*1*). Escalating clinical use of radiation, coupled with recent concerns about exposure from airport security scanners and now from the earthquake-driven radiation leaks from Japanese nuclear power plants, heightens concern about potential health effects of exposure to radiation at doses that are measurable and increasing, but still very low relative to those known to be carcinogenic. Determining the potential health consequences of low dose radiation exposure in a human population is complicated by a number of factors: the biochemical intersection between the effects of low dose radiation and many other environmental stressors, the differential effects of lifestyle variables that impact the response to stress, and the underlying genetic variation within a population. Current radiation protection guidelines are based on linear extrapolation of risk from dose. Biologically, however, low dose ionizing radiation elicits both molecular and higher order phenotypes that are not necessarily observed at higher doses (*2-5*). Therefore the physiological consequences of low dose exposures are not easily predicted using a linear model.

The immune system illustrates the challenges in delineating health effects of low dose radiation exposure. High radiation doses (>1 Gy) suppress immune function through destruction of myeloid and lymphoid cell populations in bone marrow (*6*). In contrast, several studies suggest that low doses of radiation enhance functions of various immune cell populations that could be beneficial to the organism, at least acutely (*7*). For example, LDR has been shown to increase mitogen-induced lymphocyte proliferation (*8-12*), macrophage and natural killer cell activation (*11, 13-16*), and tumor surveillance (*7, 14, 17-20*). At the molecular level, LDR alters gene expression (*21*),

cytokine secretion (*13, 22, 23*), expression of surface molecules on immune cells (*22, 24*), and apoptosis (*12*), which can lead to LDR-induced modification of leukocyte distribution (*25*). Relatively little is known, however, about how this sensitivity translates into efficacy of the immune system.

Ultimately, risk of LDR exposure must be applied to a population, which requires an understanding of the contribution of genetic variation to radiation response. Mouse models provide the most direct evidence that genetic background confers inter-individual differences in radiosensitivity that are acknowledged, but more difficult to study, in humans. Using inbred strains of mice, Roderick (1963) demonstrated the importance of genetic background on viability following daily 1 Gy doses of X-ray radiation; survival time in the most sensitive and resistant strains differed by more than two-fold (*26*). Since then, differences between strains at sublethal radiation doses have been reported for a number of outcomes, including radiation-induced apoptosis (*27-31*) and carcinogenesis (*32-35*). A more limited number of studies have reported differential effects of low radiation doses in inbred mouse strains and in cell lines derived from a panel of human donors (*36, 37*).

Panels of recombinant inbred (RI) mouse strains provide the means to study the impact of environmental factors such as low dose radiation in the context of genetic variation and to simultaneously screen for loci that contribute to differential outcomes of the exposure. The BXD (C57BL/6J X DBA/2J) RI strain set is the largest existing set of inbred mouse strains. The parental strains (C57BL/6J and DBA/2J) differ in their sensitivities to LDR exposure (*31, 38-40*), making the BXD panel attractive for both

studies of LDR in a population-based model and for identification of genetic loci linked to variation in the LDR response. Here we report the first steps in using a systems genetics framework in the BXD panel to uncover the basis for differential genetic sensitivity to LDR exposure. Systems genetics is an analysis framework that exploits correlation between traits to assemble multi-level networks, from the molecular level through intermediate traits to overlying, systems level phenotypes (*41-43*). Our overarching goal is to iteratively assemble a systems level view of LDR sensitivity that encompasses both the initial response to radiation stress and later, potentially prolonged physiological outcomes that may include both hermetic and detrimental effects. Stable genetic reference populations, such as the BXD lines, are valuable tools for this approach because they allow data to be integrated over time and from multiple experiments. We began by addressing two objectives, the results of which are described herein. Our first objective was to test the consequences of LDR on peripheral blood mononuclear cell phagocytosis of bacteria, a functional measure of the innate immune systems that could be measured ex vivo after whole body irradiation without the need for cell culture. This objective was based in part on a previous study in our lab that suggested strain-specific effects of low dose X-ray exposure on genes related to immune function (*44*). Our second objective was to determine if genetic variation significantly altered the oxidative stress defense response to LDR, based on the central role of reactive oxygen species as mediators of radiation effects at lose doses (*45-49*). Efforts were focused on a limited number of biochemical and functional endpoints that could be assayed efficiently across a large number of mice. Existing genotype data

available for the BXD panel were used to identify putative genetic loci linked to differences in LDR traits. Transcriptomic profiling of spleen from the parental strains was included to screen for additional differences in LDR response between strains that will guide future studies of differential LDR sensitivity using the BXD strain panel.

## METHODS

### RADIATION EXPOSURE

C57BL/6J and DBA/2J mice were obtained from the Jackson Laboratory (Bar Harbor, ME). A total of 39 BXD RI strains were used for this study. Stocks were obtained from The Jackson Laboratory and Drs. Lu Lu and Robert Williams at the University of Tennessee Health Science Center (UTHSC, Memphis, TN). This population represents a mixture of the strains from the original Jackson Laboratory strains (*50, 51*) and the advanced intercross strains developed at UTHSC (*52*). Mice were housed and propagated in the specific-pathogen-free (SPF) Russell Vivarium at Oak Ridge National Laboratory (ORNL) as previously described (*53*). Approximately 10 week-old mice were exposed to a single whole-body 10 cGy dose of radiation from a $^{137}$Cs source delivered at a rate of ~9 cGy/h. Each strain by treatment group consisted of an average of 4 irradiated or 4 sham-exposed control mice per strain, and each group was balanced between males and females. Only 2 mice (1 irradiated and 1 sham control) per strain were exposed on any given day, and strains were randomized across the study. Following radiation or sham exposure, mice were housed for 48 h in a satellite facility prior to dissection. Blood was collected by retro-orbital sinus puncture into heprinized

tubes for neutrophil function assays. Spleens were harvested and snap-frozen in liquid nitrogen and stored at -80°C for subsequent biochemical assays.

For spleen expression profiling, male C57BL/6J and DBA/2J mice were exposed to either a low dose (10 cGy, as described above), or a high dose (1 Gy) of whole-body γ-radiation delivered by a $^{60}$Co source with a dose rate of ~6 Gy/min. Mice were sacrificed 24 h following exposure, and spleens were stabilized in RNAlater (Sigma-Aldrich, St. Louis, MO) until RNA was extracted. All studies were approved by the Animal Care & Use Committee at Oak Ridge National Laboratory.

#### NEUTROPHIL FUNCTIONAL ASSAYS

Flow cytometry was used to assay neutrophil function in peripheral blood from 34 BXD strains 48 h after sham or radiation exposure. For both assays, red blood cells in the blood samples were lysed and leukocytes were fixed prior to flow cytometric analysis. DNA staining was used to distinguish between aggregation artifacts and murine cells. At least 10,000 leukocytes per sample were analyzed using a Beckman Coulter Epics XL flow cytometer and EXPO32 ADC Software (Beckman Coulter, Brea, CA). Neutrophils were gated for analysis based on forward and side scattering profiles. Gating based on fluorescence was set on unstimulated samples from each mouse to include approximately 10% of the evaluated cell population, and the same gating parameters were used to evaluate percentage and median channel fluorescence (MCF) of stimulated neutrophils exhibiting phagocytic or oxidative burst activity. Neutrophil phagocytosis (Phagotest Kit, Orpegen Pharma, Heidelberg, Germany) and oxidative

burst assays (Phagoburst Kit, Orpegen Pharma) were performed as previously described (*54*).

### BIOCHEMICAL ASSAYS

Response to oxidative stress was assayed by quantification of superoxide dismutase (SOD) activity, glutathione (GSH), and oxidized glutathione (GSSG) in spleens from 39 BXD strains. Spleens were homogenized in 1mL of cold HEPES buffer (20mM HEPES, pH 7.2, with 1mM EGTA, 210mM mannitol, and 70mM sucrose). The homogenate was aliquoted for SOD, GSH, and Bradford assays. SOD activity was measured in the spleen extracts using an enzymatic assay (Cayman Chemical Company, Ann Arbor, MI) that reflects the combined activity levels of all three SOD isoforms (SOD1, SOD2, and SOD3), normalized to the protein concentration of the spleen extract (Bio-Rad Protein Assay, Hercules, CA), and reported as the units of SOD activity per mg of protein (U/mg). The percentage of GSSG to total GSH in deproteinated spleen extracts was determined with a kit which utilizes an enzymatic recycling method using glutathione reductase (Cayman Chemical Company). GSSG levels were assayed separately from the determination of total GSH levels; both were assayed according to manufacturer's instructions using the end-point method. The percentage of GSSG to total GSH was determined, as well as the GSSG and total GSH concentration normalized to the protein concentration of the original spleen lysate.

### QTL MAPPING

Quantitative trait loci (QTL) mapping was performed on SOD activity in sham and LDR-exposed BXD mice. SOD activity data adjusted for assay date differences by the models described below were used for QTL mapping. QTL analysis was performed using nearly 3,800 single-nucleotide polymorphisms (SNPs) and microsatellite markers for the BXD panel obtained from GeneNetwork database (http://www.genenetwork.org/dbdoc/BXDGeno.html). The genotype information was based on the markers originally reported by Shifman et al. (*55*) which were re-aligned with National Center for Biotechnology Information (NCBI) Build 36. QTLs were identified using WebQTL (*56*), which creates a linkage map using Haley-Knott regression and interval mapping. Genome-wide significance thresholds were calculated based on 1,000 permutations (*57*), and the cut-off *p*-values for significant and suggestive loci were $P = 0.05$ and $P = 0.63$, respectively (*58*). Multiple-QTL modeling was performed using stepwise linear regression in SAS; a *p*-value of 0.05 was used as the threshold for terms to remain in the final model.

### PARENTAL GENE EXPRESSION PROFILING

Transcriptome profiling in C57BL/6J and DBA/2J spleens (24 h post-exposure) was performed by Genome Quebec (Montreal, Canada) using the Mouse WG-6 v1.1 BeadChip on the Illumina platform (San Diego, CA) as previously described (*53*). Four mice per strain per radiation exposure (i.e. control, low dose, high dose) were used for transcriptome profiling, except there were only three DBA low dose radiation exposed samples. Quantitative polymerase chain reaction (Q-PCR) was used to confirm the

microarray results of differential expression of several genes in the parental strains following radiation exposure. Reverse transcription was performed on 500ng of RNA using the Bio-Rad iScript cDNA Synthesis kit from Bio-Rad. QuantiTect primers were used in conjunction with the QuantiTect SYBR Green PCR kit (Qiagen) on a CFX96 real-time PCR detection system (Bio-Rad). All samples were analyzed in triplicate; gene expression was normalized to hypoxanthine guanine phosphoribosyl transferase (*Hprt)* expression. Microarray analysis and Q-PCR confirmed that *Hprt* expression did not vary across our experimental groups. Fold changes were calculated based on the ΔΔCt method and differences between groups were tested using ANOVA.

**Statistical testing**

Statistical testing was performed using SAS (SAS Institute, Cary, NC). For biochemical and neutrophil function assays, Proc Mixed was used to test for strain, radiation, and strain by radiation interaction effects (strain*radiation), using the assay date as a random variable. The “assay date” term was included because the radiation exposures and assay dates were performed across several weeks to accommodate the large population of mice used in this study. If the strain*radiation term was not significant ($P > 0.05$), then a reduced model was rerun using only the strain and radiation terms. All neutrophil function and biochemical data reported are least squares means from the Proc Mixed model.

Expression data from Illumina bead chips were normalized using Variance Stabilizing Transformation (VST) followed by Robust Spline Normalization (RSN) using the R/Bioconductor (*59*) package lumi (*60*). Fold changes were calculated using the

inverseVST function in the lumi package. Raw and normalized expression data are available through NCBI's GEO database (http://www.ncbi.nlm.nih.gov/projects/geo, Accession GSE21562). SAS procedure GLM was used for analysis of variance (ANOVA) to test the effects of strain, dose and their interaction on expression. Post ANOVA contrast tests were used to compare the groups. An alpha of 0.05 was used for all statistical tests. False discoveries due to multiple comparison testing were controlled by using q-value (*61*). Differential expression was considered significant if both p and q values were < 0.05 and fold change ≥ |1.5|. DAVID (*62, 63*) was used for gene ontology (GO) enrichment analysis of differentially expressed genes. Benjamini-Hochberg false discovery rate-corrected p-values are reported (*64*).

## RESULTS

### NEUTROPHIL FUNCTION

Phagocytosis of foreign material by neutrophils provides a first line of defense through which the innate immune system protects the body against invading pathogens (*65, 66*). Ex vivo measures of phagocytosis provide a means to profile functional activity of the innate immune system (*67*). Phagocytosis and oxidative burst were measured across a panel of irradiated and control BXD mice to determine if LDR altered functional activity of peripheral blood neutrophils and, if so, if genetic variation differentially impacted this response. LDR exposure significantly increased both the percentage of phagocytic neutrophils (i.e., phagocytosis of one or more FITC-labeled bacteria per cell; $P = 0.044$) and the median channel fluorescence (MCF) of phagocytic neutrophils, reflecting the

number of bacteria phagocytosed per *E.coli*-positive cell ($P$ = 0.019, Figure A4-1). The radiation effect translates into approximately a 4% increase in the number of cells undergoing phagocytosis and an 11% increase in the number of bacteria engulfed by those cells, relative to sham controls. In addition to a main effect of radiation, both measures of phagocytic activity showed significant effects of strain (% Phagocytic Neutrophils, $P$ < 0.001 and Phagocytic MCF, $P$ = 0.002), indicating genetic variation in phagocytic function in the BXD panel. Despite the wide range of baseline variation across strains, genetic background did not significantly modify the effect of LDR on phagocytic activity (strain*radiation, $P$ > 0.05).

Flow cytometry also was used to measure the generation of intracellular reactive oxygen species (ROS) generated during oxidative burst. The percentage of oxidative burst-positive neutrophils (%OB Neutrophils) was analyzed as well as the MCF of positive neutrophils (measurement of enzymatic activity). While both %OB Neutrophils and OB MCF varied significantly by strain ($P$ = 0.020 and $P$ < 0.001, respectively), radiation and strain*radiation interaction effects were not significant ($P$ > 0.05).

