IN SILICO ANALYSIS OF MUTANT p53(R249S) ONCOGENICITY IN HEPATOCELLULAR CARCINOMA

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND THE INSTITUTE OF ENGINEERING AND SCIENCE OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

By
Guvanchmurad Ovezmuradov
September, 2007

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Rengül Çetin-Atalay (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Özlen Konu

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Volkan Atalay

Approved for the Institute of Engineering and Science:

Prof. Dr. Mehmet B. Baray
Director of the Institute of Engineering and Science

ABSTRACT

IN SILICO ANALYSIS OF MUTANT p53(R249S) ONCOGENICITY IN HEPATOCELLULAR CARCINOMA

Guvanchmurad Ovezmuradov
M.S. in Molecular Biology and Genetics
Supervisor: Assist. Prof. Dr. Rengül Çetin-Atalay
September, 2007

Oncogenic properties of mutant p53 proteins still stand as an ill-known subject, and the mechanism responsible for this phenomenon remains to be uncovered. This thesis aims to uncover the effect of p53 codon R249S ((AGG→AGT, arginine to serine) mutation on the development of hepatocellular carcinoma (HCC) through high throughput transcriptomics analysis using oligonucleotide arrays. We compared the expression profiles of HepG2 cells carrying wt and mutant p53(R249S). Microarray data analysis revealed a molecular signature consisting of 84 differentially regulated genes, showing that the expression of mutant p53(R249S) in HepG2 cells resulted in a distinct expression profile. Furthermore, mapping these significant differentially-expressed genes to the p53 interaction network revealed a putative interaction network representing functional outcomes of p53(R249S) expression in the context of diverse molecular interactions. Our results clearly demonstrated that several Hepatocyte Nuclear Factors (HNF1A, HNF4A and HNF6) could play an essential role in mediating mutant p53 oncogenic activity in HCC, as the key molecules of the gene network.

Keywords: p53, hepatocellular carcinoma, microarray, gene network, bioinformatics.

ÖZET

HEPATOSELLÜLER KARSINOMADAKİ MUTANT p53(R249S) ONCOJENİSİTENİN İN SİLİCO ANALİZİ

Guvanchmurad Ovezmuradov
Moleküler Biyoloji ve Genetik, Yüksek Lisans
Tez Yöneticisi: Yard. Doç. Dr. Rengül Çetin-Atalay
Eylül, 2007

Mutant p53 proteinlerinin onkojenik özellikleri hala az bilinenen bir konudur ve bu olguyu sağlayan mekanizma hala çözülmüş değildir. Bu tezin amacı, oligonukleotid dizinlerinin kullanıldığı geniş ölçekli transkriptomik analizini yaparak, p53'ün 249. kodon (AGG→AGT, arjininden serine) mutasyonunun hepatosellüler karsinomaya (HCC) olan etkisini ortaya çıkarmak. Wild-type ve mutant p53(R249S) taşıyan HepG2 hücrelerinin ifade şekli karşılaştırıldı. Mikrodizin veri analizi sonucu ifadesi değişen 84 genden oluşan ve bir "moleküler imza" niteliğını taşıyan bir ifade değişikliği açığa çıkarılarak, mutant p53(R249S) ifadesinin Hep G2 hücrelerinde tamamen ayrı bir gen ifade şekline sebep olduğu gösterildi. Ayrıca, bu 84 genin p53 etkileşim ağına eşlestirilmesi sonucu p53(R249S) ifadesinin işlevsel sonuçlarını degişik moleküler etkileşimler bağlamında açıklayan varsayımlı bir etkileşim ağı ortaya cıkarıldı. Bu çalışmayla elde edilen sonuçlarla, birkaç Hepatocyte Nuclear Factor'ün (HNF1A, HNF4A and HF6) ilgili gen ağının kilit molekülleri olarak mutant p53'ün HCC'deki onkojenik aktivitesinin sağlanmasında önemli roller üstleniyor olabildikleri gösterildi.

Anahtar sözcükler: p53, hepatosellüler karsinoma, mikrodizin, gen ağı, biyoenformatik.

ACKNOWLEDGEMENT

My heartfelt thanks and endless gratitude to Dr. Rengül Çetin-Atalay: Her unshakeable belief in me and brilliant guidance are the key factors underlying this work.

My special thanks to Prof. Mehmet Özturk, for his invaluable advices and motivation.

My exclusive thanks to Prof. Volkan Atalay, for his crucial help during this work.

I am eternally grateful to all faculty members of our department for providing excellent education and fascinating academic medium. I am especially indebted to Dr. Özlen Konu, Dr. Tamer Yağcı and Dr. Ali Güre for their kindly support and encouragement.

I deeply appreciate assistance of Özge and Haluk Yüzügüllü, Dr. Esra Önen, Ebru Bilget, Sevgi Bağışlar and Şerif Şentürk.

My special thanks to my friends Fuat Yağcı, Ender Avcı, Koray Kaya, Şafak Çağlayan, Reshad Mamedov and Ertuğrul Dalkıç.

My best regards to all Bilkent MBG members.

Finally, of course, my infinite acknowledgement to my beloved family.

LIST OF ABBREVIATIONS

AFB1	Aflatoxin B1
DNA	Deoxyribonucleic Acid
FDR	False Discovery Rate
GO	Gene Ontology
GOF	Gain of Function
HCC	Hepatocellular Carcinoma
HNF	Hepatocyte Nuclear Factor
mRNA	Messenger RNA
p53	Tumor Protein 53
RNA	Ribonucleic Acid
R249S	249th codon, Arginine to Serine Mutation
SAM	Significance Analysis of Microarrays
TP53	Tumor Protein 53 gene

CONTENTS

Chapter 1: Introduction 1

1.1. p53: A Two-faced Cancer Gene……………………………………………………1

1.1.1. p53 History……………………………………………………………......1

1.1.2. p53 as a Guardian of the Genome………………………………………...2

1.1.3. Is Mutant p53 an Oncogenic Protein? 2

1.1.4. Hot Spot Mutations of p53……………………………………………......5

1.1.5. Role of p53 Mutations in Tumorigenesis…………………………………5

1.1.6. Mechanisms of Transcriptional Regulation by Mutant p53………………6

1.2. p53 R249S Hot-spot Mutation and Hepatocellular Carcinoma……………….....8

1.3. Microarray-based Cancer Research and Bioinformatics ………………………10

1.4. Gene Networks Analysis ………………………………………………………10

1.5. Integrated Analysis of Genomic Data …………………………………………12

Chapter 2: Aim and Approach 14

Chapter 3: Materials and Methods 16

3.1. Microarray Data Analysis……………………………………………………...16

3.1.1. Microarray Experiment…………………………………………………16

3.1.2. Normalization of Raw Data……………………………………………...16

3.1.3. Test of Differential Expression (Significance Testing)………………….16

3.1.4. Data Mining Using Functional Annotation Tools……………………….16

3.1.5. Hierarchical Clustering…………………………………………………..16

3.2. Gene Network Analysis……………………………………………………….**17**

3.2.1. Mapping Significant Genes to Human BIND Network…………………17

3.2.2. Integrating CXX1 and HNF4A to the Network…………………………17

3.2.3. Integrating Differential Expression with the Network…………………..17

3.2.4. GO Annotation of the Network…………………………………………17

3.2.5. Alternative Layouts of the Network……………………………………..18

Chapter 4: Results 19

4.1. Microarray Data Analysis……………………………………………………...19

4.1.1. Microarray Experiment…………………………………………………19

4.1.2. Normalization of Raw Data……………………………………………...19

4.1.3. Test of Differential Expression (Significance Testing)………………….19

4.1.4. Data Mining Using Functional Annotation Tools……………………….20

4.1.5. Hierarchical Clustering…………………………………………………..26
4.2. Gene Network Analysis……………………………………………………..28
4.2.1. Mapping Significant Genes to Human BIND Network…………………28
4.2.2. Integrating CXX1 and HNF4A to the Network…………………………28
4.2.3. Integrating Differential Expression with the Network…………………..28
4.2.4. GO Annotation of the Network………………………………………….30
4.2.5. Alternative Layouts of the Network……………………………………..31
Chapter 4: Discussion **33**
5.1. Discussion of the Results from Microarray Data Analysis ……………………33
5.2. Discussion of the Results from Gene Network Analysis………………………34
5.3. Conclusion and Future Perspectives…………………………………………...36
Bibliography **38**
Appendix **43**

LIST OF FIGURES

1.1. Diagrammatic illustration of the history of p53 functions since its discovery in 1979……………………………………………………………………………........1
1.2. High frequency of missense mutations affecting p53 compared to other tumor suppressors……………………………………………………………….........3
1.3. p53: a two-faced cancer gene..……………………………………………………4
1.4. Distribution of p53 mutations…………………………………………………….5
1.5. Proposed mechanisms for the role of p53 mutations in tumorigenesis…………...6
1.6. Models for mutant p53 transcriptional activity…………………………………...7
1.7. Multistage hepatocarcinogenesis……………………………………………….....8
1.8. An example of a biochemical network…………………………………………..11
1.9. Overview of integrated analysis of genomic data……………………….........12
2.1. The workflow of the thesis…………………………………………...........15
4.1. Normalization of raw expression values………………………………………...19
4.2. Significance Analysis of Microarrays……………………………….............19
4.3. Bar chart of the tissue expression pattern……………………………..............23
4.4. Chromosome distribution chart……………………………………….............23
4.5. GO Molecular Function distribution chart…………………………...............24
4.6. GO Biological Process distribution chart………………………………………..24
4.7. GO Cellular Component distribution chart……………………………………...25
4.8. GO Molecular Function distribution flat pie chart………………………………25
4.9. GO Biological Function distribution flat pie chart………………………………25
4.10. GO Cellular Component distribution flat pie chart……………………………26
4.11. Dendogram demonstrating hierarchical clustering of significant genes……….27
4.12. Interaction network showing relationship between p53 and our significant genes……………………………………………………………………………29
4.13. GO annotation of the network………………………………………………….30
4.14. Subcellular localization layout of the network…………………………………31
4.15. Hierarchical layout of the network……………………………………………..32
5.1. Control of liver gene expression by HNF transcription factors...........................35

LIST OF TABLES

1.1. Hypothesis: dietary AFB1 exposure can cause 249ser (AGG-AGT) TP53 mutations during human liver carcinogenesis……………………………..............9

4.1. List of upregulated significant genes…..20

4.2. List of downregulated significant genes………………………………………22

4.3. Schematic representation of gene network analysis of significant genes……….28

A.1. Delta table……………………………………………………………………….43

A.2. Upregulated significant probesets…………………………………................44

A.3. Downregulated significant probesets……………………………………….....45

A.4. Input data of subcellular localization labels for Cerebral plug-in………………46

CHAPTER 1: INTRODUCTION

1.1. p53: A Two-faced Cancer Gene

1.1.1. p53 History

Tumor protein 53, hereafter to be referred as p53, is one of the most extensively studied genes in molecular biology. Since its discovery in 1979, there has been a remarkable change in depicting the role of p53 in tumorigenesis [1] [2] [Fig. 1.1]. Intriguingly, p53 was initially described as an oncogene because of its higher expression in tumor cells, profound promoting effect on immortalizing certain cell types and ability to assist cellular transformation [1]. Surprisingly, 10 years after the identification of p53, it was realized that previous studies unknowingly utilized mutant forms of p53, and therefore all corresponding findings were related to mutant protein [3]. Moreover subsequent intensive studies revealed that the actual wild-type p53 gene is a tumor-suppressor gene, making it one of the most intensively studied human cancer genes [1].

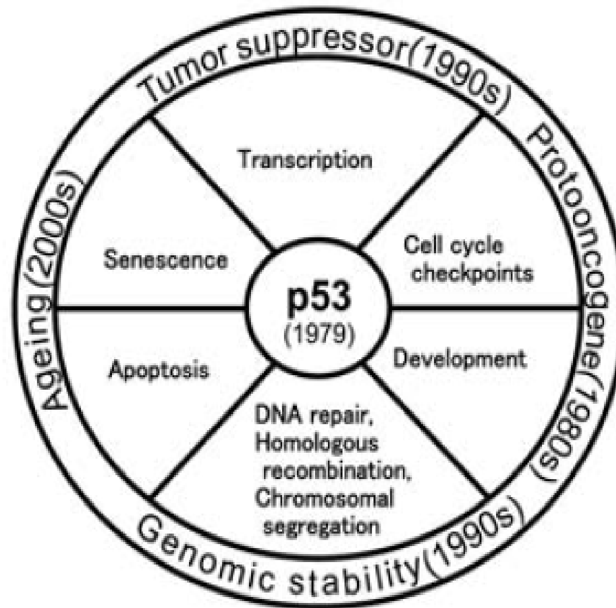


Figure 1.1: Diagrammatic illustration of the history of p53 functions since its discovery in 1979. (Hussain SP and Harris CC, 2006) [2]

Despite the fact that the most studies on p53 have concentrated on addressing the tumor suppressor functions of wild-type p53, the research on mutant p53 as an oncogenic protein still goes on yielding significant results [1].

1.1.2. p53 as a Guardian of the Genome

Recent advancements in molecular biology have revealed that close regulation of biological processes on cellular level is indispensable for life [4]. Any abnormal molecular condition is sensed by complex surveillance mechanisms and cell fate decision is made resulting in its rectification, to cell death, or to disease such as in cancer if this is not possible [4]. p53 plays an essential role as the master regulator of these events, and thus has been widely regarded as "the guardian of genome" [5]. Accordingly, TP53, the gene encoding p53, is considered to be one of the most essential genes in preventing cancer, and has been investigated intensively for more than twenty years [3]. These growing studies resulted in recognition of tumor suppressor p53 as a key element of the cell's antiproliferation machinery, accomplishing its effect by inducing either cell cycle arrest or apoptosis in response to various stress conditions [3]. These observations underscore the importance of p53 in tumor suppression and explain why p53 is prominent as the most frequently mutated gene (observed in half of all cancers) among human cancer genes [3]. Such a high frequency of mutations suggests a strong selective pressure for disruption of normal p53 activity in the process of tumorigenesis [6]. Accordingly, it's obvious that the disruption of wild-type p53 activity is vitally important for tumorigenesis [7]. Interestingly, beside of being such common as somatic mutations in human malignancies, alterations of the TP53 gene are also prominent as germline mutations in some cancer-prone families with Li-Fraumeni syndrome [8].

On the other hand, recent findings in this area revealed a broad spectrum of wild-type p53 activities including maintaining the genomic integrity (as a "Guardian of the genome"), transcription, cell cycle, apoptosis, senescence, DNA repair and development [2] [Fig. 1.1]. These findings clearly indicate that wild-type p53 acts as a key cell-growth regulator and tumor suppressor protein [9] at the crossroads of multiple cell signalling pathways.

1.1.3. Is Mutant p53 an Oncogenic Protein?

Efforts to decipher the oncogenic properties of mutant p53 proteins have yielded a considerable amount of elucidative results. The evidences supporting the idea of classifying mutant p53 as an oncogenic protein can be summarized in three parts:

1. High frequency of missense mutations: Overwhelming majority (74%) of p53 mutations are missense mutations, resulting in full-length, though mutant, proteins [10]. This

frequency of missense mutations is noticeably much higher when compared to other tumor suppressor genes [10], as the mutations striking other tumor suppressor genes (like APC, ATM and BRACA) in the majority of cases are reading-frame shifts or nonsense mutations [11] [12] [Fig. 1.2]. As a consequence, while mutations of other tumor suppressor genes result in truncated proteins, often degraded rapidly in cell, p53 mutations result in slightly altered, albeit stable proteins [11] [12]. This striking observation led scientists to the inescapable conclusion that mutant p53 proteins contribute to tumorigenesis by conferring selective advantage to cells harbouring them [10], as these cells can benefit from the presence of a slightly altered p53 protein rather than from its complete absence [11]. Consequently, it can be deduced that mutations in the p53 gene gain oncogenic functions to its protein products (oncogenic "Gain of function", GOF), besides destroying the tumor suppressor function of the wild-type protein [13].

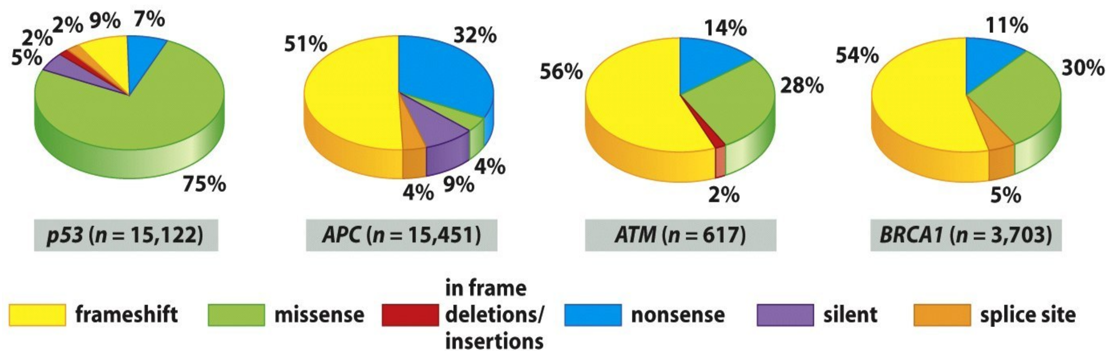


Figure 1.2: High frequency of missense mutations affecting p53 compared to other tumor suppressors. (Weinberg RA, 2006) [11]

2. Accumulation in tumor cells: Mutations of p53 result to the accumulation of high levels of mutant p53 in tumor cells [12]. This happens because only mutant p53, but not its wild-type form is stable in the nucleus of tumor cells [12] Observation that the levels of p53 protein in tumour cells are significantly higher than p53 levels in normal cells [1] strongly suggests that these mutant proteins are selectively overexpressed because of their vital oncogenic role in tumor progression [13].

3. Oncogenic activity in tumor cells: Since its early discovery, the research on p53 has revealed a solid knowledge that mutant p53 proteins have oncogenic properties which contribute to the establishment of malignant phenotype [14]. Early studies identified that consequences of expressing mutant p53 is not equivalent to the simple loss of wild-type p53: mutant p53 expression is capable of immortalizing certain cell types and of cooperating with

other oncogenes in cellular transformation [1] [7]. In recent years, a growing number of studies have demonstrated both *in-vitro* [15] [16] and *in-vivo* [17] [18] that expression of a mutant p53 can generate a broad variety of new oncogenic functions which enhance oncogenic potential of cells that express these proteins [14] [7] [Fig. 1.3]. Most remarkable are *in-vivo* studies [17] [18] utilizing mouse models in which mutant p53 expression is strongly correlated with a change in tumor spectrum in addition to enhancement of metastatic potential compared to p53-null mice [7]. All of these sophisticated studies have provided a compelling evidence of mutant p53 oncogenic activity in tumor cells [15].

Taking in account all of the evidences provided above, it's obvious that p53 doesn't fully obey the Knudson's two-hit model [19] of how tumor suppressor genes should operate [11]. Actually, since p53 is simultaneously both a tumor suppressor gene and an oncogene, it can truthfully be regarded as a two-faced cancer gene [14] [Fig. 1.3].

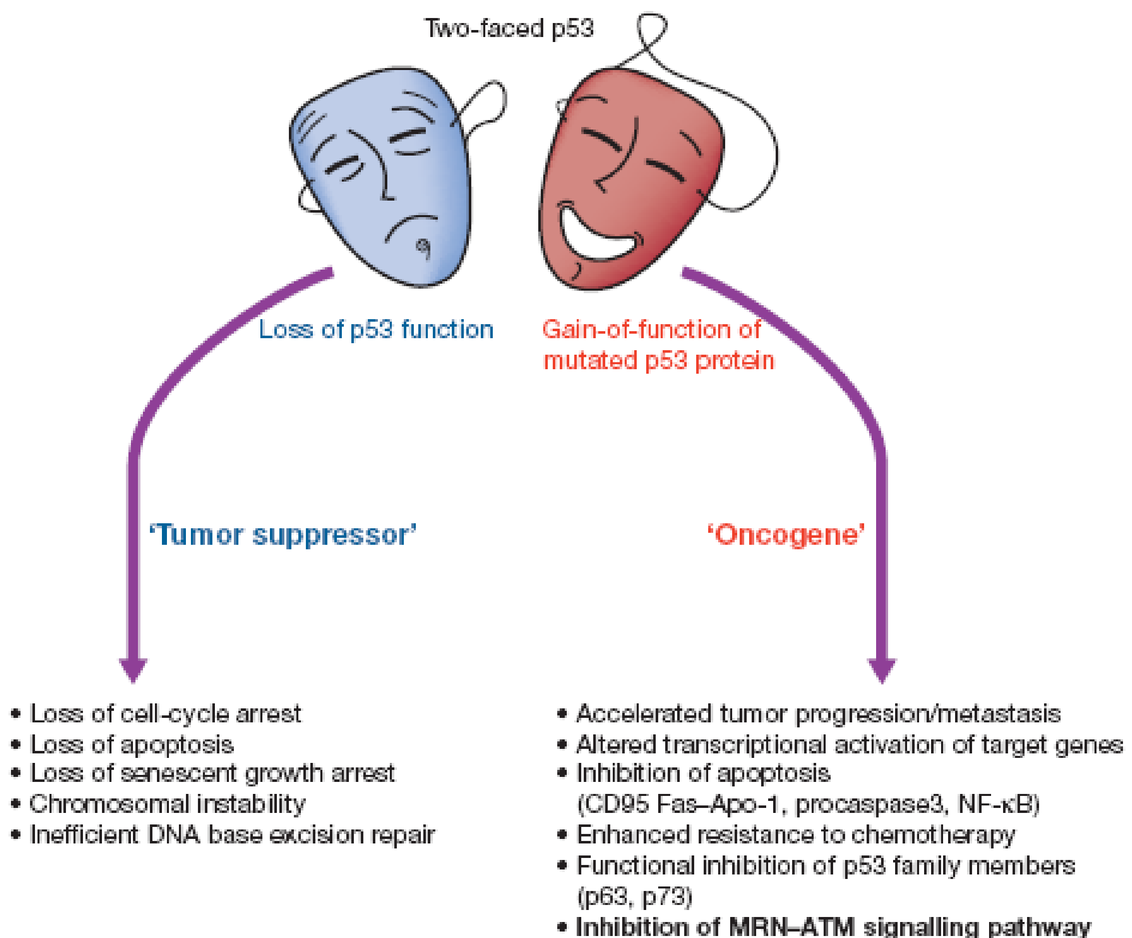


Figure 1.3: p53: a two-faced cancer gene. (Kastan MB and Berkovich E, 2007) [14]

1.1.4. Hot Spot Mutations of p53

Noticeably, overwhelming majority (>90%) of missense mutations in p53 affect the sequence-specific DNA-binding domain (DBD) of the protein, whereas more than 40% of all missense mutations alter codons R175, G248, R248, R249, R273 and R282 located within this domain [20] [Fig. 1.3]. Mutations in these six codons are the subjects in the majority of p53 studies and are commonly referred as hot-spot p53 mutations [21].