#### ANTI-OXIDANT DEFENSE SYSTEM

Superoxide dismutase, catalase and the tripeptide glutathione act as an endogenous system of defense against oxidative stress, including that which is produced by ionizing radiation (*46, 68*). Genetic variation in the ability to mitigate oxidative stress has been linked to increased susceptibility to inflammatory disorders and to the effects of environmental stimuli that increase free radical production (Reviewed in (*69*)). SOD

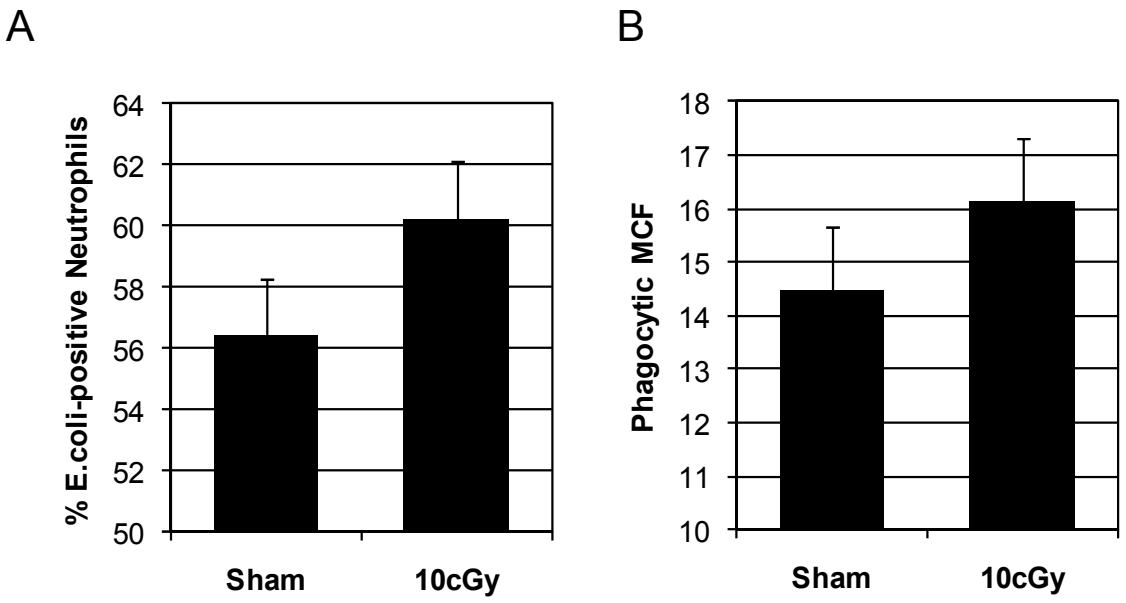



Figure A4- 1. Effect of radiation on neutrophil **function.** Peripheral neutrophils were assayed using flow cytometry 48 h after irradiation; N ≥ 66 mice / group (A) A greater percentage of neutrophils from radiation-exposed mice engulfed FITC-labeled *E. coli* compared to those of controls ($P$ = 0.044). (B) *E. coli*-positive neutrophils from radiation-exposed mice had a greater median channel fluorescence (Phagocytic MCF, indicating more bacteria engulfed per cell) compared to those of control ($P$ = 0.019). Error bars reflect the SEM.

activity and oxidized and total GSH were assayed in spleen from control and irradiated BXD mice to simultaneously screen for genetic variation in baseline oxidative defense capacity that could alter radiation sensitivity and to determine if LDR altered SOD activity, and in a manner that depended upon genetic background. Total SOD activity in spleen varied significantly across the BXD strain panel ($P < 0.001$). Genetic background further altered the SOD response to LDR, as indicated by the very significant strain*radiation interaction ($P < 0.001$). Unlike for phagocytosis, the main effect of radiation was not significant ($P > 0.05$). Neither the GSSG levels nor percentage of oxidized (GSSG) to total glutathione (GSH) levels significantly differed among strains or between radiation and sham-exposed mice within strains (all *P*-values > 0.05). When total GSH levels were normalized to the protein concentration in the spleen lysate, a significant strain effect was observed ($P = 0.001$), but there was no radiation main effect or strain*radiation interaction effect ($P > 0.05$).

The significant interactive effect of strain and radiation on SOD activity demonstrates that genetic background modulates the antioxidant response to low dose radiation exposure. We performed QTL analysis to identify loci associated with this differential response to LDR, using genotype data readily available for the BXD panel (*55*). QTL analysis (Figure A4- 2) revealed a significant QTL on Chromosome (Chr) 15 (@ 74Mb, LOD = 3.54), as well two suggestive QTLs on Chr 16 locus (@ 69Mb, LOD = 2.34 and @ 93Mb, LOD = 2.80), that were linked to SOD activity in unexposed controls. The QTL on distal Chr 16 encompasses the *Sod1* gene. In contrast, QTL analysis of LDR SOD activity identified the same Chr 15 locus as well as a LDR-specific locus on Chr 17 (@

76Mb, LOD = 1.17). Further, the Chr 16 locus containing *Sod1* was not present in the LDR QTL model. A multi-locus regression model which includes additive effects of the Chrs 15 and 17 loci explains 24% of the variance in SOD activity in the spleens of LDR-exposed mice.

The concept of genetic correlation (*70*) was applied to search for relationships between baseline gene expression in unexposed mice and SOD activity following low dose radiation exposure. . These so-called quantitative trait transcripts (QTT) (*71*) could provide insight into mechanisms of variation in LDR SOD activity, particularly if the transcript resides within the QTL interval for LDR SOD activity.

Transcriptomic data from spleens of an overlapping set of BXD strains (*53*) were integrated with SOD activity data, and all possible pair-wise Pearson correlations were computed between expressed transcripts and SOD activity in sham and LDR-exposed BXD strains. Xanthine dehydrogenase (*Xdh*) is located approximately 2Mb upstream of the maxima LOD, and its expression levels are significantly correlated with LDR-induced but not sham control SOD activity (r = -0.34, $P = 0.041$ and r = -0.19, $P > 0.05$, respectively). XDH can be converted to the superoxide-generating enzyme xanthine oxidase (XO) by reversible sulfhydryl oxidation or irreversibly by proteolytic modification (*72*), a process that has been shown to occur in response to high (> 3 Gy) doses of ionizing radiation, potentiating tissue oxidative stress beyond the initial radiochemical reactions (*73*). Outside the QTL interval, expression of *Sod2*, the inducible form of the enzyme, was significantly correlated with LDR SOD activity (r = 0.47, $P = 0.003$) but

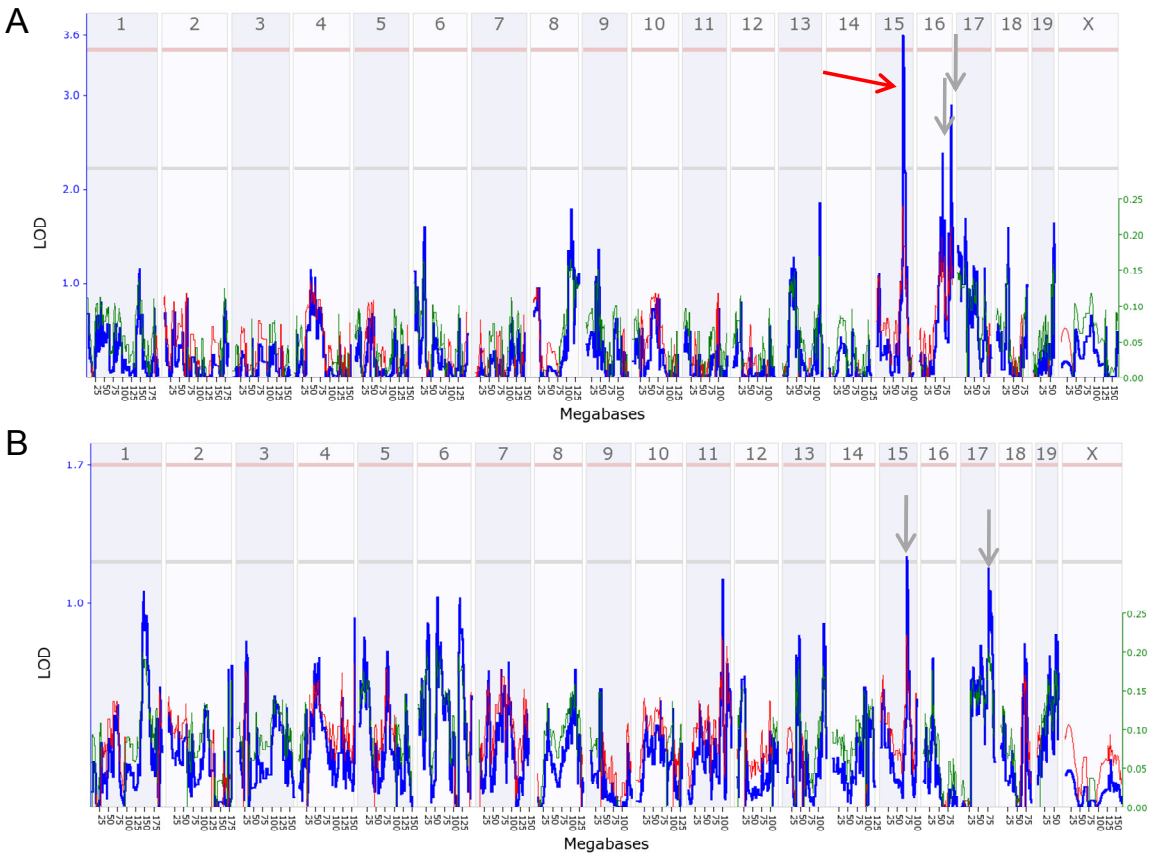



**Figure A4-2. SOD activity QTL analysis.** WebQTL interval mapping of spleen SOD activity in sham (A) and radiation-exposed (B) BXD mice. The mouse genome is portrayed along the horizontal axis, while the vertical axis shows the logarithm of odds (LOD). Significant and suggestive levels of association were determined based on permutation testing and are depicted by horizontal red and gray lines, respectively. Significant and suggestive loci are indicated by red and gray arrows, respectively. LOD scores are indicated by the blue line across the genome; the red line indicates that the C57BL/6J allele at the marker increases the SOD activity, while the green line indicates the DBA/2J allele increases activity. Strength of additive effects is indicated by the scale on the right.

not with sham SOD activity (r = 0.030, $P > 0.05$). In addition to *Sod2*, baseline expression of apoptosis-inducing factor, mitochondrion-associated 2 (*Aifm2*) was differentially correlated with LDR SOD activity (r = 0.55, $P < 0.001$) but not sham SOD activity (r = 0.16, $P > .05$). AIFM2 was originally described as a caspase-independent inducer of apoptosis that translocates from the mitochondria to the nucleus in response to damaging agents and mediates nuclear changes such as chromatin condensation (*74, 75*). Under normal conditions, AIFM2 is thought to act as an intrinsic anti-oxidant enzyme that scavenges free radicals (*76*).

#### SPLEEN GENE EXPRESSION

The parental strains (C57BL/6J and DBA/2J) that contributed alleles to the BXD panel differ in their radiation sensitivity based on classical DNA damage phenotypes (e.g., apoptosis, cell cycle control) (*31, 38-40*). In an effort to more broadly identify differences in the parental strains that also may have segregated in the BXD population, we used microarrays to compare and contrast effects of LDR on expression profiles in the spleen. A higher dose (1 Gy) exposure was included to assess differential effects of genetic background at a radiation dose known to elicit DNA damage and to compare strain differences at two levels of exposure. Low dose radiation significantly altered the expression of 964 genes in either one or both of the strains (q-value < 0.05 and fold change ≥ |1.5|). A total of 138 genes were differentially expressed with LDR exposure in C57Bl/6J but not DBA/2J; the majority of these genes (127 of 138) were down regulated

in LDR mice. Gene ontology enrichment analysis revealed that down regulated genes were significantly enriched in functions related to heme biosynthesis and nucleosome organization (Table A4-1). Considerably more genes (N = 752) were significantly altered in DBA/2J mice and not C57BL/6J 24 h after low dose irradiation, including 511 that were down regulated and 241 up regulated. In DBA/2J, LDR decreased expression of genes in a number of related to immune function, including antigen presentation and processing, B cell receptor signaling, T cell receptor signaling, and cytokine-cytokine receptor interaction (Table A4-1).

Genes significantly up regulated by LDR were enriched in functions related to cell cycle and nucleosome organization including *Chek2*, *Cdc25c*, *Cdkn1*, *Rad51c* and other genes linked to DNA damage repair and cell cycle arrest.

A total of 74 genes were significantly altered by LDR in both strains. Interestingly, all genes in this list exhibited an opposite pattern of response between strains. Close inspection of this list revealed that a number of genes were related to neutrophil function. Q-PCR was used to validate strain-dependent effects of nine neutrophil-related genes, all of which were confirmed to be significantly down regulated in C57BL/6J spleens and up regulated in DBA/2J spleens 24 h following LDR exposure (Table A4-2).A total of 562 genes were significantly different between sham and irradiated mice at the higher (1 Gy) exposure, 307 of which overlapped with the LDR group. Of the HDR genes, 547 were significantly different within C57BL/6J. Genes affected by HDR were highly enriched in functions known to be altered in response to radiation, including cell cycle and the KEGG pathway for p53-mediated transcription. The majority of genes

Table A4-1: Significant GO enrichment of genes differentially regulated in spleen 24 h following low dose radiation exposure.

| GO Biological Process | # of Genes | Benjamini |
|---|---|---|
| **C57BL/6J Down regulated** | | |
| cofactor biosynthetic process | 9 | < 0.001 |
| heme biosynthetic process | 5 | < 0.001 |
| nucleosome organization | 6 | 0.002 |
| **DBA/2J Up regulated** | | |
| cell cycle | 20 | 0.006 |
| nucleosome organization | 7 | 0.040 |
| DNA metabolic process | 14 | 0.039 |
| **DBA/2J Down regulated** | | |
| antigen processing and presentation | 16 | < 0.001 |
| hemopoietic or lymphoid organ development | 23 | < 0.001 |
| Hemopoiesis | 19 | < 0.001 |
| T cell activation | 13 | < 0.001 |
| antigen receptor-mediated signaling pathway | 8 | < 0.001 |
| regulation of apoptosis | 23 | 0.012 |
| B cell activation | 8 | 0.014 |

**Table A4-2: Q-PCR validation of selected genes in which microarray analysis indicated a strain-specific response to low dose radiation in spleen 24 h after exposure to low dose (LDR, 10 cGy) or high dose radiation (HDR, 1 Gy).**

| | | | LDR | | HDR | |
|---|---|---|---|---|---|---|
| | | | C57BL/6J | DBA/2J | C57BL/6J | DBA/2J |
| ***Mpo*** | myeloperoxidase | p-value | < 0.001 | < 0.001 | < 0.001 | > 0.05 |
| | | Fold Change | -27.6 | 9.2 | -16.2 | |
| ***Fcnb*** | ficolin B | p-value | < 0.001 | < 0.001 | < 0.001 | 0.048 |
| | | Fold Change | -17.1 | 8.4 | -14.0 | -2.2 |
| ***Cebpe*** | CCAAT/enhancer binding protein, epsilon | p-value | 0.029 | < 0.001 | < 0.001 | 0.047 |
| | | Fold Change | -2.5 | 6.2 | -4.6 | 2.0 |
| ***Elane*** | elastase, neutrophil expressed | p-value | < 0.001 | < 0.001 | < 0.001 | > 0.05 |
| | | Fold Change | -9.0 | 14.7 | -17.7 | |
| ***Prtn3*** | proteinase 3 | p-value | < 0.001 | < 0.001 | < 0.001 | > 0.05 |
| | | Fold Change | -28.9 | 9.8 | -23.2 | |
| ***Lcn2*** | lipocalin 2 | p-value | < 0.001 | < 0.001 | < 0.001 | > 0.05 |
| | | Fold Change | -15.1 | 12.4 | -10.2 | |
| ***Chi3l1*** | chitinase 3-like 1 | p-value | 0.001 | < 0.001 | < 0.001 | > 0.05 |
| | | Fold Change | -3.3 | 6.6 | -4.7 | |
| ***Camp*** | cathelicidin antimicrobial peptide | p-value | < 0.001 | < 0.001 | < 0.001 | > 0.05 |
| | | Fold Change | -10.2 | 11.9 | -19.5 | |
| ***Ngp*** | neutrophilic granule protein | p-value | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| | | Fold Change | -17.1 | 39.2 | -14.0 | -18.2 |

differentially expressed with LDR in C57BL/6J were also significant at the high dose exposure (198 of 212), and with the same direction of change. Genes involved in heme biosynthesis and neutrophil function were suppressed by both radiation doses in C57BL/6J. In DBA/2J, a total of 102 genes were differentially expressed after HDR, of which 86 were also significantly different (with the same direction of change) with HDR in C57BL/6J. These common genes include up regulation of *Cdkn1a* and down-regulation of *Gadd45a* in both strains. Unlike in C57BL/6J, overlap between LDR and HDR genesets was minimal in DBA/2J (N=8).