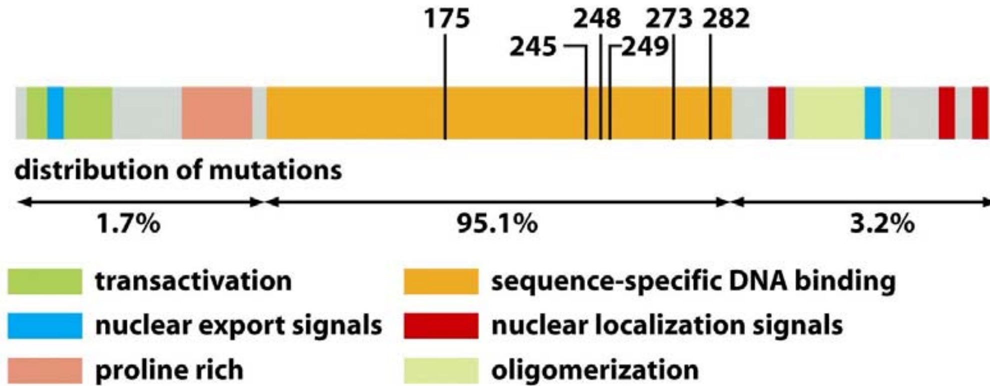


Figure 1.4: Distribution of p53 mutations. (Weinberg RA, 2006) [11]

1.1.5. Role of p53 Mutations in Tumorigenesis

Though extensive studies in this area, the exact mechanism responsible for mutant p53 contribution to tumorigenesis is not yet well defined [3]. The most promising model implies a paradigm of triple oncogenic effect caused by p53 mutations: *(a)* loss of function effect, where the tumor suppressor function of p53 is disrupted *(b)*; dominant negative effect, where wild-type p53 function is inactivated as result of hetero-oligomeric complex formation between wild-type and mutant p53 proteins, and *(c)* gain of function (dominant-positive) effect, where mutant p53 acquires novel oncogenic functions not seen in wild-type p53 that are independent of complex formation with wild-type p53 and therefore may occur in the absence of second (wild-type) p53 allele [3] [13] [22] [Fig. 1.4].

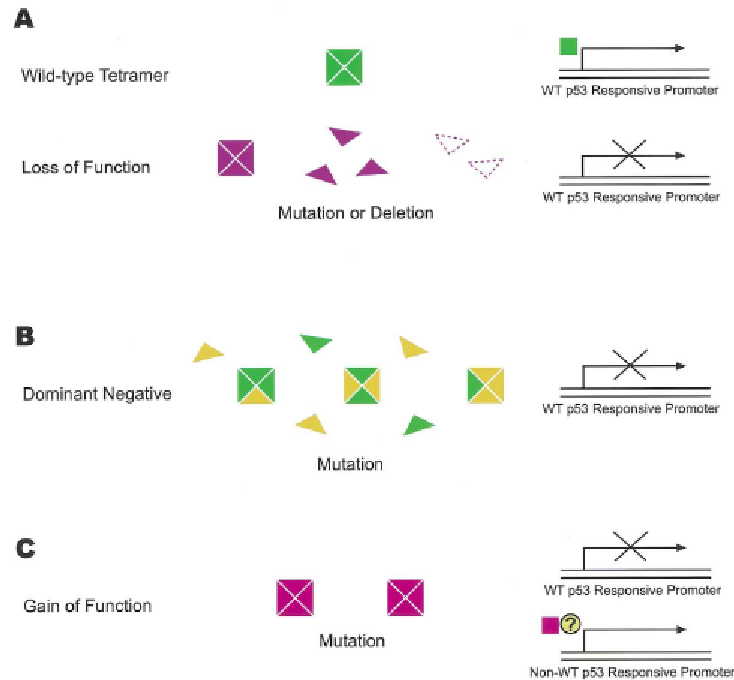


Figure 1.5: Proposed mechanisms for the role of p53 mutations in tumorigenesis.
(Cadwell C and Zambetti GP, 2001) [3]

The gain of function hypothesis for mutant p53 has been tested in cells devoid of endogenous p53 [18] [23]. Since there was no endogenous wild-type p53 activity for the mutant p53 to interfere with, it was concluded that mutant p53 must have been directly causing the stimulation of unregulated cell growth [13]. Last but not least, mutant p53 gain of function can also be linked to physical interaction of mutant p53 with other p53 family members, p73 and p63 [13]. Although wild-type p53 shows no obvious ability to bind p63 or p73, mutation can provide "gain of this ability" [7]. Indeed, recent studies have confirmed these interactions for a subset of p53 mutants and demonstrated subsequent functional outcomes: mutant p53 binding to its sibling proteins results in their inhibition [7] [21].

1.1.6. Mechanisms of Transcriptional Regulation by Mutant p53

Ability to regulate gene expression and modulate the transcriptome of the mutant cell is considered as one of the major mechanisms underlying mutant p53 GOF [1] [24]. Indeed, the series of studies have demonstrated the ability of mutant p53 to turn on and off specific sets of genes through acting as a transcription factor [25]. Noticeably, these genes are not regulated by wild-type p53 and none of them contain wild-type p53 DNA biding consensus site [3] [26]. Moreover, the repertoire of mutant p53-responsive genes is significantly distinct from that of wild-type p53 [26]. This implies that the alteration in target gene specificity is

what really responsible for mutant p53 GOF [26]. Given this, the meaningful question arises: How the specificity of mutant p53-mediated transcriptional regulation is achieved? [26]. As a consequence, it remains a challenging task to elucidate at the molecular level the mode of this transcriptional regulation [1] [24].

Currently, it seems to be a consensus on two molecular scenarios explaining function of mutant p53 as an oncogenic transcription factor [25] [Fig. 1.5]. While the first model depends on altered protein-DNA interactions of p53, the second one relies on its altered protein-protein interactions.

First model, based on "direct binding" [27], presumes direct binding of mutant p53 to the target DNA sequences, through yet unknown mechanisms which involve intrinsic DNA binding activities of mutant p53 proteins themselves [25]. The issue of sequence-specific binding is still under debate: since different mutant p53-responsive promoters show no sequence homology, the linear DNA sequence motif serving as a mutant p53-specific binding site couldn't have been defined so far [1] [26].

According to the second model, based on "passive targeting" [27], mutant p53 can be recruited to its promoters in a specific manner indirectly and independently from the presence of canonical p53 binding site [1] [28] [26] [25]. This targeting is possible through physical interaction of mutant p53 with other sequence-specific transcription factors, such as Ets-1, SP-1 and NF-Y [1] [28]. The fact that both mutant p53 and these transcription factors are direct transcriptional regulators of the set of common genes further supports this notion [26] [27]. Consequently, being a member of a transcriptional protein complex enables tethering of mutant p53 to its promoter regions [25].

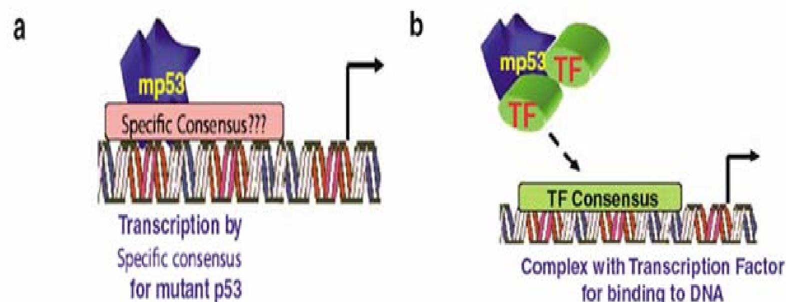


Figure 1.6: Models for mutant p53 transcriptional activity. (a) Mutant p53 binds the regulatory regions of its target genes through a specific and yet unknown DNA-binding consensus sequence; (b) mutant p53 interacts with a specific transcription factor that drives its gene target specificity. (Strano S *et al*, 2007) [25]

From the structural view, the two models can be explained through "cause-result" relationship. The structural changes resulting as response to mutation, determine three important properties of mutant p53 protein: (i) its folding state; (ii) its affinity for a range of target promoters; (iii) its affinity to others proteins [29]. While the second one explains the alteration of target DNA selectivity, the third one sheds light on the variation in protein interactions. Therefore, it is apparent that while the change in p53's affinity to DNA sequences can clarify the first model, the change in affinity to proteins may be responsible for the second.

1.2. p53 R249S Hot-spot Mutation and Hepatocellular Carcinoma

Hepatocellular carcinoma (HCC) is the most common liver cancer [30], accounting for an estimated 600 000 deaths annually [31]. Moreover, HCC is one of the most widespread malignancies worldwide [32], standing among the five leading causes of cancer death in the world [30]. Chronic infections with hepatitis B or C virus and consumption of dietary aflatoxin B1 (AFB1) or alcohol are recognized as the foremost risk factors [32] [Fig. 1.7].

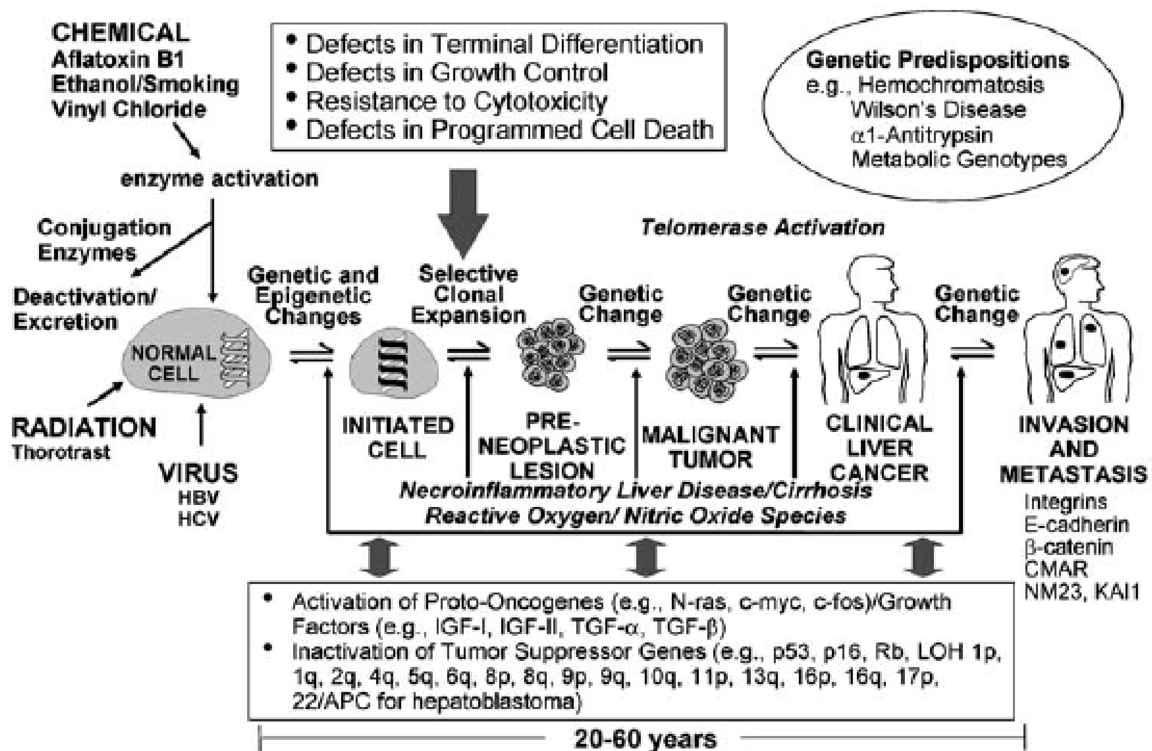


Figure 1.7: Multistage hepatocarcinogenesis. (Hussain SP *et al*, 2007) [32]

From clinical perspective, local surgical resection or liver transplantation constitute the only "curative treatment" [30] [33] for HCC patients. The fact that majority of patients have

already widespread HCC tumors at the time of diagnosis further complicates their surgery [30]. Moreover, recurrence is observed in the half of patients with localized HCC tumors who undergo surgical treatment [30]. Such a severity of HCC, inefficiency of treatment methods, and the absence of effective diagnostic markers have turned this disease into one of the most critical and challenging areas in cancer biology [31] [33].

The molecular pathogenesis of HCC, which involves multiple genetic and epigenetic changes, still remains largely unknown [32] [31] [34]. Current understanding of hepatocarcinogenesis identifies it as a multistage process accompanied by accumulation of abundant genetic alterations, like the mutation in p53 [32] [35] [Fig. 1.7]. Aflatoxin B1 (AFB1) plays a causative role in the process of hepatocarcinogenesis as a major chemical carcinogen [35] [32].

Table 1.1: Hypothesis: dietary AFB1 exposure can cause 249ser (AGG-AGT) TP53 mutations during human liver carcinogenesis. (Staib *et al.*, 2003 and Hussain SP *et al.*, 2007) [32]

Strength of association	*Biological plausibility*
• *Consistency* ○ Positive dose–response correlation between estimated dietary AFB_1 exposure and frequency of 249^{ser} TP53 mutations in three different ethnic populations on three continents. ○ 249^{ser} TP53 mutant DNA is detected in sera from individuals exposed to AFB_1 and infected with HBV. ○ 249^{ser} TP53 mutations are found in HCC from individuals both exposed to dietary AFB_1 and infection with HBV, but not with HBV alone. • *Specificity* ○ 249^{ser} TP53 mutations are uncommon in other cancer types. ○ 249^{ser} TP53 mutation in serum and plasma is a biomarker of liver cancer risk. • *Temporality* ○ 249^{ser} TP53 mutant cells are observed in nontumorous liver in high HCC incidence geographic areas.	• AFB_1 is a potent mutagen and carcinogen in laboratory studies. • AFB_1 is enzymatically activated by human hepatocytes and the 8,9-AFB_1 oxide binds to the third base (G) in codon 249. • AFB_1 exposure to human liver cells *in vitro* produces codon 249^{ser} TP53 mutations. • HBx gene expression increases the frequency of 249^{ser} TP53 mutations in cells exposed to AFB_1 *in vitro*. • 249^{ser} TP53 expression inhibits apoptosis and p53-mediated transcription, and enhances liver cell growth *in vitro*.

Abbreviations: AFB_1, aflatoxin B_1; HBV, hepatitis B virus; HBx, X gene of HBV; HCC, hepatocellular carcinoma; TP53, *p53* tumor suppressor gene.

AFB1 is a very potential mutagen inducing a hot-spot p53 mutation [Fig. 1.3] in the third position of 249th codon [32] [35] [Table 1.1]. The resultant G→T transversion leads to the amino-acid substitution R249S (arginine to serine), which is extremely specific to HCC [35] [36]. Accordingly, this hot-spot mutation predominantly occurs in patients with hepatocellular tumors from the region of eastern Asia and sub-Saharan Africa, where the AFB1 dietary intake occurs as a common food contaminant [29] [32].

Because of its remarkable specificity to HCC, R249S mutation is considered as one of the tumor-specific, ‘signature’ mutations of p53 [36]. This significant molecular link between the

exposure to a particular carcinogen and a specific type of cancer (as well as mutation) provides an elegant example of how environmental carcinogens can be implicated in the etiology of human cancers [32]. This remarkable correlation between AFB1 exposure and R249S p53 mutation can be due to at least two reasons [32]. While first explanation relies on the potential high mutability of the third base at 249th codon to AFB1, another one suggests that these R249S mutants may confer a unique growth and/or survival advantage to these liver cells, resulting in their selection in a tissue-specific manner [32].

1.3. Microarray-based Cancer Research and Bioinformatics

Performing global gene expression analysis became possible after development of expression microarrays [24]. Combined with advent of supporting bioinformatics tools, this innovation enabled for the first time a comprehensive analysis of cell transcriptome on genome wide-scale [24]. This high-throughput technology has been commonly exploited in a wide range of biological areas, such as study of cancer and neuroscience [37].

Defining molecular differences between cancerous and healthy cell is one of the major tasks in cancer biology [38]. Since microarray analysis enables tracking relative transcript levels during comparing different biological classes, it has proven to be invaluable in translational cancer research [39] [37]. Monitoring simultaneously the expression levels of numerous genes on an unbiased manner is promising to unravel the complicated gene-expression programs governing tumorigenesis [33] [40].

The goal of bioinformatics is to develop and present software programs for the use of biologists as an applicable tool in solving complicated biological problems. Since microarray technology is highly dependent on bioinformatics and biostatistics, a comprehensive understanding of the large-scale data derived from array-based experiments highly demands application of the relevant computational tools [41].

1.4. Gene Networks Analysis

It proved applicable to represent various biological datasets as “gene networks”, composed of multiple nodes (corresponding to genes or proteins) and connections (matching to physical interactions between these entities) [42]. In reality, these gene networks are a simplification of the ultimate biochemical network, which unequivocally includes all three interaction levels equivalent to three types of biological molecules (mRNA, proteins and metabolites) [43] [Fig. 1.8]. Therefore, network of interaction can be constructed on several

levels and can depict various interaction types [43]. But when the research is restrained to surveying gene expression, such as in microarray experiments, it's appropriate to limit interaction network with the representative gene network to explain the data [43].

Regulatory system of cell is a complex mechanism, involving various cell signaling mechanisms and regulatory machinery [38] [44]. Many signaling molecules are implicated in this process as participants of complicated signal transduction processes, commonly referred as cell signaling pathways [38]. Since signaling pathways never occur in isolation in cell, but function as members of large biomolecular networks [43], it became clear that signaling takes place through a regulatory network of interacting signaling pathways [38]. On the other hand, it's widely recognized that a coordinate response of a combination of genes is what responsible for most of the cellular behavior and phenotypes [43]. All these findings suggest that studying the complex architecture of signaling networks is thought to demonstrate how these complex biological traits arise and propagate [43] [38]. For the same reason, deciphering complicated regulatory program of cell through gene networks is a promising approach for combating complex diseases such as cancer [43].

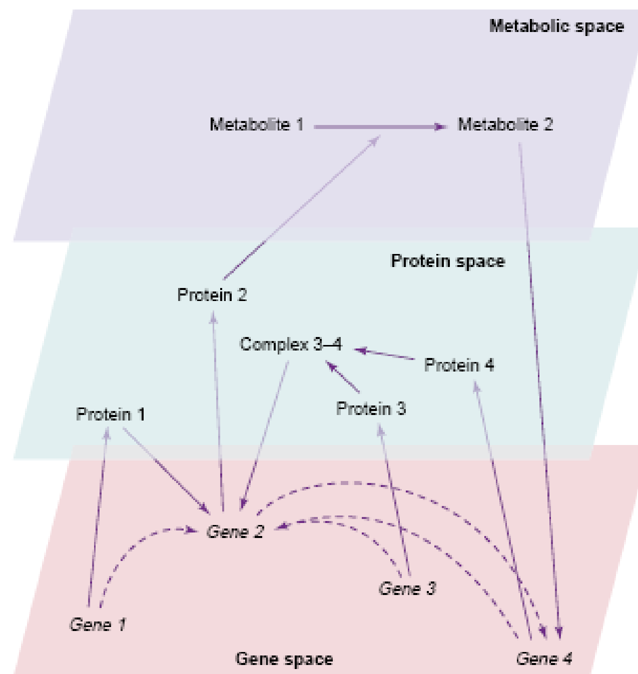


Fig. 1.8: An example of a biochemical network. Molecular constituents (nodes of the network) are organized in three levels (spaces): mRNAs, proteins, and metabolites. Solid arrows indicate interactions, the signs of which (activation or repression) are not specified in this diagram. Projections of these interactions into the 'gene space', indicated by dashed lines, constitute a corresponding gene network. (Brazhnik P *et al,* 2002) [43].

1.5. Integrated Analysis of Genomic Data

Microarray experiments currently stand as the major source for genomic high-throughput information [45]. Ultimate goal behind these experiments is to find out both differentially expressed genes and genes with similar expression pattern. The rationale of searching for similar expression patterns using clustering algorithms is that co-clustering genes are supposed to be functionally related to each other [43]. As a consequence, their products should preferentially interact with each other in order to execute common molecular functions [38]. From researchers' perspective, since genes of interacting proteins are predisposed to share similar expression patterns, it's reasonable to integrate both information sources in solving complex biological problems [45]. Indeed, linking the transcription pattern similarities of co-expressed genes to corresponding molecular interactions between their products has become one of the most appealing concepts of systems biology [45] [Fig. 1.9].

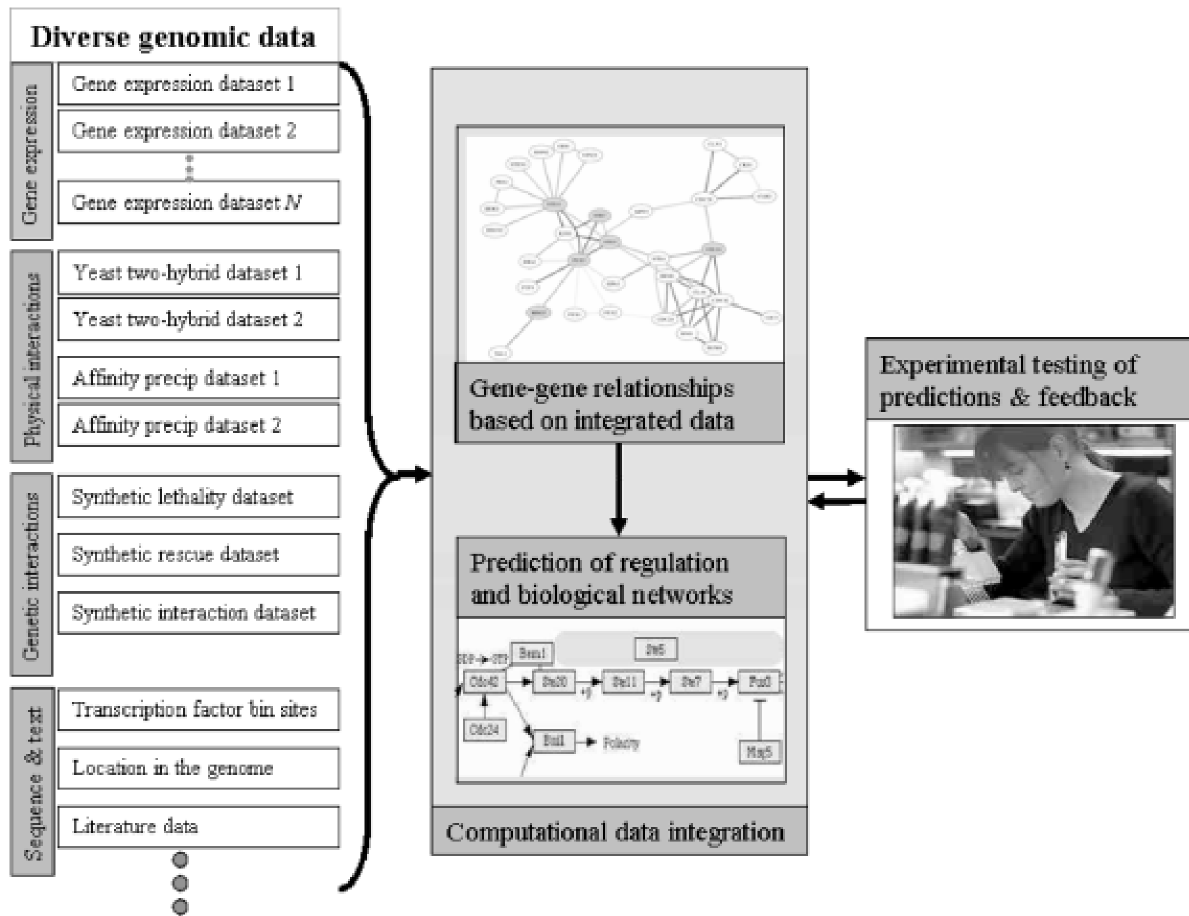


Fig. 1.9: Overview of integrated analysis of genomic data. (Troyanskaya OG, 2005) [49]

Applications have already demonstrated that analysis of the experimental data in the context of molecular interactions leads to better elucidation of interrelations among the discovered differentially expressed genes [46]. Progress of bioinformatics gave rise to numerous knowledge databases and computational tools that enable integrating massive high-throughput expression data with accumulating molecular interaction data into the united conceptual framework [47] [48] [49] [Fig. 1.9]. In conclusion, this integrative approach might provide valuable clues and lead to new ideas for comprehensive elucidation of multiple molecular mechanisms that govern cell behaviour.