## DISCUSSION

Across the BXD population, a single exposure to 10 cGy of radiation significantly enhanced both the numbers of cells that engaged in phagocytosis and the phagocytic activity of those cells. These data suggest that, at least acutely, LDR might increase the ability to respond to invading pathogens. Our findings are consistent with a study of immune function in residents of two villages in Iran, Taleshmahaleh and Chaparsar, who are exposed to background radiation levels 13 times greater than normal due to elevated natural levels of radiation exposure. Residents of these two villages were shown to have increased neutrophil phagocytosis and motility, as well as differences in circulating cytokines such as IL-2, IL-4 and IL-10 (77). It is important to note, however, that both radiation quality and dose rate differ significantly between these two studies. One limitation of our study is that only one time point was analyzed (48 hours post-exposure), which was chosen to fit the overall study design characterizing LDR-effects

that are downstream of the initial radiation stress. This time point may not be optimal for this phenotype; greater enhancement of phagocytosis might be observed at earlier or later time points following irradiation. Because the half-life of murine neutrophils in peripheral blood is approximately 8 hours (*78*), the cells assayed by flow cytometry were irradiated while undergoing maturation in bone marrow. Therefore the mechanism of increased phagocytosis could include maturational effects on cells prior to their release into circulation. Alternatively, LDR may have increased phagocytosis by altering levels of cytokines and chemokines such as TNFα, IL-8, and IFNγ that act on neutrophils in circulation (*13, 23, 79*). Phagocytosis is one of a series of steps that lead to bacterial killing. Follow-up studies are thus necessary to define more broadly the functional effects of LDR on neutrophils.

Microarray data collected from the BXD parental strains further support significant effects of LDR on neutrophils, and in a manner that varied according to strain, based on inverse regulation by LDR of *Mpo* and other genes (Table A4-2). Mature neutrophils were once thought to be transcriptionally inert, but are now known to respond to a number of stressors and cellular signals through changes in gene expression (rev. in (*66*)). We do not have parallel functional data in these two strains, and thus cannot determine if phagocytosis was also differentially impacted by LDR in C57BL/6J and DBA/2J. Current efforts are directed to collecting these data and expanding the scope of LDR-induced neutrophil phenotypes, including chemotaxis and cell killing. We should also point out that the expression data were collected in spleen rather than in isolated neutrophils or in bone marrow. Divergent effects of LDR on gene expression could be

due to indirect effects, such as differential neutrophil migration into spleen or clearance of apoptotic neutrophils. Further experiments will be necessary to test these possibilities.

In contrast to phagocytosis, SOD activity in response to LDR varied significantly between strains. SOD activity increases after radiation exposure to mitigate oxidative stress resulting from the radiolysis of intracellular water, a response that is largely due to increased activity of mitochondrial SOD (SOD2) (*49, 80*). We interpret the significant interaction between strain and treatment to reflect genetic differences in the kinetics of SOD activity and its repletion following radiation stress, as opposed to opposite regulation of the enzyme across strains. Increased SOD activity in spleen occurs within hours of LDR-exposure (*46*), but less is known about the persistence of the response over time. Our data indicate that certain individuals within a population mount a more persistent antioxidant defense to LDR, or that the supply of SOD available is rapidly depleted in some individuals, while others have intrinsically greater response to oxidative LDR stress. The significant positive genetic correlation we observed between *Sod2* expression in spleen of unexposed mice and SOD activity after LDR exposure suggests that heritable differences in *Sod2* expression may contribute to strain differences in the amount or persistence of SOD activity after LDR exposure. *Sod2* deficient mice (C57BL/6J *Sod2*(+/-)), which are more sensitive to radiation than wild type littermates, illustrate the importance of *Sod2* in radiation outcomes (*47*). To further explore this relationship, it would be interesting to determine if BXD strains with the highest heritable levels of *Sod2* expression are less susceptible to LDR responses that

have been shown to be influenced by SOD2 activity, including adaptive radio-resistance and DNA damage (*81*). Similar relationships between baseline gene expression and radiation sensitivity were reported by Amundson et al. using the National Cancer Institute Anticancer Drug Screen (NCI-60) panel of cell lines (*82*).QTL mapping identified a region on Chr 17 associated with LDR-induced but not control SOD activity. To our knowledge, this is the first identification of a QTL for differential responses to LDR (≤10 cGy), although a QTL for radiation-induced thymocytes apoptosis has been reported for a moderate dose (50 cGy) (*83*). The gene encoding *Sod2* is located on Chr 17, but is positioned >60Mb downstream of the maximum LDR SOD activity QTL. Therefore it does not appear that genetic variation within the *Sod2* locus itself, or in proximal regulatory regions, contributes to the variation in SOD activity between irradiated strains. The QTL for SOD activity in control mice, however, encompassed the locus containing *Sod1*, the constitutive form of the enzyme. Using microarray data that we previously collected from BXD spleens (*53*), we identified a gene, xanthine dehydrogenase (*Xdh*), for LDR SOD activity. Because XDH can be converted into xanthine oxidase (XO, the free-radical generating form) by ionizing radiation, the relationship between baseline *Xdh* expression, XDH to XO ratios, SOD activity, and oxidative and peroxidative damage following radiation exposure is likely complicated. Additional experiments are needed to investigate the strong negative correlation between baseline *Xdh* expression and LDR-induced SOD activity.

Comparison of microarray data from low (10 cGy) and high (1 Gy) radiation exposures illustrates that LDR is not simply a more subtle version of HDR, as has been suggested

by other studies (*2-4*). Only 32% (307 of 964 genes) of the genes differentially expressed following LDR-exposure were also changed in response to HDR. Further, the impact of genetic variation on radiation response was more apparent at the lower dose based on strain-dependent patterns of gene expression. At the low dose, 1200 genes showed a significant strain-dependent pattern of expression (based on the radiation*strain interaction term in the ANOVA model), while only five genes met this criterion at HDR (data not shown). These differential effects across a rather modest (10-fold) increase in dose suggest that the response to radiation between individuals qualitatively becomes more similar at higher doses, perhaps as the demand to abate DNA damage overrides other, less critical consequences. As dose is lowered to a level that imposes oxidative stress rather than immediate damage, genetic variation in mechanisms for coping with stress and the consequential effects on other pathways begin to emerge. Taken together, these array data further confirm the biological uniqueness of low dose vs. high dose responses. They also highlight the need to consider genetic variation when assessing LDR outcomes, perhaps even more so than for higher radiation doses.

Differences in radiation sensitivity between the BXD parental strains were first described by Roderick more than 45 years ago, with DBA/2J succumbing more quickly than C57BL/6J to a lethal dose of radiation (*26*). At more modest doses, C57BL/6J mice were shown to be more resistant to radiation-induced genomic instability than DBA/2J (*38, 84, 85*). Wright and colleagues described differential apoptotic responses between the two strains after 1 Gy radiation, with C57BL/6J favoring apoptosis through rapid

induction of p53 and up regulation of pro-apoptotic *Bax*, and DBA/2J having a delayed but prolonged p53 activation with more emphasis on p21 activation and cell cycle arrest (*31, 39, 40*). Microarray data collected from spleen further illustrates that genetic variation plays a major role in how these two strains respond to radiation, particularly at low doses and for genes related to immune function. One intriguing question prompted by these collective results is if there is a mechanistic link between differences in the initial response to radiation (promotion of apoptosis in C57BL/6J and cell cycle arrest in DBA/2J) and the later changes in expression of immune-related genes. The p53 signaling cascade regulates a number of transcriptional targets that further impact downstream pathways and cellular responses. A recent report by Tavana et al. suggests that, in response to cellular stress, preferential activation of p53-mediated apoptosis or cell cycle arrest alters later phenotypes and differentially impacts inflammation and immune activation (*86*). These authors used a p53 mutant mouse model that could not transactivate pro-apoptotic genes but retained the ability to up regulate *Cdkn1a* in response to DNA damage. When exposed to UV radiation, the mutant mice were more susceptible to inflammation and immunosuppression than wild type controls. If causative links can be made from the initial differences in response to radiation stress (as described by Wright and others) and later changes in cellular pathways (as our array data suggest for the immune system), it will enhance our understanding of mechanisms of heritable radiation sensitivity. Further, if the response phenotypes (apoptosis vs. cell cycle arrest) of these two “individuals” are more generalizable to subsets of the human population, understanding radiation sensitivity in

these two strains and in the BXD progeny should inform parallel investigations in humans.

The results we have presented represent our initial efforts to utilize genetic reference populations and a systems genetics approach to understand the basis for and consequences of heritable differences in sensitivity to low dose radiation exposure. Genetic reference populations like the BXD panel and a systems genetics framework are ideal for assembling linkages between genetic variants, intermediate phenotypes and outcomes of environmental exposures (*87*). Using a genetically stable population model allows us to integrate data from multiple studies as if they were taken from the same animals. This is valuable because it enables complex connections between molecular, cellular and higher order radiation responses to be assembled across time, using tools of systems genetics. Systems genetics is founded on the concept of genetic correlation among traits and provides a framework for extracting interrelationships between phenotypes that might not otherwise be suspected. As an example, we identified a significant genetic correlation between phagocytosis in irradiated mice from this study and peripheral CD4+:CD8+ ratio from a previous study (data not shown; (*53*)). Our current efforts are focused on genetic sensitivity to low dose radiation; more broadly, the same framework could be used to relate radiation sensitivity to susceptibility to other environmental exposures or to disease. At the translational level, systems genetics can potentially improve the assessment of risk through identification of phenotypes, whether molecular, cellular or biochemical, that signal differential sensitivity to radiation and other environmental challenges of concern. Thus far, we use

the BXD strain panel for studies of radiation sensitivity because it is the largest set of inbred mouse RI strains. As with typical F2 populations that are often used for QTL mapping, one limitation of the BXD panel is that the input allelic diversity is limited to those found in two strains of mice. The Collaborative Cross, an RI panel being produced from eight, rather than two, parental lines will provide a model much more representative of a population with respect to genetic diversity (*88, 89*). The CC will both expand the range of heritable sensitivity to LDR and the ability to map radiation sensitivity loci with much greater resolution and precision (*90*). In summary, our results demonstrate responses to low dose exposure that are robust to genetic variation (enhanced neutrophil phagocytosis) as well as other responses for which genetic background significantly impacts the response (SOD activity). These data provide groundwork for expanding our approach to radiation sensitivity using systems genetics.

# REFERENCES TO APPENDIX

*1*. NCRP, National Council on Radiation Protection and Measurements: Evaluation of the Linear-Nonthreshold Dose-Response Model for Ionizing Radiation, Report No. 136., Bethesda, MD, 2001.

*2*. M. Tubiana, The 2007 Marie Curie prize: the linear no threshold relationship and advances in our understanding of carcinogenesis. *International Journal of Low Dose Radiation* **5**, 173-204 (2008).

*3*. T. K. Day, G. Zeng, A. M. Hooker, M. Bhat, D. R. Turner and P. J. Sykes, Extremely low doses of x-radiation can induce adaptive responses in mouse prostate. *Dose Response* **5**, 315-322 (2007).

*4*. Y. Ina and K. Sakai, Activation of immunological network by chronic low-dose-rate irradiation in wild-type mouse strains: analysis of immune cell populations and surface molecules. *International journal of radiation biology* **81**, 721-729 (2005).

*5*. M. Yonezawa, A. Takeda and J. Misonoh, Acquired radioresistance after low dose X-irradiation in mice. *J Radiat Res (Tokyo)* **31**, 256-262 (1990).

*6*. R. E. Goans, E. C. Holloway, M. E. Berger and R. C. Ricks, Early dose assessment following severe radiation accidents. *Health Phys* **72**, 513-518 (1997).

*7*. S. Z. Liu, On radiation hormesis expressed in the immune system. *Crit Rev Toxicol* **33**, 431-441 (2003).

*8*. M. Nogami, J. T. Huang, L. T. Nakamura and T. Makinodan, T cells are the cellular target of the proliferation-augmenting effect of chronic low-dose ionizing radiation in mice. *Radiat Res* **139**, 47-52 (1994).

*9*. R. E. Anderson and G. M. Troup, Effects of irradiation upon the response of murine spleen cells to mitogens. *Am J Pathol* **109**, 169-178 (1982).

*10*. K. Ishii, K. Yamaoka, Y. Hosoi, T. Ono and K. Sakamoto, Enhanced mitogen-induced proliferation of rat splenocytes by low-dose whole-body X-irradiation. *Physiol Chem Phys Med NMR* **27**, 17-23 (1995).

*11*. R. Pandey, B. S. Shankar, D. Sharma and K. B. Sainis, Low dose radiation induced immunomodulation: effect on macrophages and CD8+ T cells. *International journal of radiation biology* **81**, 801-812 (2005).

*12*. B. Shankar, S. Premachandran, S. D. Bharambe, P. Sundaresan and K. B. Sainis, Modification of immune response by low dose ionizing radiation: role of apoptosis. *Immunol Lett* **68**, 237-245 (1999).

*13*. Y. Ibuki and R. Goto, Contribution of inflammatory cytokine release to activation of resident peritoneal macrophages after in vivo low-dose gamma-irradiation. *J Radiat Res (Tokyo)* **40**, 253-262 (1999).

*14*. E. M. Nowosielska, J. Wrembel-Wargocka, A. Cheda, E. Lisiak and M. K. Janiak, Enhanced cytotoxic activity of macrophages and suppressed tumor metastases in mice irradiated with low doses of X- rays. *J Radiat Res (Tokyo)* **47**, 229-236 (2006).

*15*. X. D. Liu, S. M. Ma and S. Z. Liu, Effects of 0.075 Gy x-ray irradiation on the expression of IL-10 and IL-12 in mice. *Phys Med Biol* **48**, 2041-2049 (2003).

*16*. Y. X. Shan, S. Z. Jin, X. D. Liu, Y. Liu and S. Z. Liu, Ionizing radiation stimulates secretion of pro-inflammatory cytokines: dose-response relationship, mechanisms and implications. *Radiat Environ Biophys* **46**, 21-29 (2007).

*17*. Z. A. Cao, D. Daniel and D. Hanahan, Sub-lethal radiation enhances anti-tumor immunotherapy in a transgenic mouse model of pancreatic cancer. *BMC Cancer* **2**, 11 (2002).

*18*. S. Hashimoto, H. Shirato, M. Hosokawa, T. Nishioka, Y. Kuramitsu, K. Matushita, M. Kobayashi and K. Miyasaka, The suppression of metastases and the change in host immune response after low-dose total-body irradiation in tumor-bearing rats. *Radiat Res* **151**, 717-724 (1999).

*19*. Y. Hosoi and K. Sakamoto, Suppressive effect of low dose total body irradiation on lung metastasis: dose dependency and effective period. *Radiother Oncol* **26**, 177-179 (1993).

*20*. S. Kojima, K. Nakayama and H. Ishida, Low dose gamma-rays activate immune functions via induction of glutathione and delay tumor growth. *J Radiat Res (Tokyo)* **45**, 33-39 (2004).

*21*. D. S. Gridley, M. J. Pecaut, A. Rizvi, G. B. Coutrakon, X. Luo-Owen, A. Y. Makinde and J. M. Slater, Low-dose, low-dose-rate proton radiation modulates CD4(+) T cell gene expression. *International journal of radiation biology* **85**, 250-261 (2009).

*22*. S. Z. Liu, S. Z. Jin, X. D. Liu and Y. M. Sun, Role of CD28/B7 costimulation and IL-12/IL-10 interaction in the radiation-induced immune changes. *BMC Immunol* **2**, 8 (2001).