CHAPTER 2: AIM AND APPROACH

Mutations in p53 are considered among the major cancer-causing genetic alterations in the process of carcinogenesis. In addition to loss of function of the p53 tumor suppressor, the resulting mutant p53 proteins contribute to the malignancies by enhancing tumorigenic properties of cells. Currently, the oncogenic properties of mutant p53 proteins still stand as an ill-known subject, and the mechanism responsible for this phenomenon remains to be uncovered. Investigating the role and the underlying mechanism of mutant p53 oncogenicity in the course of hepatocellular carcinoma (HCC) was the main objective of this thesis. p53 mutation is one of the most carcinogenic steps in development of HCC, but overall impact of this mutation on the gene regulatory networks of liver cells is not well understood. Since we were interested in what effect this mutation has on development of HCC, p53 249th codon (AGG→AGT, arginine to serine) mutation, which is specific to HCC, was selected for our research. The specific aim of the present study was to find out the list of differentially expressed genes and the associated gene network affected by the expression of p53(R249S) mutant proteins.

Searching for genes that change expression in response to mutant p53 expression was a promising approach to unravel the mechanism underlying mutant p53 oncogenicity in HCC. Remarkably, it was the first time when genome-wide gene expression profiling was chosen as a means to discover a set of genes involved in this process. "Comparative genomic approach" using two isogenic HCC cell lines (HepG2 and its counterpart stably expressing p53(R249S) named HepG2-249.1) was selected as the experimental approach in our microarray experiment. Analysis of the raw data derived from this experiment constitutes first part of this study.

There were different approaches until now to elucidate the nature of mutant p53 oncogenicity, most of which included gene expression analyses. However, this phenomenon has not been investigated yet in terms of gene networks, which confers additional originality to our approach. Mapping our significant differentially-expressed genes to the p53 interaction network and subsequent computational analysis of the resultant network represents second part of this study. It was strongly anticipated that interpretation of microarray data in context of diverse molecular interactions would lead to better elucidation of the interrelations among

the discovered differentially expressed genes and aid comprehensive cross-validation of our findings with the existing knowledge about the related molecular mechanisms.

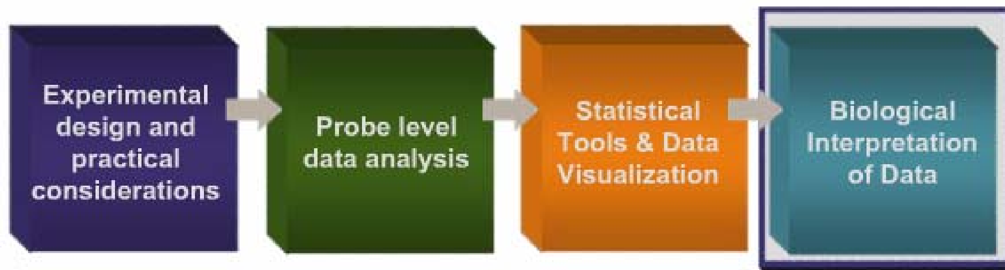


Figure 2.1: The workflow of the thesis. (Adopted from Affymetrix web site) [50]

CHAPTER 3: MATERIALS AND METHODS

3.1. Microarray Data Analysis

3.1.1. Microarray Experiment

"Comparative genomic approach" using two isogenic HCC cell lines (HepG2 and its counterpart stably expressing codon 249 mutant p53 named HepG2-249.1) was exploited as a model for the microarray experiment. In order to achieve higher fidelity of the results, microarray analysis was performed using quadruplets of total RNA samples extracted from these cell lines. Affymetrix "HGU95Av2" Gene Chip (screening with 12.625 probe sets) [50] was exploited as the microarray platform of this expression analysis.

3.1.2. Normalization of Raw Data

In silico analysis of acquired microarray data was made using R, which is a software environment for statistical computing and graphics. [51]. Quantile normalization [52] method was applied to normalize the raw expression data.

3.1.3. Test of Differential Expression (Significance Testing)

We applied SAM ("Significance Analysis of Microarrays"), which is the most popular statistical method used for significance analysis [53], in order to test the differential expression and identify significant genes. SAMR package [54] for R was utilized for this purpose.

3.1.4. Data Mining Using Functional Annotation Tools

Multiple annotation tools were used to interpret our microarray data results. NetAffx [55], WebGestalt [57], Onto-express [56] and Fatigo+ [58] were our top priority since these tools enable simultaneous input and analysis of multiple genes. List of our significant genes was given as an input for annotation analyses by these databases.

3.1.5. Hierarchical Clustering

Gene expression patterns observed in microarray experiments can be interpreted as indications of the status of cellular processes and may provide a further insight to the coexpressed genes of unknown function [59]. Hierarchical clustering of significant genes according to their gene expressions was performed using Cluster [59] and Java Treeview [60] software.

3.2. Gene Network Analysis

3.2.1. Mapping Significant Genes to Human BIND Network

Cytoscape [61] is open-source software for network visualization and analysis. Application of Cytoscape is most powerful when applied in combination with large databases of biomolecular interactions [61].

In order to analyse our findings on systems biology level, the list of significance genes was mapped onto BIND human protein interaction network [62] using Cytoscape software [61]. (BIND network is composed of experimentally proven biomolecular interactions (protein-protein and protein-DNA) [62]) Since we were interested in mutant p53 GOF, we also added p53 gene to the input list in order to discover the relationship between the significant genes and this gene. Consequently, the data set used for core network construction consisted of our significant genes, p53 and the molecules in the neighborhood.

3.2.2. Integrating CXX1 and HNF4A to the Network

The resulting network was expanded by integrating CXX1 gene to the core network. Interaction data regarding CXX1 was retrieved from MINT database [63]. Moreover, additional direct interaction between HNF4A and p53 was adopted from String database [64] and added to the core network.

3.2.3. Integrating Differential Expression with the Network

Cytoscape allows visual integration of biomolecular interaction networks with expression profiles derived from high-throughput expression data. Interactions of our resulting network were integrated with gene expression data obtained from microarray analysis. Visual Mapper feature of Cytoscape was used for this purpose,

3.2.4. GO Annotation of the Network

The software “Core” of Cytoscape has been extensively extended through development of numerous plug-ins, allowing application of additional computational analyses and features [65]. Using such plug-ins may facilitate linking the network to databases of functional annotations.

Interactions of our resulting network were integrated with Gene Ontology (GO) Biological Process data [66] using GOlorize plug-in [67]. All nodes of the network were clustered and colored according to their corresponding GO category.

3.2.5. Alternative Layouts of the Network

Cerebral (Cell Region-Based Rendering And Layout) plug-in [68] for Cytoscape enables the visual integration of the network with subcellular localization data. This plug-in was used to generate an alternative layout of the interaction network according to the subcellular localization of the participating molecules. Related subcellular localization data was retrieved from Entrez Gene [8], UniProt [69] and MEP2SL [70] databases [Table A.4; Appendix]. Apart from this, the hierarchical layout of Cytoscape interface was used to determine hierarchical architecture of the network.

CHAPTER 4: RESULTS

4.1. Microarray Data Analysis

4.1.1. Microarray Experiment

Microarray experiment generated raw expression data, which was the starting point for the subsequent *in silico* analyses.

4.1.2. Normalization of Raw Data

Raw expression data was normalized using Quantile normalization method [Fig. 4.1].

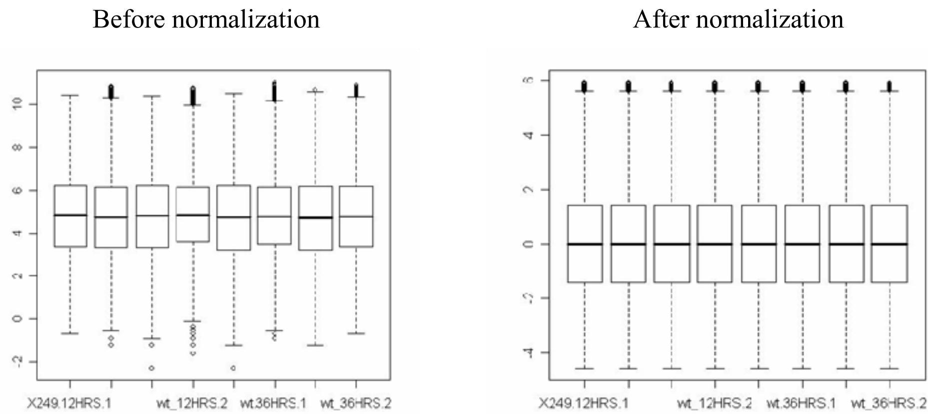


Figure 4.1: Normalization of raw expression values.

4.1.3. Test of Differential Expression (Significance Testing)

Significance analysis of microarrays was used to identify significant probesets [Fig. 4.2]. Calculated delta table [Table A.1; Appendix] was used to generate the list of significant probesets and FDR equal to 0.049 was chosen as the significance level. As result differentially expressed 110 probesets (FDR<0.05) were selected for further analysis [Table A.2-3; Appendix].

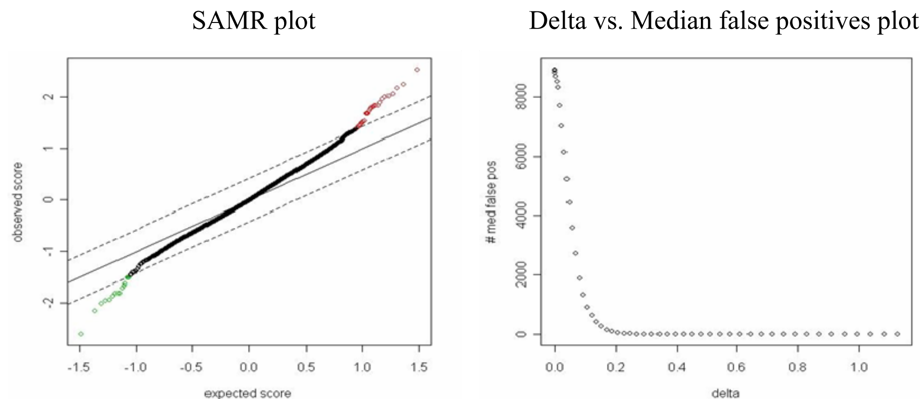


Figure 4.2: Significance Analysis of Microarrays.

4.1.4. Data Mining Using Functional Annotation Tools

Functional annotation of 110 significant probesets using NetAffx, WebGestalt and Babelomics databases showed that they correspond to 84 known genes (63 up- and 21 downregulated) of various functions and properties [Table 4.1 and 4.2].

Table 4.1: List of upregulated significant genes.