*23*. D. S. Gridley, M. J. Pecaut, G. M. Miller, M. F. Moyers and G. A. Nelson, Dose and dose rate effects of whole-body gamma-irradiation: II. Hematological variables and cytokines. *In Vivo* **15**, 209-216 (2001).

*24*. S. Z. Liu, SuXu, Y. C. Zhang and Y. Zhao, Signal transduction in lymphocytes after low dose radiation. *Chin Med J (Engl)* **107**, 431-436 (1994).

*25*. D. S. Gridley, A. Rizvi, X. Luo-Owen, A. Y. Makinde and M. J. Pecaut, Low dose, low dose rate photon radiation modifies leukocyte distribution and gene expression in CD4(+) T cells. *J Radiat Res (Tokyo)* **50**, 139-150 (2009).

*26*. T. H. Roderick, The Response Of Twenty-Seven Inbred Strains Of Mice To Daily Doses Of Whole-Body X-Irradiation. *Radiat Res* **20**, 631-639 (1963).

*27*. N. Mori, M. Okumoto, J. Morimoto, S. Imai, T. Matsuyama, Y. Takamori and O. Yagasaki, Genetic analysis of susceptibility to radiation-induced apoptosis of thymocytes in mice. *International journal of radiation biology* **62**, 153-159 (1992).

*28*. N. Mori, M. Okumoto, A. A. Hart and P. Demant, Apoptosis susceptibility genes on mouse chromosome 9 (Rapop2) and chromosome 3 (Rapop3). *Genomics* **30**, 553-557 (1995).

*29*. N. Mori, M. Okumoto, M. A. van der Valk, S. Imai, S. Haga, K. Esaki, A. A. Hart and P. Demant, Genetic dissection of susceptibility to radiation-induced apoptosis of thymocytes and mapping of Rapop1, a novel susceptibility gene. *Genomics* **25**, 609-614 (1995).

*30*. N. Mori, T. van Wezel, M. van der Valk, J. Yamate, S. Sakuma, M. Okumoto and P. Demant, Genetics of susceptibility to radiation-induced apoptosis in colon: two loci on chromosomes 9 and 16. *Mamm Genome* **9**, 377-380 (1998).

*31*. M. Wallace, P. J. Coates, E. G. Wright and K. L. Ball, Differential post-translational modification of the tumour suppressor proteins Rb and p53 modulate the rates of radiation-induced apoptosis in vivo. *Oncogene* **20**, 3597-3608 (2001).

*32*. N. Mori, Y. Matsumoto, M. Okumoto, N. Suzuki and J. Yamate, Variations in Prkdc encoding the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) and susceptibility to radiation-induced apoptosis and lymphomagenesis. *Oncogene* **20**, 3609-3619 (2001).

*33*. I. R. Major and R. H. Mole, Myeloid leukaemia in x-ray irradiated CBA mice. *Nature* **272**, 455-456 (1978).

*34*. R. L. Ullrich and R. J. Preston, Radiation induced mammary cancer. *J Radiat Res (Tokyo)* **32 Suppl 2**, 104-109 (1991).

*35*. Y. Yu, R. Okayasu, M. M. Weil, A. Silver, M. McCarthy, R. Zabriskie, S. Long, R. Cox and R. L. Ullrich, Elevated breast cancer risk in irradiated BALB/c mice associates with unique functional polymorphism of the Prkdc (DNA-dependent protein kinase catalytic subunit) gene. *Cancer Res* **61**, 1820-1824 (2001).

*36*. T. Kataoka, Y. Mizuguchi, K. Notohara, T. Taguchi and K. Yamaoka, Histological changes in spleens of radio-sensitive and radio-resistant mice exposed to low-dose X-ray irradiation. *Physiol Chem Phys Med NMR* **38**, 21-29 (2006).

*37*. P. F. Wilson, P. B. Nham, S. S. Urbin, J. M. Hinz, I. M. Jones and L. H. Thompson, Inter-individual variation in DNA double-strand break repair in human

fibroblasts before and after exposure to low doses of ionizing radiation. *Mutat Res* **683**, 91-97 (2010).

*38*. G. E. Watson, S. A. Lorimore, S. M. Clutton, M. A. Kadhim and E. G. Wright, Genetic factors influencing alpha-particle-induced chromosomal instability. *International journal of radiation biology* **71**, 497-503 (1997).

*39*. K. J. Lindsay, P. J. Coates, S. A. Lorimore and E. G. Wright, The genetic basis of tissue responses to ionizing radiation. *Br J Radiol* **80 Spec No 1**, S2-6 (2007).

*40*. P. J. Coates, S. A. Lorimore, K. J. Lindsay and E. G. Wright, Tissue-specific p53 responses to ionizing radiation and their genetic modification: the key to tissue-specific tumour susceptibility? *The Journal of pathology* **201**, 377-388 (2003).

*41*. B. H. Voy, Systems genetics: a powerful approach for gene-environment interactions. *J Nutr* **141**, 515-519 (2011).

*42*. D. W. Threadgill, Meeting report for the 4th annual Complex Trait Consortium meeting: from QTLs to systems genetics. *Mamm Genome* **17**, 2-4 (2006).

*43*. T. F. Mackay, E. A. Stone and J. F. Ayroles, The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-577 (2009).

*44*. B. H. Voy, J. A. Scharff, A. D. Perkins, A. M. Saxton, B. Borate, E. J. Chesler, L. K. Branstetter and M. A. Langston, Extracting gene networks for low-dose radiation using graph theoretical algorithms. *PLoS Comput Biol* **2**, e89 (2006).

*45*. D. R. Spitz, E. I. Azzam, J. J. Li and D. Gius, Metabolic oxidation/reduction reactions and cellular responses to ionizing radiation: a unifying concept in stress response biology. *Cancer Metastasis Rev* **23**, 311-322 (2004).

*46*. K. Yamaoka, S. Kojima, M. Takahashi, T. Nomura and K. Iriyama, Change of glutathione peroxidase synthesis along with that of superoxide dismutase synthesis in mice spleens after low-dose X-ray irradiation. *Biochim Biophys Acta* **1381**, 265-270 (1998).

*47*. M. W. Epperly, C. J. Epstein, E. L. Travis and J. S. Greenberger, Decreased pulmonary radiation resistance of manganese superoxide dismutase (MnSOD)-deficient mice is corrected by human manganese superoxide dismutase-Plasmid/Liposome (SOD2-PL) intratracheal gene therapy. *Radiat Res* **154**, 365-374 (2000).

*48*. M. Fan, K. M. Ahmed, M. C. Coleman, D. R. Spitz and J. J. Li, Nuclear factor-kappaB and manganese superoxide dismutase mediate adaptive radioresistance in low-dose irradiated mouse skin epithelial cells. *Cancer Res* **67**, 3220-3228 (2007).

*49*. G. Guo, Y. Yan-Sanders, B. D. Lyn-Cook, T. Wang, D. Tamae, J. Ogi, A. Khaletskiy, Z. Li, C. Weydert, et al., Manganese superoxide dismutase-mediated gene

expression in radiation-induced adaptive responses. *Mol Cell Biol* **23**, 2362-2378 (2003).

*50*. B. A. Taylor, C. Wnek, B. S. Kotlus, N. Roemer, T. MacTaggart and S. J. Phillips, Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. *Mamm Genome* **10**, 335-348 (1999).

*51*. B. A. Taylor, H. G. Bedigian and H. Meier, Genetic studies of the Fv-1 locus of mice: linkage with Gpd-1 in recombinant inbred lines. *J Virol* **23**, 106-109 (1977).

*52*. J. L. Peirce, L. Lu, J. Gu, L. M. Silver and R. W. Williams, A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet* **5**, 7 (2004).

*53*. R. M. Lynch, S. Naswa, G. L. Rogers, Jr., S. A. Kania, S. Das, E. J. Chesler, A. M. Saxton, M. A. Langston and B. H. Voy, Identifying genetic loci and spleen gene coexpression networks underlying immunophenotypes in BXD recombinant inbred mice. *Physiol Genomics* (2010).

*54*. C. J. Leblanc, A. K. Leblanc, M. M. Jones, J. W. Bartges and S. A. Kania, Evaluation of peripheral blood neutrophil function in tumor-bearing dogs. *Vet Clin Pathol* (2010).

*55*. S. Shifman, J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams, R. Mott and J. Flint, A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol* **4**, e395 (2006).

*56*. E. J. Chesler, L. Lu, J. Wang, R. W. Williams and K. F. Manly, WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nat Neurosci* **7**, 485-486 (2004).

*57*. G. A. Churchill and R. W. Doerge, Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971 (1994).

*58*. E. Lander and L. Kruglyak, Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241-247 (1995).

*59*. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, et al., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).

*60*. P. Du, W. A. Kibbe and S. M. Lin, lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-1548 (2008).

*61*. J. D. Storey and R. Tibshirani, Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445 (2003).

*62*. W. Huang da, B. T. Sherman and R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).

*63*. G. Dennis, Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki, DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).

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

*65*. P. Delves, S. Martin, D. Burton and I. Roitt, *Roitt's Essential Immunology*. Wiley-Blackwell, Hoboken, 2006.

*66*. S. Kobayashi and F. DeLeo, Role of neutrophils in innate immunity: a systems biology-level approach. *WIREs Systems Biology and Medicine* **1**, 309-333 (2009).

*67*. U. Bilitewski, Determination of immunomodulatory effects: focus on functional analysis of phagocytes as representatives of the innate immune system. *Anal Bioanal Chem* **391**, 1545-1554 (2008).

*68*. C. M. Pathak, P. K. Avti, S. Kumar, K. L. Khanduja and S. C. Sharma, Whole body exposure to low-dose gamma radiation promotes kidney antioxidant status in Balb/c mice. *J Radiat Res (Tokyo)* **48**, 113-120 (2007).

*69*. L. Miao and D. K. St Clair, Regulation of superoxide dismutase genes: implications in disease. *Free Radic Biol Med* **47**, 344-356 (2009).

*70*. V. M. Philip, S. Duvvuru, B. Gomero, T. A. Ansah, C. D. Blaha, M. N. Cook, K. M. Hamre, W. R. Lariviere, D. B. Matthews, et al., High-throughput behavioral phenotyping in the expanded panel of BXD recombinant inbred strains. *Genes Brain Behav* (2009).

*71*. G. Passador-Gurgel, W. P. Hsieh, P. Hunt, N. Deighton and G. Gibson, Quantitative trait transcripts for nicotine resistance in Drosophila melanogaster. *Nat Genet* **39**, 264-268 (2007).

*72*. M. Srivastava and R. K. Kale, Effect of radiation on the xanthine oxidoreductase system in the liver of mice. *Radiat Res* **152**, 257-264 (1999).

*73*. M. Srivastava, D. Chandra and R. K. Kale, Modulation of radiation-induced changes in the xanthine oxidoreductase system in the livers of mice by its inhibitors. *Radiat Res* **157**, 290-297 (2002).

*74*. M. Wu, L. G. Xu, T. Su, Y. Tian, Z. Zhai and H. B. Shu, AMID is a p53-inducible gene downregulated in tumors. *Oncogene* **23**, 6815-6819 (2004).

*75*. K. R. Marshall, M. Gong, L. Wodke, J. H. Lamb, D. J. Jones, P. B. Farmer, N. S. Scrutton and A. W. Munro, The human apoptosis-inducing protein AMID is an oxidoreductase with a modified flavin cofactor and DNA binding activity. *J Biol Chem* **280**, 30735-30740 (2005).

*76*. N. Joza, J. A. Pospisilik, E. Hangen, T. Hanada, N. Modjtahedi, J. M. Penninger and G. Kroemer, AIF: not just an apoptosis-inducing factor. *Ann N Y Acad Sci* **1171**, 2-11 (2009).

*77*. M. Attar, Y. Molaie Kondolousy and N. Khansari, Effect of high dose natural ionizing radiation on the immune system of the exposed residents of Ramsar Town, Iran. *Iran J Allergy Asthma Immunol* **6**, 73-78 (2007).

*78*. B. I. Lord, G. Molineux, Z. Pojda, L. M. Souza, J. J. Mermod and T. M. Dexter, Myeloid cell kinetics in mice treated with recombinant interleukin-3, granulocyte colony-stimulating factor (CSF), or granulocyte-macrophage CSF in vivo. *Blood* **77**, 2154-2159 (1991).

*79*. M. Galdiero, G. Cipollaro de l'Ero, A. Folgore, M. Cappello, A. Giobbe and F. S. Sasso, Effects of irradiation doses on alterations in cytokine release by monocytes and lymphocytes. *J Med* **25**, 23-40 (1994).

*80*. K. Otsuka, T. Koana, H. Tauchi and K. Sakai, Activation of antioxidative enzymes induced by low-dose-rate whole-body gamma irradiation: adaptive response in terms of initial DNA damage. *Radiat Res* **166**, 474-478 (2006).

*81*. S. M. de Toledo, N. Asaad, P. Venkatachalam, L. Li, R. W. Howell, D. R. Spitz and E. I. Azzam, Adaptive responses to low-dose/low-dose-rate gamma rays in normal human fibroblasts: the role of growth architecture and oxidative metabolism. *Radiat Res* **166**, 849-857 (2006).

*82*. S. A. Amundson, K. T. Do, L. C. Vinikoor, R. A. Lee, C. A. Koch-Paiz, J. Ahn, M. Reimers, Y. Chen, D. A. Scudiero, et al., Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer Res* **68**, 415-424 (2008).

*83*. M. M. Weil, X. Xia, Y. Lin, L. C. Stephens and C. I. Amos, Identification of quantitative trait loci controlling levels of radiation-induced thymocyte apoptosis in mice. *Genomics* **45**, 626-628 (1997).

*84*. T. Nomura, M. Kinuta, T. Hongyo, H. Nakajima and T. Hatanaka, Programmed cell death in whole body and organ systems by low dose radiation. *J Radiat Res (Tokyo)* **33 Suppl**, 109-123 (1992).

*85*. E. G. Wright, Radiation-induced genomic instability in haemopoietic cells. *International journal of radiation biology* **74**, 681-687 (1998).

*86*. O. Tavana, C. L. Benjamin, N. Puebla-Osorio, M. Sang, S. E. Ullrich, H. N. Ananthaswamy and C. Zhu, Absence of p53-dependent apoptosis leads to UV radiation hypersensitivity, enhanced immunosuppression and cellular senescence. *Cell Cycle* **9**, 3328-3336.

*87*. J. F. Ayroles, M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollmann, L. H. Duncan, F. Lawrence, et al., Systems genetics of complex traits in Drosophila melanogaster. *Nat Genet* **41**, 299-307 (2009).

*88*. G. A. Churchill, D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, J. Beatty, W. D. Beavis, J. K. Belknap, B. Bennett, et al., The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* **36**, 1133-1137 (2004).

*89*. E. J. Chesler, D. R. Miller, L. R. Branstetter, L. D. Galloway, B. L. Jackson, V. M. Philip, B. H. Voy, C. T. Culiat, D. W. Threadgill, et al., The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm Genome* **19**, 382-389 (2008).

*90*. D. L. Aylor, W. Valdar, W. Foulds-Mathes, R. J. Buus, R. A. Verdugo, R. S. Baric, M. T. Ferris, J. A. Frelinger, M. Heise, et al., Genetic analysis of complex traits in the emerging collaborative cross. *Genome Res* **21**, 000-000 (2011).

Acknowledgements

We would like to thank KT Cain for breeding the BXD mice at ORNL, Ginger D. Shaw and Marc Garland for mouse radiation exposures, Darla Miller for project coordination, and Dianne J. Trent for flow cytometry.

Grants

This project was supported by the Low Dose Radiation Research and EPSCoR Laboratory Partnership Programs of the Office of Biological and Environmental Research, Office of Science, Department of Energy ERKP650 and ERKP804.

# CHAPTER 5 : CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation illustrates the influence of LDIR on gene networks. It has been established that in comparison to differential expression methods such as ANOVA differential correlation is more suitable for detecting the effects of relatively mild stimuli like LDIR especially in genetically diverse population. The two stage statistical filter based differential correlation method used here revealed gene networks highly enriched with radiation sensitive genes. GO enrichment of genes from the differentially correlated network revealed influence of LDIR on many biological processes relevant to radiation responses such as apoptosis, DNA damage, oxidative stress, signal transduction, immune system and cell cycle. Differential correlation also detected perturbations in putative networks involving cancer genes. LDIR affected differential correlation among a network of interconnected hubs enriched in well known BRCA complex genes (BRCC3, BRCC45 and RNF168). Exposure to LDIR also influenced genes belonging to MAP Kinase Pathway (another cancer pathway) and hematopoietic system. In contrast to differential correlation method the differential expression method (ANOVA) could not detect these LDIR induced changes. The inability of ANOVA can be explained by presence of inter-strain genetic variability and mild nature of treatment (LDIR). Hence biological validation of LDIR induced differentially co-expressed candidate genes of interest should be done using sensitive biological assays such as quantitative PCR. Moreover, comparison of irradiated and sham irradiated cells in a controlled environment using spleen cell lines may reveal differential expression of genes because of low variability of data. LDIR induced changes in correlation of genes could also be validated by complete and partial knockouts of interesting candidate

genes using RNA interference experiments. RNA interference experiments could also reveal the causality of relationship among differentially correlated genes.

The strict two stage statistical filter used here produced a sparse graph of differentially correlated vertices (genes). Dense portions of this sparse graph such as connected components, interconnected hubs etc. can provide a small list of candidate genes to determine the directionality of network for control as well as irradiated data using methods like structure equation modeling and Bayesian statistics. The information from RNA interference experiments and existing literature can be used to test the validity of directions in the proposed models.

The gene networks and pathways influenced by the LDIR raise questions about net effect of these changes on the well being of an individual. Are these responses specific to LDIR? Do these changes represent deleterious effects of radiation at a lower scale? Are they adaptive responses to radiation? Gene networks influenced by low dose and high dose of radiation can be compared to specify those influenced by LDIR alone. The comparison may also enable us to predict gene networks participating in hormesis and adaptive responses actuated by LDIR.

Though differential correlation could identify the influence of LDIR on gene networks in this study, there is limitation to this method. It would be hard to use this method if the number of observations (microarrays) is too low as in many case control studies. In such cases differential expression would be more appropriate if the treatment is not too mild. In case of LDIR transcription profiling of mice exposed to more than one dose of

radiation may make its effects more pronounced that could be detected by differential expression methods. Such a study would address concerns regarding increased and repetitive usage of LDIR for medical procedures.

Another limitation of the differential co-expression method used here is its ability to detect only differences in linear correlations between gene pairs. However biological networks including gene networks can be nonlinear. LDIR induced differences in co-expression should further be detected by using nonlinear methods like differences in mutual information between genes. Mutual information based method should result in denser differential co-expression networks since it will consider both linear and nonlinear relations between genes.

In this dissertation, we have also established that baseline expression profiles obtained from a reference population of genetically different strains of inbred BXD mice can be used for extraction of gene networks at a lower threshold of correlation than the conventionally used high threshold. Graphs filtered with a statistically significant but low Pearson's correlation threshold ($r = 0.413$) resulted in paracliques significantly enriched in genes involved in various biological processes. Pearson correlation among the genes belonging to various KEGG pathways and regulatory networks also suggested a low threshold. The enrichment of biologically related genes in the gene networks obtained at low correlation thresholds can be explained by the absence of a stimulus driving these genes to higher correlations coupled with higher sensitivity of detection resulting from higher sample size in system genetics studies. Here we employed only linear model (Pearson correlation) to explain the relationships between genes. Networks based on

Pearson correlation should therefore be compared with networks generated by modeling nonlinear relations among genes. Methods like mutual information or polynomial regression can be used for detection of nonlinear relationships among genes. The comparisons may help in filling some of the missing links in gene networks and may also reveal that many of the low magnitude linear relationships (low Pearson correlation) are better explained by nonlinear models.

# REFERENCES

1. IAEA. *Radiation, People and the Environment*. 2004; Available from: http://www.iaea.org/Publications/Booklets/RadPeopleEnv/radiation_booklet.html.

2. Cuttler, J.M. and M. Pollycove, *Nuclear energy and health: and the benefits of low-dose radiation hormesis.* Dose Response, 2009. 7(1): p. 52-89.

3. UNSCEAR, *United Nations. Scientific Committee on the Effects of Atomic Radiation: Sources and effects of ionizing radiation Radiation:Volume I: Sources - Report to the General Assembly Scientific Annexes A and B.* United Nations, New York, 2008. 1.

4. Sanders, C.L.L., *Radiation Hormesis and the Linear-No-Threshold Assumption*2010: Springer Verlag Berlin Heidelberg.

5. Nair, M.K., et al., *Population study in the high natural background radiation area in Kerala, India.* Radiat Res, 1999. 152(6 Suppl): p. S145-8.

6. NRC, *Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation, National Academies Press.*, 2006.

7. CDRH, *Initiative to Reduce Unnecessary Radiation Exposure from Medical Imaging: White paper.* Center for Devices and Radiological Health,U.S. Food and Drug Administration, 2010(February).

8. Spitz, D.R., et al., *Metabolic oxidation/reduction reactions and cellular responses to ionizing radiation: a unifying concept in stress response biology.* Cancer Metastasis Rev, 2004. 23(3-4): p. 311-22.

9. Kamiya, K. and M. Sasatani, *[Effects of radiation exposure on human body].* Nihon Rinsho, 2012. 70(3): p. 367-74.

10. Sigurdson, A.J., et al., *Routine diagnostic X-ray examinations and increased frequency of chromosome translocations among U.S. radiologic technologists.* Cancer Res, 2008. 68(21): p. 8825-31.

11. Preston, D.L., et al., *Radiation effects on breast cancer risk: a pooled analysis of eight cohorts.* Radiat Res, 2002. 158(2): p. 220-35.

12. Little, M.P., et al., *Risks associated with low doses and low dose rates of ionizing radiation: why linearity may be (almost) the best we can do.* Radiology, 2009. 251(1): p. 6-12.

13. Olivieri, G., J. Bodycote, and S. Wolff, *Adaptive response of human lymphocytes to low concentrations of radioactive thymidine.* Science, 1984. 223(4636): p. 594-7.

14. Boothman, D.A., et al., *Altered G1 checkpoint control determines adaptive survival responses to ionizing radiation.* Mutat Res, 1996. 358(2): p. 143-53.

15. Wolff, S., et al., *Adaptive response of human lymphocytes for the repair of radon-induced chromosomal damage.* Mutat Res, 1991. 250(1-2): p. 299-306.

16. Prise, K.M., *New advances in radiation biology.* Occup Med (Lond), 2006. 56(3): p. 156-61.

17. Matsumoto, H., et al., *A new paradigm in radioadaptive response developing from microbeam research.* J Radiat Res (Tokyo), 2009. 50 Suppl A: p. A67-79.

18. Bravard, A., et al., *Contribution of antioxidant enzymes to the adaptive response to ionizing radiation of human lymphoblasts.* Int J Radiat Biol, 1999. 75(5): p. 639-45.

19. Murley, J.S., et al., *SOD2-mediated effects induced by WR1065 and low-dose ionizing radiation on micronucleus formation in RKO human colon carcinoma cells.* Radiat Res, 2011. 175(1): p. 57-65.

20. Azzam, E.I., et al., *Low-dose ionizing radiation decreases the frequency of neoplastic transformation to a level below the spontaneous rate in C3H 10T1/2 cells.* Radiat Res, 1996. 146(4): p. 369-73.

21. Cohen, B.L., *Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon decay products.* Health Phys, 1995. 68(2): p. 157-74.

22. Liang, X., et al., *The low-dose ionizing radiation stimulates cell proliferation via activation of the MAPK/ERK pathway in rat cultured mesenchymal stem cells.* J Radiat Res (Tokyo), 2011. 52(3): p. 380-6.

23. Shin, S.C., et al., *Alteration of cytokine profiles in mice exposed to chronic low-dose ionizing radiation.* Biochem Biophys Res Commun, 2010. 397(4): p. 644-9.

24. Shankar, B., et al., *Modification of immune response by low dose ionizing radiation: role of apoptosis.* Immunol Lett, 1999. 68(2-3): p. 237-45.

25. Ghiassi-Nejad, M., et al., *Long-term immune and cytogenetic effects of high level natural radiation on Ramsar inhabitants in Iran.* J Environ Radioact, 2004. 74(1-3): p. 107-16.

26. Mohammadi, S., et al., *Adaptive response of blood lymphocytes of inhabitants residing in high background radiation areas of ramsar- micronuclei, apoptosis and comet assays.* J Radiat Res (Tokyo), 2006. 47(3-4): p. 279-85.

27. Maffei, F., et al., *Spectrum of chromosomal aberrations in peripheral lymphocytes of hospital workers occupationally exposed to low doses of ionizing radiation.* Mutat Res, 2004. 547(1-2): p. 91-9.

28. Tucker, J.D., *Low-dose ionizing radiation and chromosome translocations: a review of the major considerations for human biological dosimetry.* Mutat Res, 2008. 659(3): p. 211-20.

29. Krueger, S.A., et al., *Role of apoptosis in low-dose hyper-radiosensitivity.* Radiat Res, 2007. 167(3): p. 260-7.

30. Marples, B., et al., *Low-dose hyper-radiosensitivity: a consequence of ineffective cell cycle arrest of radiation-damaged G2-phase cells.* Radiat Res, 2004. 161(3): p. 247-55.

31. Nagasawa, H. and J.B. Little, *Induction of sister chromatid exchanges by extremely low doses of alpha-particles.* Cancer Res, 1992. 52(22): p. 6394-6.

32. Azzam, E.I., et al., *Oxidative metabolism modulates signal transduction and micronucleus formation in bystander cells from alpha-particle-irradiated normal human fibroblast cultures.* Cancer Res, 2002. 62(19): p. 5436-42.

33. Nagasawa, H. and J.B. Little, *Unexpected sensitivity to the induction of mutations by very low doses of alpha-particle radiation: evidence for a bystander effect.* Radiat Res, 1999. 152(5): p. 552-7.

34. Zhou, H., et al., *Induction of a bystander mutagenic effect of alpha particles in mammalian cells.* Proc Natl Acad Sci U S A, 2000. 97(5): p. 2099-104.

35. Lewis, D.A., et al., *Production of delayed death and neoplastic transformation in CGL1 cells by radiation-induced bystander effects.* Radiat Res, 2001. 156(3): p. 251-8.

36. Zhou, H., et al., *Effects of irradiated medium with or without cells on bystander cell responses.* Mutat Res, 2002. 499(2): p. 135-41.

37. Ponnaiya, B., et al., *Detection of chromosomal instability in bystander cells after Si490-ion irradiation.* Radiat Res, 2011. 176(3): p. 280-90.

38. Iyer, R., B.E. Lehnert, and R. Svensson, *Factors underlying the cell growth-related bystander responses to alpha particles.* Cancer Res, 2000. 60(5): p. 1290-8.

39. Azzam, E.I., S.M. de Toledo, and J.B. Little, *Oxidative metabolism, gap junctions and the ionizing radiation-induced bystander effect.* Oncogene, 2003. 22(45): p. 7050-7.

40. Suzuki, M. and C. Tsuruoka, *Heavy charged particles produce a bystander effect via cell-cell junctions.* Biol Sci Space, 2004. 18(4): p. 241-6.

41. Bishayee, A., et al., *Free radical-initiated and gap junction-mediated bystander effect due to nonuniform distribution of incorporated radioactivity in a three-dimensional tissue culture model.* Radiat Res, 2001. 155(2): p. 335-44.

42. Finkel, T. and N.J. Holbrook, *Oxidants, oxidative stress and the biology of ageing.* Nature, 2000. 408(6809): p. 239-47.

43. Maslov, A.Y. and J. Vijg, *Genome instability, cancer and aging.* Biochim Biophys Acta, 2009. 1790(10): p. 963-9.

44. Negrini, S., V.G. Gorgoulis, and T.D. Halazonetis, *Genomic instability--an evolving hallmark of cancer.* Nat Rev Mol Cell Biol, 2010. 11(3): p. 220-8.

45. Marder, B.A. and W.F. Morgan, *Delayed chromosomal instability induced by DNA damage.* Mol Cell Biol, 1993. 13(11): p. 6667-77.

46. Weissenborn, U. and C. Streffer, *The one-cell mouse embryo: cell cycle-dependent radiosensitivity and development of chromosomal anomalies in postradiation cell cycles.* Int J Radiat Biol, 1988. 54(4): p. 659-74.

47. Kadhim, M.A., et al., *Radiation-induced genomic instability: delayed cytogenetic aberrations and apoptosis in primary human bone marrow cells.* Int J Radiat Biol, 1995. 67(3): p. 287-93.

48. Seymour, C.B., C. Mothersill, and T. Alper, *High yields of lethal mutations in somatic mammalian cells that survive ionizing radiation.* Int J Radiat Biol Relat Stud Phys Chem Med, 1986. 50(1): p. 167-79.

49. Selvanayagam, C.S., et al., *Latent expression of p53 mutations and radiation-induced mammary cancer.* Cancer Res, 1995. 55(15): p. 3310-7.

50. Mendonca, M.S., R.J. Antoniono, and J.L. Redpath, *Delayed heritable damage and epigenetics in radiation-induced neoplastic transformation of human hybrid cells.* Radiat Res, 1993. 134(2): p. 209-16.

51. Wright, E.G., *Manifestations and mechanisms of non-targeted effects of ionizing radiation.* Mutat Res, 2010. 687(1-2): p. 28-33.

52. Aypar, U., W.F. Morgan, and J.E. Baulch, *Radiation-induced genomic instability: are epigenetic mechanisms the missing link?* Int J Radiat Biol, 2011. 87(2): p. 179-91.

53. Kim, G.J., G.M. Fiskum, and W.F. Morgan, *A role for mitochondrial dysfunction in perpetuating radiation-induced genomic instability.* Cancer Res, 2006. 66(21): p. 10377-83.

54. Dayal, D., et al., *Mitochondrial complex II dysfunction can contribute significantly to genomic instability after exposure to ionizing radiation.* Radiat Res, 2009. 172(6): p. 737-45.

55. Pogribny, I., et al., *Fractionated low-dose radiation exposure leads to accumulation of DNA damage and profound alterations in DNA and histone methylation in the murine thymus.* Mol Cancer Res, 2005. 3(10): p. 553-61.

56. Pogribny, I., et al., *Dose-dependence, sex- and tissue-specificity, and persistence of radiation-induced genomic DNA methylation changes.* Biochem Biophys Res Commun, 2004. 320(4): p. 1253-61.

57. Wakabayashi, T., et al., *Studies of the mortality of A-bomb survivors, report 7. Part III. incidence of cancer in 1959-1978, based on the tumor registry, Nagasaki.* Radiat Res, 1983. 93(1): p. 112-46.

58. Ron, E., et al., *Skin tumor risk among atomic-bomb survivors in Japan.* Cancer Causes Control, 1998. 9(4): p. 393-401.

59. Dal Maso, L., et al., *Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors.* Cancer Causes Control, 2009. 20(1): p. 75-86.