Order	Symbol	Gene Name	Fold Change
1	CXX1	CAAX box 1	23.22800715
2	CD9	CD9 molecule	10.66870392
3	TRIB2	tribbles homolog 2 (Drosophila)	10.73610878
4	SPINK1	serine peptidase inhibitor, Kazal type 1	6.836368805
5	FGL1	fibrinogen-like 1	6.910175987
6	SCGN	secretagogin, EF-hand calcium binding protein	7.394791124
7	TUBB2B	tubulin, beta 2B	6.631740161
8	IGFBP2	insulin-like growth factor binding protein 2, 36kDa	5.618615288
9	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	5.566360188
10	TRIB2	tribbles homolog 2 (Drosophila)	8.830691762
11	GAS7	growth arrest-specific 7	7.459127617
12	GPC3	glypican 3	4.385885356
13	CRIP1	cysteine-rich protein 1 (intestinal)	5.06460286
14	TFF3	trefoil factor 3 (intestinal)	5.646708948
15	CYP7A1	cytochrome P450, family 7, subfamily A, polypeptide 1	4.75273648
16	NT5E	5'-nucleotidase, ecto (CD73)	4.821628984
17	SEPT6	septin 6	3.892380553
18	IGFBP2	insulin-like growth factor binding protein 2, 36kDa	4.975579655
19	VTN	vitronectin	3.497558863
20	ATP9A	ATPase, Class II, type 9A	3.657102437
21	SALL1	sal-like 1 (Drosophila)	3.509610598
22	NFE2	nuclear factor (erythroid-derived 2), 45kDa	4.418474567
23	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	4.044187193
24	CDKL5	cyclin-dependent kinase-like 5	3.056986098
25	CD24	CD24 molecule	3.528537463
26	EMP2	epithelial membrane protein 2	3.292796742
27	DIP2C	DIP2 disco-interacting protein 2 homolog C (Drosophila)	3.680506015
28	TIMP2	TIMP metallopeptidase inhibitor 2	3.791424047
29	IGSF4	immunoglobulin superfamily, member 4	3.126117522
30	DPH4	DPH4 homolog (JJJ3, S. cerevisiae)	3.145138992

31	DIP2C	DIP2 disco-interacting protein 2 homolog C (Drosophila)	2.814455399
32	RASSF2	Ras association (RalGDS/AF-6) domain family 2	3.483605186
33	CAMK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma	3.20015185
34	MYO10	myosin X	6.26521794
35	GNMT	glycine N-methyltransferase	2.833020285
36	PLA2G1B	phospholipase A2, group IB (pancreas)	2.674845104
37	PDGFA	platelet-derived growth factor alpha polypeptide	4.366942565
38	LY6E	lymphocyte antigen 6 complex, locus E	2.812146331
39	ZNF185	zinc finger protein 185 (LIM domain)	3.302367067
40	ICAM2	intercellular adhesion molecule 2	3.290637749
41	GC	group-specific component (vitamin D binding protein)	3.457322198
42	ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1	2.574092795
43	OPHN1	oligophrenin 1	2.67282291
44	COMP	cartilage oligomeric matrix protein	3.24477113
45	AFM	afamin	4.403802207
46	FUT8	fucosyltransferase 8 (alpha (1,6) fucosyltransferase)	3.699118026
47	KNG1	kininogen 1	3.280106829
48	SP110	SP110 nuclear body protein	3.236990624
49	PGC	progastricsin (pepsinogen C)	2.410266847
50	ARMC8	armadillo repeat containing 8	2.645574822
51	SH3BGRL	SH3 domain binding glutamic acid-rich protein like	3.178723258
52	CITED2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	2.447360881
53	LRP3	low density lipoprotein receptor-related protein 3	4.569480394
54	ALDH3A1	aldehyde dehydrogenase 3 family, memberA1	2.987214461
55	TSPAN7	tetraspanin 7	2.315269713
56	AADAC	arylacetamide deacetylase (esterase)	2.531105015
57	CTSL2	cathepsin L2	3.491246648
58	PBXIP1	pre-B-cell leukemia transcription factor interacting protein 1	2.562377452
59	TNNI3	troponin I type 3 (cardiac)	3.109776838
60	FGA	fibrinogen alpha chain	2.486721324
61	RBP1	retinol binding protein 1, cellular	3.438083597
62	FRK	fyn-related kinase	3.62128114
63	PHYH	phytanoyl-CoA 2-hydroxylase	2.503697831
64	PVRL3	poliovirus receptor-related 3	2.363370451
65	AGTR1	angiotensin II receptor, type 1	2.97706393
66	KIAA0649	KIAA0649	2.935727094
67	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	2.840251507

Table 4.2: List of downregulated significant genes.

Order	Symbol	Gene Name	Fold Change
1	MYL9	myosin, light chain 9, regulatory	0.06911942
2	PLP2	proteolipid protein 2 (colonic epithelium-enriched)	0.12872334
3	PRAME	preferentially expressed antigen in melanoma	0.15274071
4	PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	0.16698822
5	GLT25D2	glycosyltransferase 25 domain containing 2	0.16702497
6	APOL1	apolipoprotein L, 1	0.18826925
7	PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	0.19549314
8	MFNG	manic fringe homolog (Drosophila)	0.20099952
9	DLK1	delta-like 1 homolog (Drosophila)	0.2233338
10	CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1	0.17863134
11	CPVL	carboxypeptidase, vitellogenic-like	0.2460938
12	ASNS	asparagine synthetase	0.24480766
13	ECGF1	endothelial cell growth factor 1 (platelet-derived)	0.24343787
14	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	0.27102712
15	ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5	0.24614508
16	CD22	CD22 molecule	0.27344226
17	ABLIM3	actin binding LIM protein family, member 3	0.32283383
18	EREG	epiregulin	0.26010424
19	IL18	interleukin 18 (interferon-gamma-inducing factor)	0.25988872
20	BICC1	bicaudal C homolog 1 (Drosophila)	0.30711096
21	MCC	mutated in colorectal cancers	0.28560867
22	MCC	mutated in colorectal cancers	0.34042488
23	TNFSF4	tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated glycoprotein 1, 34kDa)	0.29873043

It was remarkable that 78 out of 84 significant genes were specific to liver which could be due to the anticipated tissue specificity of p53 R249S mutation to HCC [Fig. 4.3].

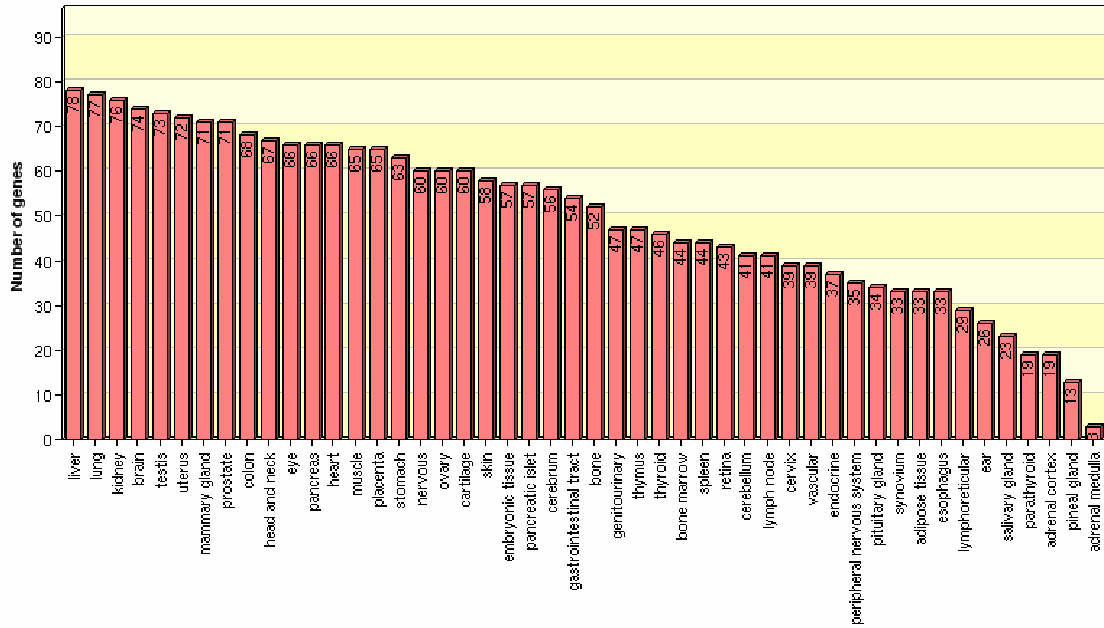


Figure 4.3: Bar chart of the tissue expression pattern. Each tissue is represented by a bar in the chart. The height of the bar represents the number of genes that are expressed in the tissue (from WebGestalt).

Significant genes were equally distributed in terms of chromosomal location.

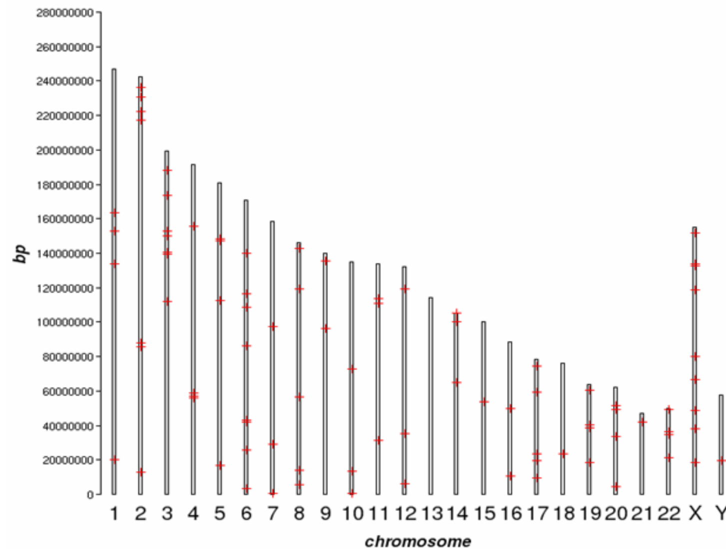


Figure 4.4: Chromosome distribution chart. Each chromosome is represented by a bar in the chart. Each gene is represented by a red cross symbol and located on the chromosome based on its location (from WebGestalt).

Classification of our significant genes according to molecular function demonstrated a noticeable high prevalence of “binding proteins” among them [Fig. 4.5 and 4.8]. This finding may reflect the possible role of these proteins in a signal transduction resulting from mutant p53 expression.

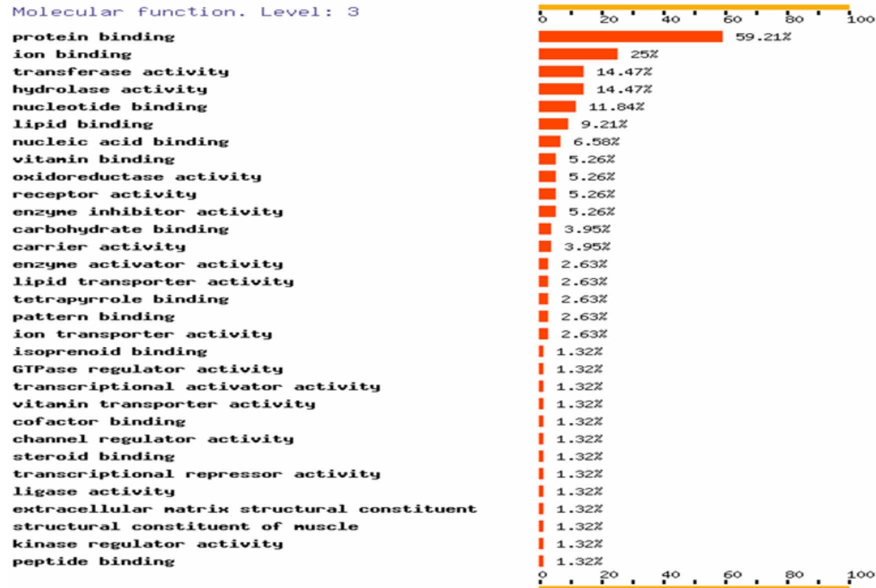


Figure 4.5: GO Molecular Function distribution chart (from Fatigo +).

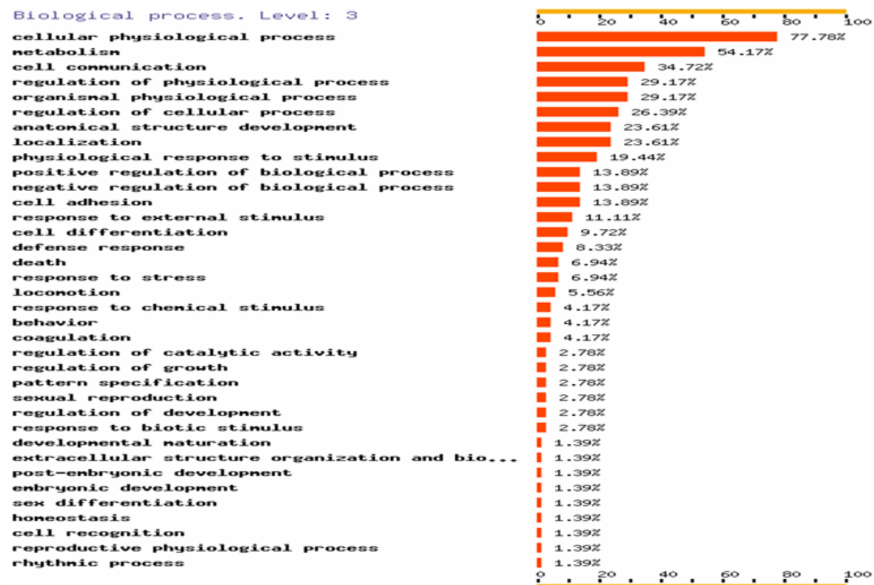


Figure 4.6: GO Biological Process distribution chart (from Fatigo +).

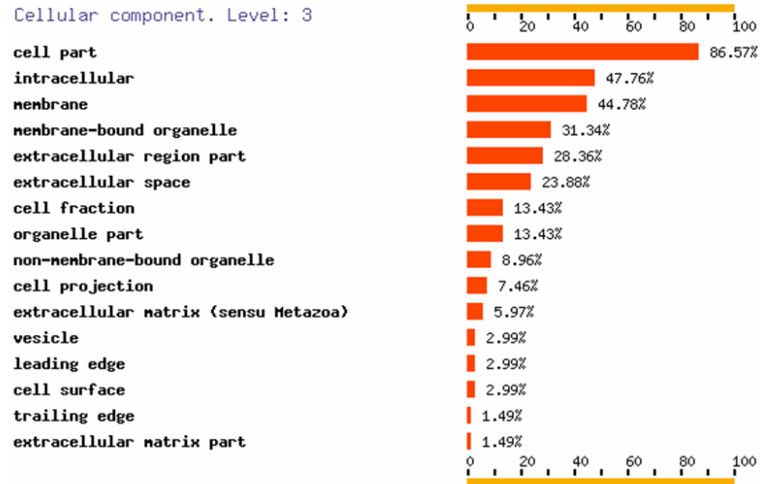


Figure 4.7: GO Cellular Component distribution chart (from Fatigo +).

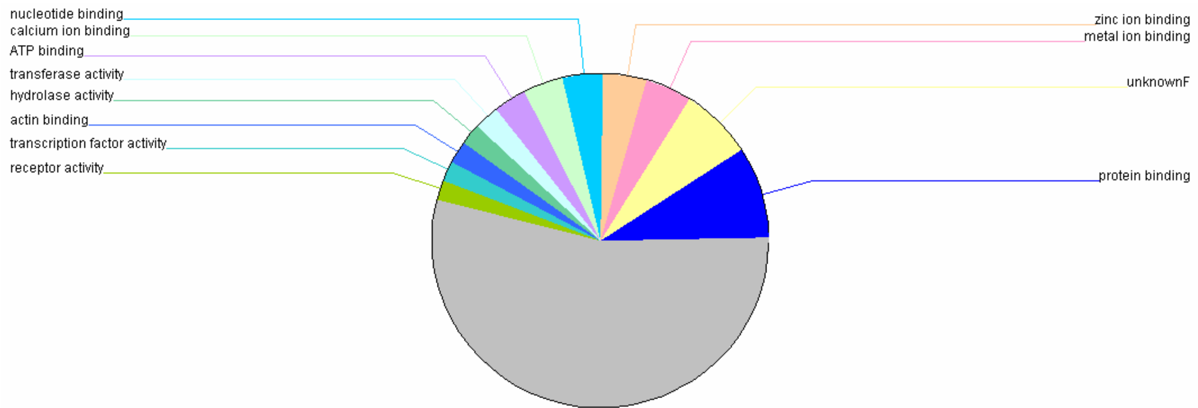


Figure 4.8: GO Molecular Function distribution flat pie chart (from Onto-express).

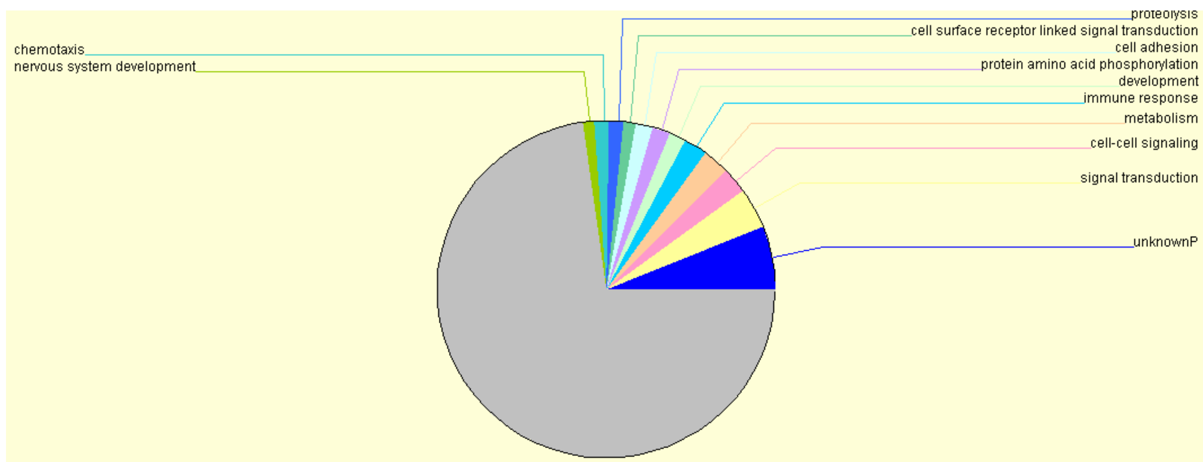


Figure 4.9: GO Biological Function distribution flat pie chart (from Onto-express).

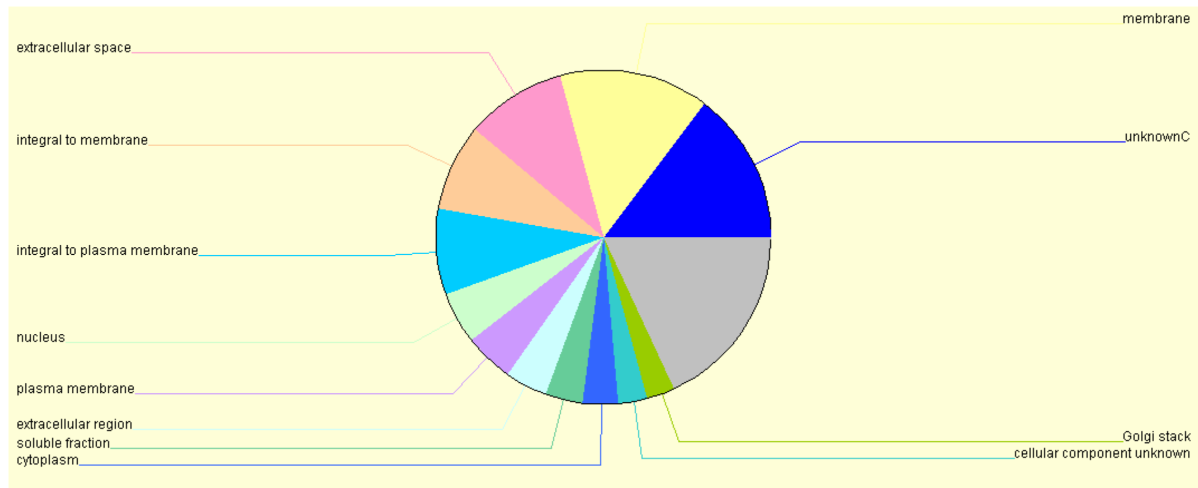


Figure 4.10: GO Cellular Component distribution flat pie chart (from Onto-express).

4.1.5. Hierarchical Clustering

Hierarchical clustering of significant genes according to their gene expressions was demonstrated using Dendogram (Tree view) image of clustering genes. This analysis revealed that the expression of our significant genes was significantly correlated with p53 status [Fig. 4.11].

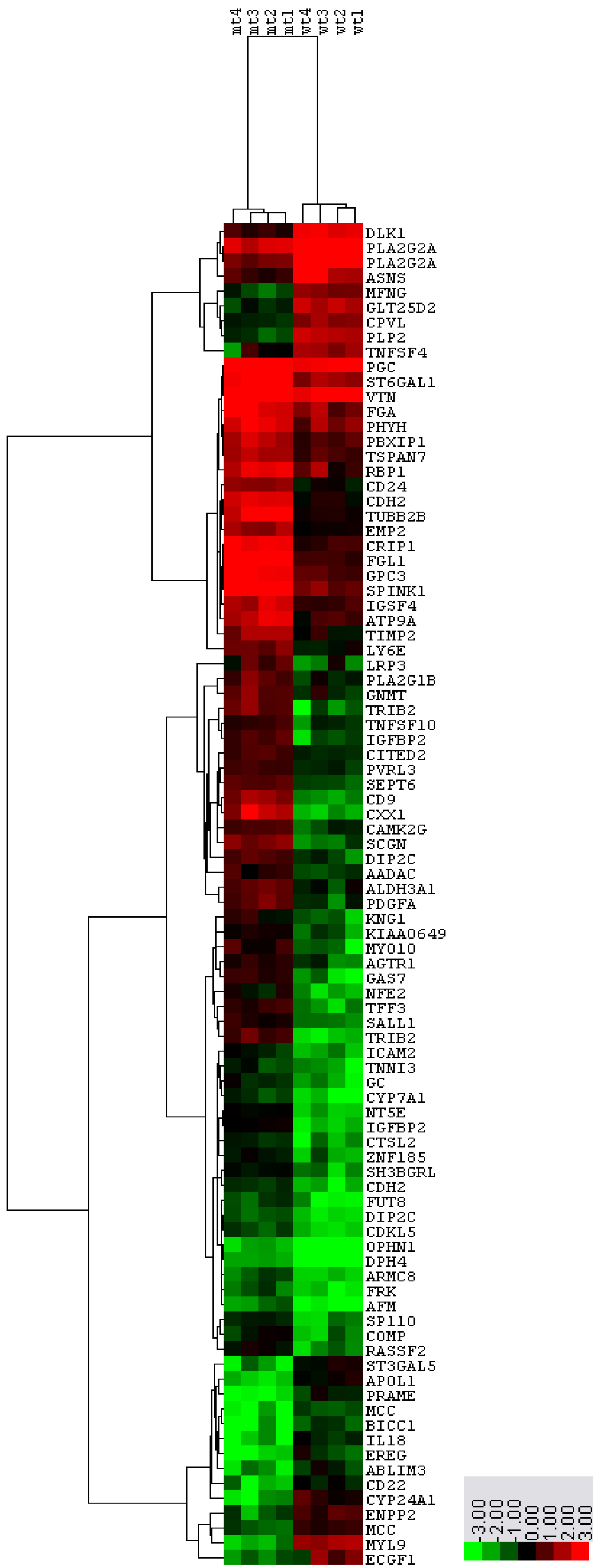


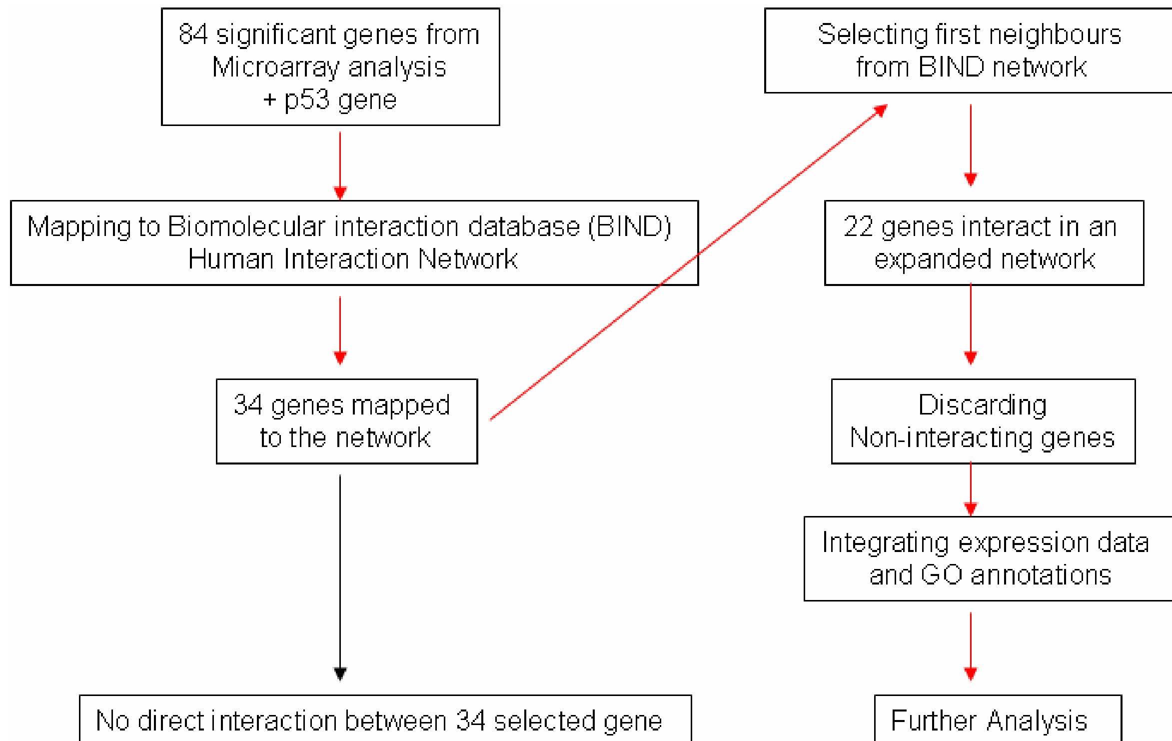
Figure 4.11: Dendogram demonstrating hierarchical clustering of significant genes.