60. Cardis, E., et al., *Risk of thyroid cancer after exposure to 131I in childhood.* J Natl Cancer Inst, 2005. 97(10): p. 724-32.

61. Cardis, E. and M. Hatch, *The Chernobyl accident--an epidemiological perspective.* Clin Oncol (R Coll Radiol), 2011. 23(4): p. 251-60.

62. Shore, R.E., et al., *Skin cancer incidence among children irradiated for ringworm of the scalp.* Radiat Res, 1984. 100(1): p. 192-204.

63. Aslan, G., et al., *Basal cell carcinoma of the scalp 70 years after irradiation.* Ann Plast Surg, 2002. 48(2): p. 216-7.

64. Sadetzki, S., et al., *Risk of thyroid cancer after childhood exposure to ionizing radiation for tinea capitis.* J Clin Endocrinol Metab, 2006. 91(12): p. 4798-804.

65. Ron, E., et al., *Radiation-induced skin carcinomas of the head and neck.* Radiat Res, 1991. 125(3): p. 318-25.

66. Hall, E.J., *Intensity-modulated radiation therapy, protons, and the risk of second cancers.* Int J Radiat Oncol Biol Phys, 2006. 65(1): p. 1-7.

67. Bartkowiak, D., et al., *Second cancer after radiotherapy, 1981-2007.* Radiother Oncol, 2011.

68. Fazel, R., et al., *Exposure to low-dose ionizing radiation from medical imaging procedures.* N Engl J Med, 2009. 361(9): p. 849-57.

69. Bhargavan, M., *Trends in the utilization of medical procedures that use ionizing radiation.* Health Phys, 2008. 95(5): p. 612-27.

70. Cuttler, J.M., *Editorial: is airport body-scan radiation a health risk?* Dose Response, 2011. 9(1): p. 1-5.

71. Berrington de Gonzalez, A. and S. Darby, *Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries.* Lancet, 2004. 363(9406): p. 345-51.

72. Doody, M.M., et al., *Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study.* Spine (Phila Pa 1976), 2000. 25(16): p. 2052-63.

73. Eisenberg, M.J., et al., *Cancer risk related to low-dose ionizing radiation from cardiac imaging in patients after acute myocardial infarction.* CMAJ, 2011. 183(4): p. 430-6.

74. Tondel, M., et al., *Increased incidence of malignancies in Sweden after the Chernobyl accident--a promoting effect?* Am J Ind Med, 2006. 49(3): p. 159-68.

75. Maenhaut, C., et al., *Gene expression profiles for radiation-induced thyroid cancer.* Clin Oncol (R Coll Radiol), 2011. 23(4): p. 282-8.

76. Tubiana, M., *Dose-effect relationship and estimation of the carcinogenic effects of low doses of ionizing radiation: the joint report of the Academie des Sciences (Paris) and of the Academie Nationale de Medecine.* Int J Radiat Oncol Biol Phys, 2005. 63(2): p. 317-9.

77. Ogata, H., *A review of some epidemiological studies on cancer risk from low-dose radiation or other carcinogenic agents.* Radiat Prot Dosimetry, 2011. 146(1-3): p. 268-71.

78. Hofman, M., et al., *Cancer-related fatigue: the scale of the problem.* Oncologist, 2007. 12 Suppl 1: p. 4-10.

79. York, J.M., et al., *The biobehavioral and neuroimmune impact of low-dose ionizing radiation.* Brain Behav Immun, 2011.

80. Martinez, A., et al., *An assessment of immediate DNA damage to occupationally exposed workers to low dose ionizing radiation by using the comet assay.* Rev Invest Clin, 2010. 62(1): p. 23-30.

81. Cheriyan, V.D., et al., *Genetic monitoring of the human population from high-level natural radiation areas of Kerala on the southwest coast of India. II. Incidence of numerical and structural chromosomal aberrations in the lymphocytes of newborns.* Radiat Res, 1999. 152(6 Suppl): p. S154-8.

82. Wang, S.S., et al., *Chromosomal aberrations in peripheral blood lymphocytes and risk for non-Hodgkin lymphoma.* J Natl Cancer Inst Monogr, 2008(39): p. 78-82.

83. Chen, D. and L. Wei, *Chromosome aberration, cancer mortality and hormetic phenomena among inhabitants in areas of high background radiation in China.* J Radiat Res (Tokyo), 1991. 32 Suppl 2: p. 46-53.

84. Albrecht, H., et al., *Transcriptional response of ex vivo human skin to ionizing radiation: comparison between low- and high-dose effects.* Radiat Res, 2012. 177(1): p. 69-83.

85. Lu, X., et al., *Radiation-induced changes in gene expression involve recruitment of existing messenger RNAs to and away from polysomes.* Cancer Res, 2006. 66(2): p. 1052-61.

86. Boveris A, C.E., ed. *Production of superoxide radicals and hydrogen peroxide in mitochondria*. Superoxide Dismutase, ed. L.W. Oberley. Vol. II. 1982, CRC Press Inc., Boca Raton, Florida,.

87. Sies, H., *Oxidative stress: oxidants and antioxidants.* Exp Physiol, 1997. 82(2): p. 291-5.

88. Ward, J.E., *Nature of Lesions Formed by Ionizing Radiation*. DNA Damage and Repair, ed. J.A. Nickoloff and M.F. Hoekstra. Vol. 2. 1998, Totowa NJ: Human Press.

89. Ward, J.F., W.F. Blakely, and E.I. Joner, *Mammalian cells are not killed by DNA single-strand breaks caused by hydroxyl radicals from hydrogen peroxide.* Radiat Res, 1985. 103(3): p. 383-92.

90. Ward, J.F., *The complexity of DNA damage: relevance to biological consequences.* Int J Radiat Biol, 1994. 66(5): p. 427-32.

**91.** **Mahaney, B.L., K. Meek, and S.P. Lees-Miller, *Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining.* Biochem J, 2009. 417(3): p. 639-50.**

**92.** **Parikh, S.S., et al., *Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA.* EMBO J, 1998. 17(17): p. 5214-26.**

**93.** **Kubota, Y., et al., *Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein.* EMBO J, 1996. 15(23): p. 6662-70.**

**94.** **Mourgues, S., M.E. Lomax, and P. O'Neill, *Base excision repair processing of abasic site/single-strand break lesions within clustered damage sites associated with XRCC1 deficiency.* Nucleic Acids Res, 2007. 35(22): p. 7676-87.**

**95.** **Hutton, R.D., et al., *PCNA stimulates catalysis by structure-specific nucleases using two distinct mechanisms: substrate targeting and catalytic step.* Nucleic Acids Res, 2008. 36(21): p. 6720-7.**

**96.** **Dianov, G.L. and J.L. Parsons, *Co-ordination of DNA single strand break repair.* DNA Repair (Amst), 2007. 6(4): p. 454-60.**

**97.** **Strom, C.E., et al., *Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate.* Nucleic Acids Res, 2011. 39(8): p. 3166-75.**

**98.** **Kastan, M.B. and D.S. Lim, *The many substrates and functions of ATM.* Nat Rev Mol Cell Biol, 2000. 1(3): p. 179-86.**

**99.** **Bakkenist, C.J. and M.B. Kastan, *DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation.* Nature, 2003. 421(6922): p. 499-506.**

**100.** **M.F. Lavin, D.D., L. Chessa. *ATM and the DNA damage response-Workshop on ataxia-telangiectasia and related syndromes*. 2006. EMBO Rep.**

**101.** **van Attikum, H. and S.M. Gasser, *Crosstalk between histone modifications during the DNA damage response.* Trends Cell Biol, 2009. 19(5): p. 207-17.**

**102.** **Harris, J.L. and K.K. Khanna, *BRCA1 A-complex fine tunes repair functions of BRCA1.* Aging (Albany NY), 2011. 3(5): p. 461-3.**

**103.** **Cheng, Q., et al., *ATM activates p53 by regulating MDM2 oligomerization and E3 processivity.* EMBO J, 2009. 28(24): p. 3857-67.**

**104.** **Fabbro, M., et al., *BRCA1-BARD1 complexes are required for p53Ser-15 phosphorylation and a G1/S arrest following ionizing radiation-induced DNA damage.* J Biol Chem, 2004. 279(30): p. 31251-8.**

**105.** **He, G., et al., *Induction of p21 by p53 following DNA damage inhibits both Cdk4 and Cdk2 activities.* Oncogene, 2005. 24(18): p. 2929-43.**

106. Wu, X., et al., *The p53-mdm-2 autoregulatory feedback loop.* Genes Dev, 1993. 7(7A): p. 1126-32.

107. Vogelstein, B., D. Lane, and A.J. Levine, *Surfing the p53 network.* Nature, 2000. 408(6810): p. 307-10.

108. Falck, J., et al., *The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways.* Nat Genet, 2002. 30(3): p. 290-4.

109. Peng, C.Y., et al., *Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216.* Science, 1997. 277(5331): p. 1501-5.

110. Lieber, M.R., et al., *Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system.* Cell Res, 2008. 18(1): p. 125-33.

111. Fattah, F., et al., *Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells.* PLoS Genet, 2010. 6(2): p. e1000855.

112. Zha, S., C. Boboila, and F.W. Alt, *Mre11: roles in DNA repair beyond homologous recombination.* Nat Struct Mol Biol, 2009. 16(8): p. 798-800.

113. Wang, M., et al., *PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways.* Nucleic Acids Res, 2006. 34(21): p. 6170-82.

114. Williams, R.S., et al., *Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair.* Cell, 2008. 135(1): p. 97-109.

115. Wold, M.S., *Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism.* Annu Rev Biochem, 1997. 66: p. 61-92.

116. Li, X. and W.D. Heyer, *Homologous recombination in DNA repair and DNA damage tolerance.* Cell Res, 2008. 18(1): p. 99-113.

117. Shrivastav, M., L.P. De Haro, and J.A. Nickoloff, *Regulation of DNA double-strand break repair pathway choice.* Cell Res, 2008. 18(1): p. 134-47.

118. Branzei, D. and M. Foiani, *Regulation of DNA repair throughout the cell cycle.* Nat Rev Mol Cell Biol, 2008. 9(4): p. 297-308.

119. Rothkamm, K., et al., *Pathways of DNA double-strand break repair during the mammalian cell cycle.* Mol Cell Biol, 2003. 23(16): p. 5706-15.

120. Li, Q., et al., *Daxx cooperates with the Axin/HIPK2/p53 complex to induce cell death.* Cancer Res, 2007. 67(1): p. 66-74.

121. Thornborrow, E.C., et al., *A conserved intronic response element mediates direct p53-dependent transcriptional activation of both the human and murine bax genes.* Oncogene, 2002. 21(7): p. 990-9.

122. Yu, J., et al., *PUMA mediates the apoptotic response to p53 in colorectal cancer cells.* Proc Natl Acad Sci U S A, 2003. 100(4): p. 1931-6.

123. Oda, E., et al., *Noxa, a BH3-only member of the Bcl-2 family and candidate mediator of p53-induced apoptosis.* Science, 2000. 288(5468): p. 1053-8.

124. Yin, X.M., Z.N. Oltvai, and S.J. Korsmeyer, *BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax.* Nature, 1994. 369(6478): p. 321-3.

125. Owen-Schaub, L.B., et al., *Wild-type human p53 and a temperature-sensitive mutant induce Fas/APO-1 expression.* Mol Cell Biol, 1995. 15(6): p. 3032-40.

126. Salvesen, G.S. and V.M. Dixit, *Caspase activation: the induced-proximity model.* Proc Natl Acad Sci U S A, 1999. 96(20): p. 10964-7.

127. Yang, J. and P.J. Duerksen-Hughes, *Activation of a p53-independent, sphingolipid-mediated cytolytic pathway in p53-negative mouse fibroblast cells treated with N-methyl-N-nitro-N-nitrosoguanidine.* J Biol Chem, 2001. 276(29): p. 27129-35.

128. Santana, P., et al., *Acid sphingomyelinase-deficient human lymphoblasts and mice are defective in radiation-induced apoptosis.* Cell, 1996. 86(2): p. 189-99.

129. Verheij, M., et al., *Requirement for ceramide-initiated SAPK/JNK signalling in stress-induced apoptosis.* Nature, 1996. 380(6569): p. 75-9.

130. Lee, Y.J., et al., *Protective role of Bcl2 in metabolic oxidative stress-induced cell death.* J Cell Sci, 2001. 114(Pt 4): p. 677-84.

131. Manome, Y., et al., *Effect of Bcl-2 on ionizing radiation and 1-beta-D-arabinofuranosylcytosine-induced internucleosomal DNA fragmentation and cell survival in human myeloid leukemia cells.* Oncol Res, 1993. 5(3): p. 139-44.

132. Deng, G., et al., *Bcl-2 facilitates recovery from DNA damage after oxidative stress.* Exp Neurol, 1999. 159(1): p. 309-18.

133. Hockenbery, D., et al., *Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death.* Nature, 1990. 348(6299): p. 334-6.

134. Beham, A., et al., *Bcl-2 inhibits p53 nuclear import following DNA damage.* Oncogene, 1997. 15(23): p. 2767-72.

135. Strasser, A., et al., *DNA damage can induce apoptosis in proliferating lymphoid cells via p53-independent mechanisms inhibitable by Bcl-2.* Cell, 1994. 79(2): p. 329-39.

136. Froesch, B.A., et al., *Inhibition of p53 transcriptional activity by Bcl-2 requires its membrane-anchoring domain.* J Biol Chem, 1999. 274(10): p. 6469-75.

137. Hemann, M.T. and S.W. Lowe, *The p53-Bcl-2 connection.* Cell Death Differ, 2006. 13(8): p. 1256-9.

138. Barnes, M., et al., *Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms.* Nucleic Acids Res, 2005. 33(18): p. 5914-23.

139. Gunderson, K.L., et al., *Decoding randomly ordered DNA arrays.* Genome Res, 2004. 14(5): p. 870-7.

140. Dobbin, K.K., et al., *Characterizing dye bias in microarray experiments.* Bioinformatics, 2005. 21(10): p. 2430-7.

141. Churchill, G.A., *Fundamentals of experimental design for cDNA microarrays.* Nat Genet, 2002. 32 Suppl: p. 490-5.

142. Pedotti, P., et al., *Can subtle changes in gene expression be consistently detected with different microarray platforms?* BMC Genomics, 2008. 9: p. 124.

143. Huang, S., et al., *At what scale should microarray data be analyzed?* Am J Pharmacogenomics, 2004. 4(2): p. 129-39.

144. Durbin, B.P., et al., *A variance-stabilizing transformation for gene-expression microarray data.* Bioinformatics, 2002. 18 Suppl 1: p. S105-10.

145. Lin, S.M., et al., *Model-based variance-stabilizing transformation for Illumina microarray data.* Nucleic Acids Res, 2008. 36(2): p. e11.

146. Quackenbush, J., *Microarray data normalization and transformation.* Nat Genet, 2002. 32 Suppl: p. 496-501.

147. Affymetrix. *Statistical Algorithms Description Document*. 2002 [cited 2012; Available from: http://media.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf.

148. 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.

149. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. 4(2): p. 249-64.

150. Wu, Z., et al., *A Model-Based Background Adjustment for Oligonucleotide Expression Arrays.* Journal of the American Statistical Association, 2004. 99(468): p. 909-917.

151. Du, P., W.A. Kibbe, and S.M. Lin, *lumi: a pipeline for processing Illumina microarray.* Bioinformatics, 2008. 24(13): p. 1547-8.

152. Du P, K.W., Lin S:. *Using lumi, a package processing Illumina Microarray.* Bioconductor package 2007; Available from: http://www.bioconductor.org/packages/2.6/bioc//vignettes/lumi/inst/doc/lumi.pdf.

153. Liang, S., S. Fuhrman, and R. Somogyi, *Reveal, a general reverse engineering algorithm for inference of genetic network architectures.* Pac Symp Biocomput, 1998: p. 18-29.

154. Friedman, N., et al., *Using Bayesian networks to analyze expression data.* J Comput Biol, 2000. 7(3-4): p. 601-20.

155. Djebbari, A. and J. Quackenbush, *Seeded Bayesian Networks: constructing genetic networks from microarray data.* BMC Syst Biol, 2008. 2: p. 57.

156. Kasturi, J., R. Acharya, and M. Ramanathan, *An information theoretic approach for analyzing temporal patterns of gene expression.* Bioinformatics, 2003. 19(4): p. 449-58.

157. Allison, D.B., et al., *Microarray data analysis: from disarray to consolidation and consensus.* Nat Rev Genet, 2006. 7(1): p. 55-65.

158. Tavazoie, S., et al., *Systematic determination of genetic network architecture.* Nat Genet, 1999. 22(3): p. 281-5.

159. 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.

160. Bezdek, J.C., *Pattern RecognitionWith Fuzzy Objective Function Algorithms*1981: Plenum Press,New York.

161. Gasch, A.P. and M.B. Eisen, *Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering.* Genome Biol, 2002. 3(11): p. RESEARCH0059.

162. Alsabti, K.R., S. ; and Singh, V. *An efficient k-means clustering algorithm*. in *Proceedings of IPPS 11th International Parallel Processing Symposium.* 1998.

163. Kaufmann, L. and P.J. Rousseeuw, *Clustering by means of medoids*, in *Statistical Data Analysis based on the L 1 Norm and Related Methods*, Y. Dodge, Editor 1987, Elsevier Science. p. 405–416.

164. Shannon, W., R. Culverhouse, and J. Duncan, *Analyzing microarray data using cluster analysis.* Pharmacogenomics, 2003. 4(1): p. 41-52.

165. Shyamsundar, R., et al., *A DNA microarray survey of gene expression in normal human tissues.* Genome Biol, 2005. 6(3): p. R22.

166. Lapointe, J., et al., *Gene expression profiling identifies clinically relevant subtypes of prostate cancer.* Proc Natl Acad Sci U S A, 2004. 101(3): p. 811-6.

167. Butte, A. and I. Kohane, *Relevance Networks: A First Step Toward Finding Genetic Regulatory Networks Within Microarray Data.* In The Analysis of Gene Expression Data2003, New York: Springer-Verlag.

168. Langston, M.A., et al., *Innovative computational methods for transcriptomic data analysis*, in *Proceedings of the 2006 ACM symposium on Applied computing*2006, ACM: Dijon, France. p. 190-194.

169. Voy, B.H., et al., *Extracting gene networks for low-dose radiation using graph theoretical algorithms.* PLoS Comput Biol, 2006. 2(7): p. e89.

170. Moriyama, M., et al., *Relevance network between chemosensitivity and transcriptome in human hepatoma cells.* Mol Cancer Ther, 2003. 2(2): p. 199-205.

171. Butte, A.J., et al., *Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks.* Proc Natl Acad Sci U S A, 2000. 97(22): p. 12182-6.

172. Grindley, H.M., et al., *Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm.* J Mol Biol, 1993. 229(3): p. 707-21.

173. Mitchell, E.M., et al., *Use of techniques derived from graph theory to compare secondary structure motifs in proteins.* J Mol Biol, 1990. 212(1): p. 151-66.

174. Koch, I. and T. Lengauer, *Detection of distant structural similarities in a set of proteins using a fast graph-based method.* Proc Int Conf Intell Syst Mol Biol, 1997. 5: p. 167-78.

175. Ullah, E., L. Kyongbum, and S. Hassoun. *An algorithm for identifying dominant-edge metabolic pathways*. in *Computer-Aided Design - Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on*. 2009.

176. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. 25(1): p. 25-9.

177. Ashburner, M. and S. Lewis, *On ontologies for biologists: the Gene Ontology--untangling the web.* Novartis Found Symp, 2002. 247: p. 66-80; discussion 80-3, 84-90, 244-52.

178. Harris, M.A., et al., *The Gene Ontology (GO) database and informatics resource.* Nucleic Acids Res, 2004. 32(Database issue): p. D258-61.

179. Baker, E.J., et al., *GeneWeaver: a web-based system for integrative functional genomics.* Nucleic Acids Res, 2012. 40(Database issue): p. D1067-76.

180. Kohl, M., S. Wiese, and B. Warscheid, *Cytoscape: software for visualization and analysis of biological networks.* Methods Mol Biol, 2011. 696: p. 291-303.

181. Baker, E.J., et al., *Ontological Discovery Environment: a system for integrating gene-phenotype associations.* Genomics, 2009. 94(6): p. 377-87.

182. Abu Khzam, F.N., M.A. Langston, and W.H. Suters, *Effective Vertex Cover Kernelization: A Tale of Two Algorithms*, in *Proceedings of Computer Systems and Applications, 2005. The 3rd ACS/IEEE International Conference*2005, AICCSA: Cairo, Egypt.

183. Borate, B.R., et al., *Comparison of threshold selection methods for microarray gene co-expression matrices.* BMC Res Notes, 2009. 2: p. 240.

184. Langston, M.A. *Practical FPT Implementations and Applications Proceedings of Parameterized and Exact Computation: First International Workshop*. in *WPEC 2004*. 2004. Bergen, Norway.

185. Perkins, A.D. and M.A. Langston, *Threshold selection in gene co-expression networks using spectral graph theory techniques.* BMC Bioinformatics, 2009. 10 Suppl 11: p. S4.

186. Zhang Y., et al., *Genome-Scale Computational Approaches to Memory-Intensive Applications in Systems Biology*, in *Proceedings of the ACM/IEEE SC*2005.

187. Lynch, R.M., et al., *Identifying genetic loci and spleen gene coexpression networks underlying immunophenotypes in BXD recombinant inbred mice.* Physiol Genomics, 2010.

188. Bruhn, S., et al., *Increased expression of IRF4 and ETS1 in CD4+ cells from patients with intermittent allergic rhinitis.* Allergy, 2012. 67(1): p. 33-40.

189. Zotenko, E., et al., *Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality.* PLoS Comput Biol, 2008. 4(8): p. e1000140.

190. He, X. and J. Zhang, *Why do hubs tend to be essential in protein networks?* PLoS Genet, 2006. 2(6): p. e88.

191. Jeong H, M.S., Barabási AL,Oltvai ZN, *Lethality and centrality in protein networks.* Nature, 2001. 411(6833): p. 41-42.

192. Joy, M.P., et al., *High-betweenness proteins in the yeast protein interaction network.* J Biomed Biotechnol, 2005. 2005(2): p. 96-103.

193. Shen-Orr, S.S., et al., *Network motifs in the transcriptional regulation network of Escherichia coli.* Nat Genet, 2002. 31(1): p. 64-8.

194. Zhang, B., M.C. Chambers, and D.L. Tabb, *Proteomic parsimony through bipartite graph analysis improves accuracy and transparency.* J Proteome Res, 2007. 6(9): p. 3549-57.

195. Bohland, J.W., et al., *The brain atlas concordance problem: quantitative comparison of anatomical parcellations.* PLoS One, 2009. 4(9): p. e7200.

196. Bleakley, K. and Y. Yamanishi, *Supervised prediction of drug-target interactions using bipartite local models.* Bioinformatics, 2009. 25(18): p. 2397-403.

197. Cohen-Gihon, I., R. Nussinov, and R. Sharan, *Comprehensive analysis of co-occurring domain sets in yeast proteins.* BMC Genomics, 2007. 8: p. 161.

198. Yamanishi, Y., et al., *Prediction of drug-target interaction networks from the integration of chemical and genomic spaces.* Bioinformatics, 2008. 24(13): p. i232-40.

199. Basler, G., et al., *Evolutionary significance of metabolic network properties.* J R Soc Interface, 2012. 9(71): p. 1168-76.

200. Storey, J.D., *The positive False Discovery Rate: A Bayesian. Interpretation and the q-value.* The Annals of Statistics, 2003. 31 (6): p. 2013–2035.

201. Benjamini Y. and Hochberg Y., *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.* . Journal of the Royal Statistical Society Series B-Methodological 1995. 57: 289-300, : p. 289-300.

202. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, 2003. 100(16): p. 9440-5.

203. Berriz, G.F., et al., *Characterizing gene sets with FuncAssociate.* Bioinformatics, 2003. 19(18): p. 2502-4.

204. Maere, S., K. Heymans, and M. Kuiper, *BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.* Bioinformatics, 2005. 21(16): p. 3448-9.

205. Beißbarth, T. and T.P. Speed, *GOstat: find statistically overrepresented Gene Ontologies within a group of genes.* Bioinformatics, 2004. 20(9): p. 1464-1465.

206. Sevilla, J.L., et al., *Correlation between gene expression and GO semantic similarity.* IEEE/ACM Trans Comput Biol Bioinform, 2005. 2(4): p. 330-8.

207. Tesson, B.M., R. Breitling, and R.C. Jansen, *DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules.* BMC Bioinformatics, 2010. 11: p. 497.

208. Cho, S.B., J. Kim, and J.H. Kim, *Identifying set-wise differential co-expression in gene expression microarray data.* BMC Bioinformatics, 2009. 10: p. 109.

209. Choi, J.K., et al., *Differential coexpression analysis using microarray data and its application to human cancer.* Bioinformatics, 2005. 21(24): p. 4348-55.

210. Choi, Y. and C. Kendziorski, *Statistical methods for gene set co-expression analysis.* Bioinformatics, 2009. 25(21): p. 2780-6.

211. Lai, Y., et al., *A statistical method for identifying differential gene-gene co-expression patterns.* Bioinformatics, 2004. 20(17): p. 3146-55.

212. Southworth, L.K., A.B. Owen, and S.K. Kim, *Aging mice show a decreasing correlation of gene expression within genetic modules.* PLoS Genet, 2009. 5(12): p. e1000776.

213. Altay, G., et al., *Differential C3NET reveals disease networks of direct physical interactions.* BMC Bioinformatics, 2011. 12: p. 296.

214. van Nas, A., et al., *Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks.* Endocrinology, 2009. 150(3): p. 1235-49.

215. Yu, H., et al., *Link-based quantitative methods to identify differentially coexpressed genes and gene pairs.* BMC Bioinformatics, 2011. 12: p. 315.

**216.** **Bruhn S, et al., *Increased expression of IRF4 and ETS1 in CD4+ cells from patients with intermittent allergic rhinitis. .* Allergy 2011, 2011.**

**217.** **Imamura, H., et al., *Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data.* BMC Bioinformatics, 2010. 11: p. 232.**

**218.** **Lee, A.B., et al., *Discovering genetic ancestry using spectral graph theory.* Genet Epidemiol, 2010. 34(1): p. 51-9.**

**219.** **Brenner, D.J., et al., *Cancer risks attributable to low doses of ionizing radiation: assessing what we really know.* Proc Natl Acad Sci U S A, 2003. 100(24): p. 13761-6.**

**220.** **Luckey, T.D., *Physiological benefits from low levels of ionizing radiation.* Health Phys, 1982. 43(6): p. 771-89.**

**221.** **Bachmann, H.S., et al., *Regulatory BCL2 promoter polymorphism (-938C>A) is associated with adverse outcome in patients with prostate carcinoma.* Int J Cancer, 2011. 129(10): p. 2390-9.**

**222.** **Bachmann, H.S., et al., *The AA genotype of the regulatory BCL2 promoter polymorphism ( 938C>A) is associated with a favorable outcome in lymph node negative invasive breast cancer patients.* Clin Cancer Res, 2007. 13(19): p. 5790-7.**

**223.** **Heubner, M., et al., *Association of the AA genotype of the BCL2 (-938C>A) promoter polymorphism with better survival in ovarian cancer.* Int J Biol Markers, 2009. 24(4): p. 223-9.**

**224.** **Lehnerdt, G.F., et al., *The regulatory BCL2 promoter polymorphism (-938C>A) is associated with relapse and survival of patients with oropharyngeal squamous cell carcinoma.* Ann Oncol, 2009. 20(6): p. 1094-9.**

**225.** **Eun, Y.G., et al., *A Polymorphism (rs1801018, Thr7Thr) of BCL2 is Associated with Papillary Thyroid Cancer in Korean Population.* Clin Exp Otorhinolaryngol, 2011. 4(3): p. 149-54.**

**226.** **Jiang, S.X., et al., *Bcl-2 protein expression in lung cancer and close correlation with neuroendocrine differentiation.* Am J Pathol, 1996. 148(3): p. 837-46.**

**227.** **Lee, S., et al., *Profiling of transcripts and proteins modulated by K-ras oncogene in the lung tissues of K-ras transgenic mice by omics approaches.* Int J Oncol, 2009. 34(1): p. 161-72.**

**228.** **Byrne, S.N., et al., *The immune-modulating cytokine and endogenous Alarmin interleukin-33 is upregulated in skin exposed to inflammatory UVB radiation.* Am J Pathol, 2011. 179(1): p. 211-22.**

**229.** **Volarevic, V., et al., *Protective role of IL-33/ST2 axis in Con A-induced hepatitis.* J Hepatol, 2011.**

**230.** **Seki, K., et al., *Interleukin-33 prevents apoptosis and improves survival after experimental myocardial infarction through ST2 signaling.* Circ Heart Fail, 2009. 2(6): p. 684-91.**

**231.** **Kang, C.B., et al., *FKBP family proteins: immunophilins with versatile biological functions.* Neurosignals, 2008. 16(4): p. 318-25.**

**232.** **Romano, S., et al., *Role of FK506-binding protein 51 in the control of apoptosis of irradiated melanoma cells.* Cell Death Differ, 2010. 17(1): p. 145-57.**

**233.** **Tamatani, M., et al., *Tumor necrosis factor induces Bcl-2 and Bcl-x expression through NFkappaB activation in primary hippocampal neurons.* J Biol Chem, 1999. 274(13): p. 8531-8.**

**234.** **Shirane, M. and K.I. Nakayama, *[Immunophilin FKBP38, an inherent inhibitor of calcineurin, targets Bcl-2 to mitochondria and inhibits apoptosis].* Nihon Rinsho, 2004. 62(2): p. 405-12.**

**235.** **Umezu, K., et al., *Genetic analysis of yeast RPA1 reveals its multiple functions in DNA metabolism.* Genetics, 1998. 148(3): p. 989-1005.**

**236.** **Balajee, A.S. and C.R. Geard, *Replication protein A and gamma-H2AX foci assembly is triggered by cellular response to DNA double-strand breaks.* Exp Cell Res, 2004. 300(2): p. 320-34.**

**237.** **Michiels, S., et al., *Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers.* Carcinogenesis, 2007. 28(8): p. 1731-9.**

**238.** **Wang, Y., et al., *Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice.* Nat Genet, 2005. 37(7): p. 750-5.**

**239.** **Leandro-Garcia, L.J., et al., *Tumoral and tissue-specific expression of the major human beta-tubulin isotypes.* Cytoskeleton (Hoboken), 2010. 67(4): p. 214-23.**

**240.** **Huser, C.A., et al., *TSC-22D1 isoforms have opposing roles in mammary epithelial cell survival.* Cell Death Differ, 2010. 17(2): p. 304-15.**

**241.** **Lu, Y., et al., *Identification of TSC-22 as a potential tumor suppressor that is upregulated by Flt3-D835V but not Flt3-ITD.* Leukemia, 2007. 21(11): p. 2246-57.**

**242.** **Berndt, A., et al., *Identification of fat4 and tsc22d1 as novel candidate genes for spontaneous pulmonary adenomas.* Cancer Res, 2011. 71(17): p. 5779-91.**

**243.** **May, W.A., et al., *EWS/FLI1-induced manic fringe renders NIH 3T3 cells tumorigenic.* Nat Genet, 1997. 17(4): p. 495-7.**

**244.** **Doil, C., et al., *RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins.* Cell, 2009. 136(3): p. 435-46.**

**245.** **Stewart, G.S., et al., *The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage.* Cell, 2009. 136(3): p. 420-34.**

**246.** **Sy, S.M., et al., *Critical roles of ring finger protein RNF8 in replication stress responses.* J Biol Chem, 2011. 286(25): p. 22355-61.**

247. Devgan, S.S., et al., *Homozygous deficiency of ubiquitin-ligase ring-finger protein RNF168 mimics the radiosensitivity syndrome of ataxia-telangiectasia.* Cell Death Differ, 2011. 18(9): p. 1500-6.

248. Bohgaki, T., et al., *Genomic instability, defective spermatogenesis, immunodeficiency, and cancer in a mouse model of the RIDDLE syndrome.* PLoS Genet, 2011. 7(4): p. e1001381.

249. Baldeyron, C., et al., *HP1alpha recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair.* J Cell Biol, 2011. 193(1): p. 81-95.

250. Dinant, C. and M.S. Luijsterburg, *The emerging role of HP1 in the DNA damage response.* Mol Cell Biol, 2009. 29(24): p. 6335-40.

251. Luijsterburg, M.S., et al., *Heterochromatin protein 1 is recruited to various types of DNA damage.* J Cell Biol, 2009. 185(4): p. 577-86.

252. Okamoto, K. and D. Beach, *Cyclin G is a transcriptional target of the p53 tumor suppressor protein.* EMBO J, 1994. 13(20): p. 4816-22.

253. Kimura, S.H. and H. Nojima, *Cyclin G1 associates with MDM2 and regulates accumulation and degradation of p53 protein.* Genes Cells, 2002. 7(8): p. 869-80.

254. Sasaki, M., et al., *Regulation of the MDM2-P53 pathway and tumor growth by PICT1 via nucleolar RPL11.* Nat Med, 2011. 17(8): p. 944-51.

255. Mizushima, N., et al., *Dissection of autophagosome formation using Apg5-deficient mouse embryonic stem cells.* J Cell Biol, 2001. 152(4): p. 657-68.

256. Kabeya, Y., et al., *LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing.* EMBO J, 2000. 19(21): p. 5720-8.

257. Sillje, H.H., et al., *Mammalian homologues of the plant Tousled gene code for cell-cycle-regulated kinases with maximal activities linked to ongoing DNA replication.* EMBO J, 1999. 18(20): p. 5691-702.

258. Narod, S.A. and W.D. Foulkes, *BRCA1 and BRCA2: 1994 and beyond.* Nat Rev Cancer, 2004. 4(9): p. 665-76.

259. Dong, Y., et al., *Regulation of BRCC, a holoenzyme complex containing BRCA1 and BRCA2, by a signalosome-like subunit and its role in DNA repair.* Mol Cell, 2003. 12(5): p. 1087-99.

260. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.* Nucleic Acids Res, 2010. 38(Web Server issue): p. W214-20.

261. Li, Q., et al., *A death receptor-associated anti-apoptotic protein, BRE, inhibits mitochondrial apoptotic pathway.* J Biol Chem, 2004. 279(50): p. 52106-16.

262. Chan, J.Y., et al., *Differential expression of a novel gene BRE (TNFRSF1A modulator/BRCC45) in response to stress and biological signals.* Mol Biol Rep, 2010. 37(1): p. 363-8.

263. Basu, A. and S. Krishnamurthy, *Cellular responses to Cisplatin-induced DNA damage.* J Nucleic Acids, 2010. 2010.

264. Dejmek, J., J.D. Iglehart, and J.B. Lazaro, *DNA-dependent protein kinase (DNA-PK)-dependent cisplatin-induced loss of nucleolar facilitator of chromatin transcription (FACT) and regulation of cisplatin sensitivity by DNA-PK and FACT.* Mol Cancer Res, 2009. 7(4): p. 581-91.

265. Yu, Z., et al., *A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation.* Mol Cell Biol, 2009. 29(11): p. 2982-96.

266. An, W., J. Kim, and R.G. Roeder, *Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53.* Cell, 2004. 117(6): p. 735-48.

267. Mao, H., et al., *RhoBTB2 (DBC2) functions as tumor suppressor via inhibiting proliferation, preventing colony formation and inducing apoptosis in breast cancer cells.* Gene, 2011. 486(1-2): p. 74-80.

268. Ling, L.J., et al., *Ectopic expression of RhoBTB2 inhibits migration and invasion of human breast cancer cells.* Cancer Biol Ther, 2010. 10(11): p. 1115-22.

269. Pahl, H.L., *Activators and target genes of Rel/NF-kappaB transcription factors.* Oncogene, 1999. 18(49): p. 6853-66.

270. Barberi, T. and M.A.S. Moore, *The LYL-1 oncoprotein is ubiquitously expressed in the mouse.* Experimental hematology, 2000. 28(7): p. 45.

271. Lukov, G.L., et al., *The expansion of T-cells and hematopoietic progenitors as a result of overexpression of the lymphoblastic leukemia gene, Lyl1 can support leukemia formation.* Leuk Res, 2011. 35(3): p. 405-12.

272. Zhong, Y., et al., *Overexpression of a transcription factor LYL1 induces T- and B-cell lymphoma in mice.* Oncogene, 2007. 26(48): p. 6937-47.

273. Friedberg, E.C., A.R. Lehmann, and R.P. Fuchs, *Trading places: how do DNA polymerases switch during translesion DNA synthesis?* Mol Cell, 2005. 18(5): p. 499-505.

274. Masuda, Y., et al., *Mechanisms of dCMP transferase reactions catalyzed by mouse Rev1 protein.* J Biol Chem, 2002. 277(4): p. 3040-6.

275. Jansen, J.G., et al., *Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice.* J Exp Med, 2006. 203(2): p. 319-23.

276. Chen, C.C., et al., *Genetic analysis of ionizing radiation-induced mutagenesis in Saccharomyces cerevisiae reveals TransLesion Synthesis (TLS) independent of PCNA K164 SUMOylation and ubiquitination.* DNA Repair (Amst), 2006. 5(12): p. 1475-88.

277. Sharma, S., et al., *REV1 and polymerase {zeta} facilitate homologous recombination repair.* Nucleic Acids Res, 2011.

278. Simpson, L.J. and J.E. Sale, *Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line.* EMBO J, 2003. 22(7): p. 1654-64.

279. Hanoux, V., et al., *Caspase-2 involvement during ionizing radiation-induced oocyte death in the mouse ovary.* Cell Death Differ, 2007. 14(4): p. 671-81.

280. Harvey, N.L., A.J. Butt, and S. Kumar, *Functional activation of Nedd2/ICH-1 (caspase-2) is an early process in apoptosis.* J Biol Chem, 1997. 272(20): p. 13134-9.

281. Hafer, K., et al., *Adaptive response to gamma radiation in mammalian cells proficient and deficient in components of nucleotide excision repair.* Radiat Res, 2007. 168(2): p. 168-74.

282. Veugelers, K., et al., *Granule-mediated killing by granzyme B and perforin requires a mannose 6-phosphate receptor and is augmented by cell surface heparan sulfate.* Mol Biol Cell, 2006. 17(2): p. 623-33.

283. Conticello, S.G., *The AID/APOBEC family of nucleic acid mutators.* Genome Biol, 2008. 9(6): p. 229.

284. Morgan, H.D., et al., *Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming.* J Biol Chem, 2004. 279(50): p. 52353-60.

285. Anant, S., et al., *Apobec-1 protects intestine from radiation injury through posttranscriptional regulation of cyclooxygenase-2 expression.* Gastroenterology, 2004. 127(4): p. 1139-49.

286. Yamanaka, S., et al., *Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals.* Proc Natl Acad Sci U S A, 1995. 92(18): p. 8483-7.

287. Rong, R., et al., *TC21 mediates transformation and cell survival via activation of phosphatidylinositol 3-kinase/Akt and NF-kappaB signaling pathway.* Oncogene, 2002. 21(7): p. 1062-70.

288. Bos, J.L., *Ras-like GTPases.* Biochim Biophys Acta, 1997. 1333(2): p. M19-31.

289. Orr, S.J., et al., *Reducing MCM levels in human primary T cells during the G(0)-->G(1) transition causes genomic instability during the first cell cycle.* Oncogene, 2010. 29(26): p. 3803-14.

290. Tsao, C.C., C. Geisen, and R.T. Abraham, *Interaction between human MCM7 and Rad17 proteins is required for replication checkpoint signaling.* EMBO J, 2004. 23(23): p. 4660-9.

291. Hayakawa, F. and M.L. Privalsky, *Phosphorylation of PML by mitogen-activated protein kinases plays a key role in arsenic trioxide-mediated apoptosis.* Cancer Cell, 2004. 5(4): p. 389-401.

292. Mann, K.K. and W.H. Miller, Jr., *Death by arsenic: implications of PML sumoylation.* Cancer Cell, 2004. 5(4): p. 307-9.

293. Yang, S., et al., *PML-dependent apoptosis after DNA damage is regulated by the checkpoint kinase hCds1/Chk2.* Nat Cell Biol, 2002. 4(11): p. 865-70.

294. Reineke, E.L. and H.Y. Kao, *Targeting promyelocytic leukemia protein: a means to regulating PML nuclear bodies.* Int J Biol Sci, 2009. 5(4): p. 366-76.

295. Miller, K.M., et al., *Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining.* Nat Struct Mol Biol, 2010. 17(9): p. 1144-51.

296. Tan, J., et al., *Protein kinase B/Akt-dependent phosphorylation of glycogen synthase kinase-3beta in irradiated vascular endothelium.* Cancer Res, 2006. 66(4): p. 2320-7.

297. Thotala, D.K., D.E. Hallahan, and E.M. Yazlovitskaya, *Glycogen synthase kinase 3beta inhibitors protect hippocampal neurons from radiation-induced apoptosis by regulating MDM2-p53 pathway.* Cell Death Differ, 2011.

298. Trivedi, C.M., et al., *Hdac2 regulates the cardiac hypertrophic response by modulating Gsk3 beta activity.* Nat Med, 2007. 13(3): p. 324-31.

299. Neye, H. and E.J. Verspohl, *The FK506 binding protein 13 kDa (FKBP13) interacts with the C-chain of complement C1q.* BMC Pharmacol, 2004. 4: p. 19.

300. Boiadzhian, A.S., et al., *[The influence of aromatic amino acid derivatives on the complement system under ionizing radiation].* Radiats Biol Radioecol, 2010. 50(4): p. 472-4.

301. Madigan, J.P., H.L. Chotkowski, and R.L. Glaser, *DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis.* Nucleic Acids Res, 2002. 30(17): p. 3698-705.

302. MacArthur, C.A. and G.M. Shackleford, *Npm3: a novel, widely expressed gene encoding a protein related to the molecular chaperones nucleoplasmin and nucleophosmin.* Genomics, 1997. 42(1): p. 137-40.

303. Dalenc, F., et al., *Increased expression of a COOH-truncated nucleophosmin resulting from alternative splicing is associated with cellular resistance to ionizing radiation in HeLa cells.* Int J Cancer, 2002. 100(6): p. 662-8.

304. Kanaar, R., et al., *Human and mouse homologs of the Saccharomyces cerevisiae RAD54 DNA repair gene: evidence for functional conservation.* Curr Biol, 1996. 6(7): p. 828-38.

305. Essers, J., et al., *Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination.* Cell, 1997. 89(2): p. 195-204.

306. Dronkert, M.L., et al., *Mouse RAD54 affects DNA double-strand break repair and sister chromatid exchange.* Mol Cell Biol, 2000. 20(9): p. 3147-56.

307. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. 5(10): p. R80.

308. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Res, 2002. 30(1): p. 207-10.

309. Sanoudou, D., et al., *Expression profiling reveals altered satellite cell numbers and glycolytic enzyme transcription in nemaline myopathy muscle.* Proc Natl Acad Sci U S A, 2003. 100(8): p. 4666-71.

310. Ala, U., et al., *Prediction of human disease genes by human-mouse conserved coexpression analysis.* PLoS Comput Biol, 2008. 4(3): p. e1000043.

311. Brem, R.B., et al., *Genetic dissection of transcriptional regulation in budding yeast.* Science, 2002. 296(5568): p. 752-5.

312. Bystrykh, L., et al., *Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'.* Nat Genet, 2005. 37(3): p. 225-32.

313. Chesler, E.J., et al., *Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function.* Nat Genet, 2005. 37(3): p. 233-42.

314. Gatti, D., et al., *Genome-level analysis of genetic regulation of liver gene expression networks.* Hepatology, 2007. 46(2): p. 548-57.

315. Stark, C., et al., *BioGRID: a general repository for interaction datasets.* Nucleic Acids Res, 2006. 34(Database issue): p. D535-9.

316. Linhart, C., Y. Halperin, and R. Shamir, *Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets.* Genome Res, 2008. 18(7): p. 1180-9.

317. Zheng, G., et al., *ITFP: an integrated platform of mammalian transcription factors.* Bioinformatics, 2008. 24(20): p. 2416-7.

318. Childress, P.J., R.L. Fletcher, and N.B. Perumal, *LymphTF-DB: a database of transcription factors involved in lymphocyte development.* Genes Immun, 2007. 8(4): p. 360-5.

319. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. 28(1): p. 27-30.

320. Evsikov, A.V., et al., *MouseCyc: a curated biochemical pathways database for the laboratory mouse.* Genome Biol, 2009. 10(8): p. R84.

321. Blake, J.A., et al., *The Mouse Genome Database (MGD). A comprehensive public resource of genetic, phenotypic and genomic data. The Mouse Genome Informatics Group.* Nucleic Acids Res, 1997. 25(1): p. 85-91.

**322.** **Georgescu, C., et al., *A gene regulatory network armature for T lymphocyte specification.* Proc Natl Acad Sci U S A, 2008. 105(51): p. 20100-5.**

**323.** **Zhao, F., et al., *TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies.* Nucleic Acids Res, 2005. 33(Database issue): p. D103-7.**

**324.** **Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. 13(11): p. 2498-504.**

**325.** **Camon, E., et al., *The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology.* Nucleic Acids Res, 2004. 32(Database issue): p. D262-6.**

**326.** **JMP9, 2010, SAS Institute Inc.: Cary, NC.**

**327.** **Schmitt, W.A., Jr., R.M. Raab, and G. Stephanopoulos, *Elucidation of gene interaction networks through time-lagged correlation analysis of transcriptional data.* Genome Res, 2004. 14(8): p. 1654-63.**

**328.** **Ji, L. and K.L. Tan, *Identifying time-lagged gene clusters using gene expression data.* Bioinformatics, 2005. 21(4): p. 509-16.**

**329.** **Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. 75(5): p. 843-54.**

# VITA

Sudhir Naswa received his Masters in Biotechnology from School of Biotechnology, Banaras Hindu University, India. Thereafter, he served Government of India in an administrative capacity. Following his interest in computational biology he joined University of Skovde, Sweden in 2003 to pursue Masters in Bioinformatics. Sudhir joined University of Tennessee in 2005 and earned Ph.D. from graduate school of Genome Science and Technology. Sudhir also completed Masters in statistics while pursuing his doctoral studies. Sudhir is married to Babita and is blessed with a wonderful daughter Shachi.