4.2. Gene Network Analysis

4.2.1. Mapping Significant Genes to Human BIND Network

22 of the genes given in the input were observed to be tightly related to each other and to TP53 with specific DNA-DNA or DNA-protein interactions, thus forming a putative interaction network with TP53 at the centre. While remaining genes were discarded, the genes contributing to the network were selected for future analysis.

Table 4.3: Schematic representation of gene network analysis.



4.2.2. Integrating CXX1 and HNF4A to the Network

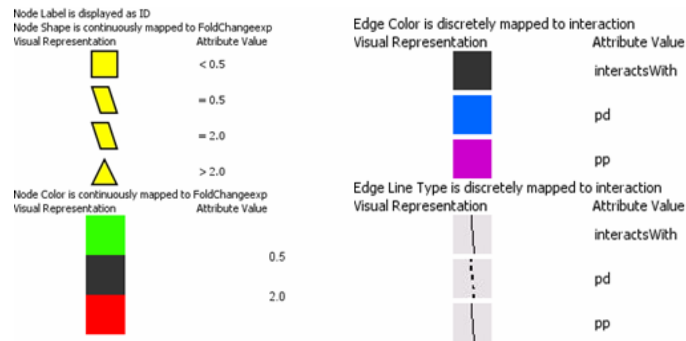
The resulting network was expanded by integrating CXX1 gene to the core network. Moreover, additional direct interaction between HNF4A and p53 was also included in the core network [Fig. 4.12].

4.2.3. Integrating Differential Expression with the Network

Our resulting network data was integrated with corresponding gene expression data. Significant genes were colored according to their expression changes [Fig. 4.12].

Figure 4.12: Putative interaction network showing relationship between p53 and our significant genes.

Visual style (the legend) of the network graphics is as below:



4.2.4. GO Annotation of the Network

Interactions of the resulting network were integrated with GO Biological Process data using GOlorize plug-in. All nodes of the network were clustered and colored according to their corresponding GO category. GO annotation of the obtained interaction network showed that the network was highly enriched with genes involved in cancer-related biological processes such as apoptosis, cell cycle, cell communication, and regulation of angiogenesis. Furthermore, the network was found to be overrepresented with the genes playing role in development and regulation of nucleobase metabolism [Fig. 4.13].

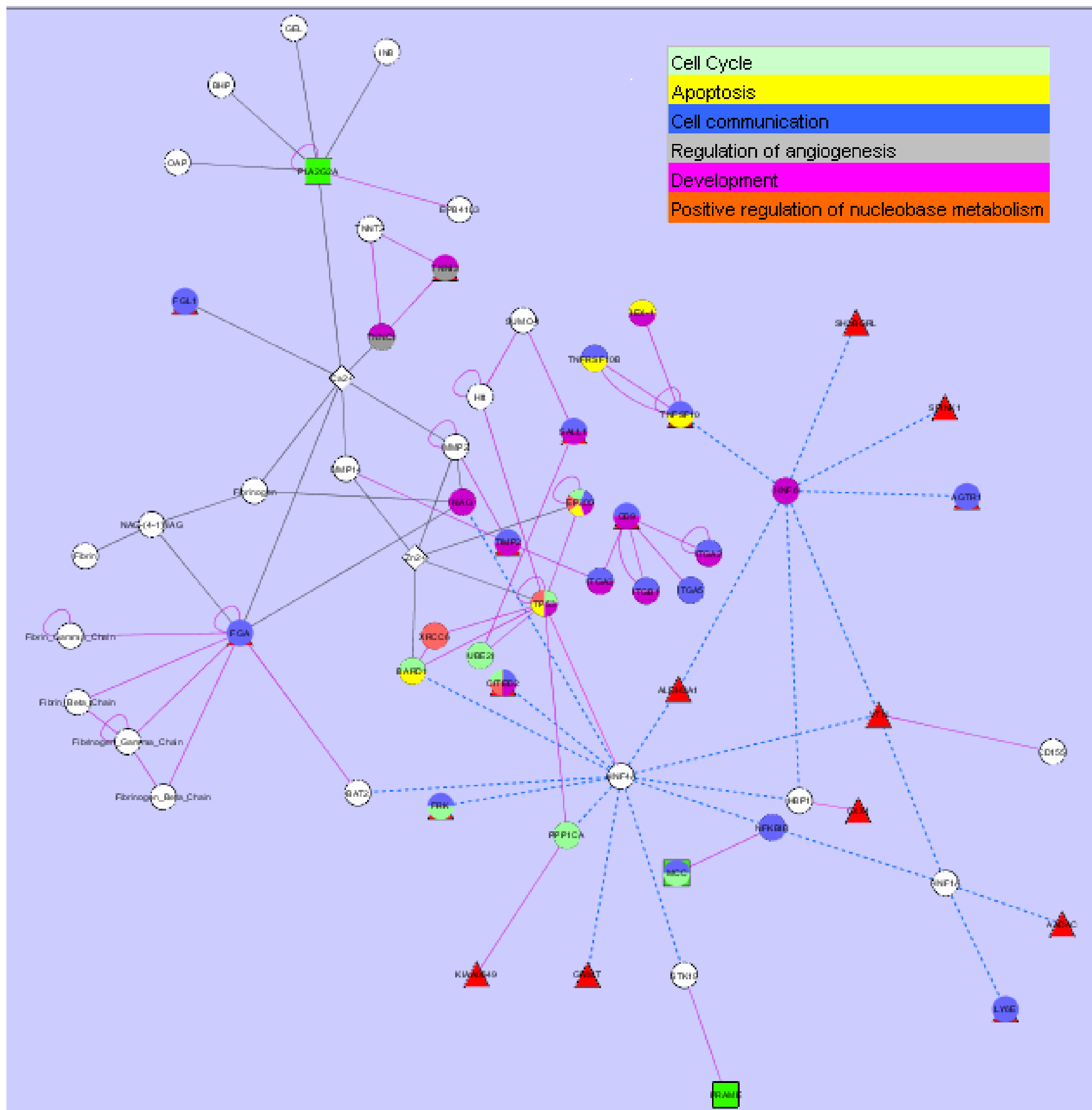


Figure 4.13: GO annotation of the network.

4.2.5. Alternative Layouts of the Network

Cerebral plug-in for Cytoscape was used to generate a layout of the network based on the subcellular localization of the participating molecules. This analysis revealed the distribution pattern of significant genes across the network. Noticeably, most of the direct neighbors of p53 were localized to nucleus and/or cytoplasm. The localization and interactions of HNF transcription factors in nucleus were also apparent [Fig. 1.14].

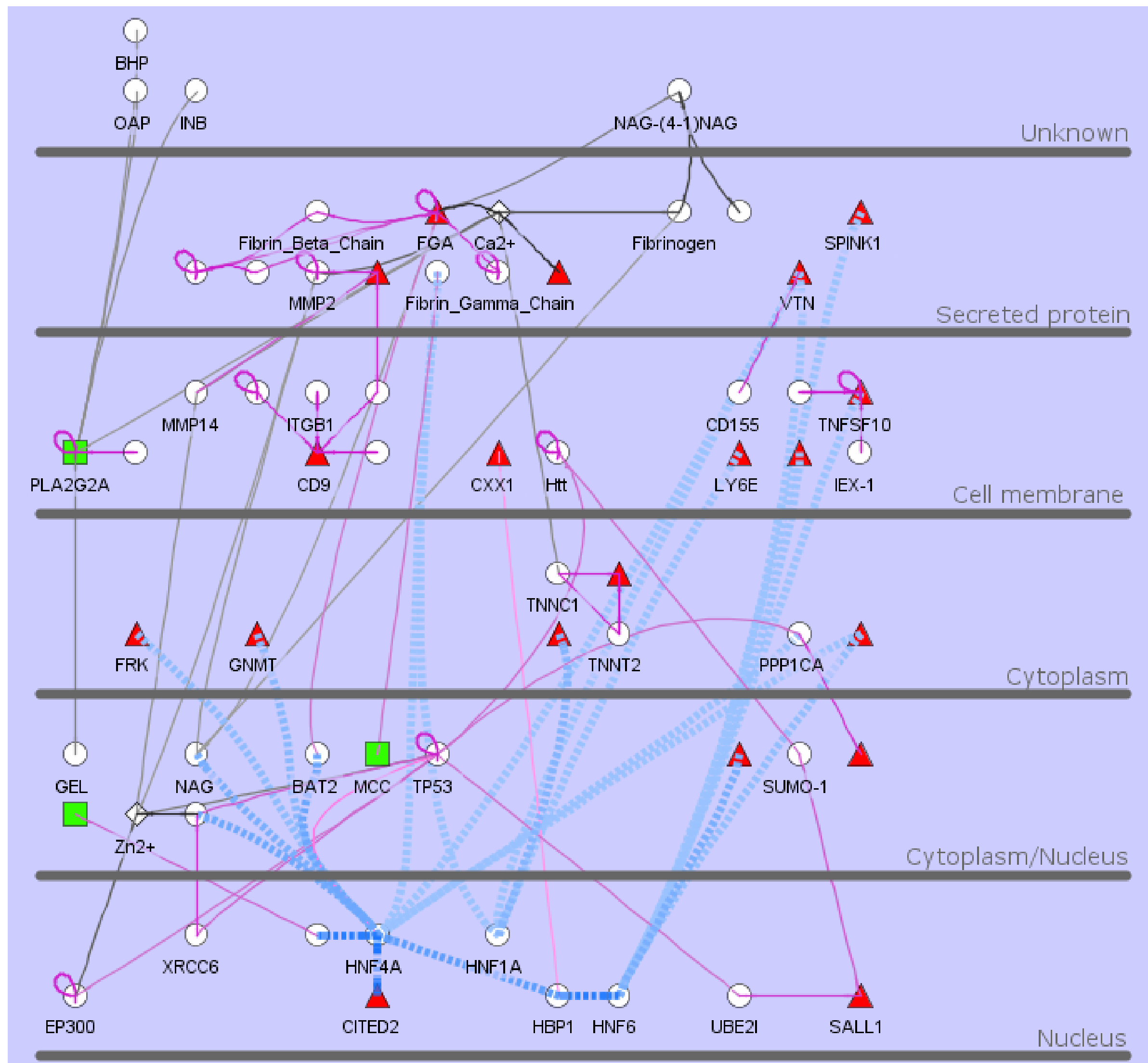


Figure 4.14: Subcellular localization layout of the network.

In addition, Hierarchical layout was used to determine the hierarchy of the network. This alternative layout of the network clearly demonstrated the significance and centrality of both p53 and HNF4A in the network [Fig. 1.15].

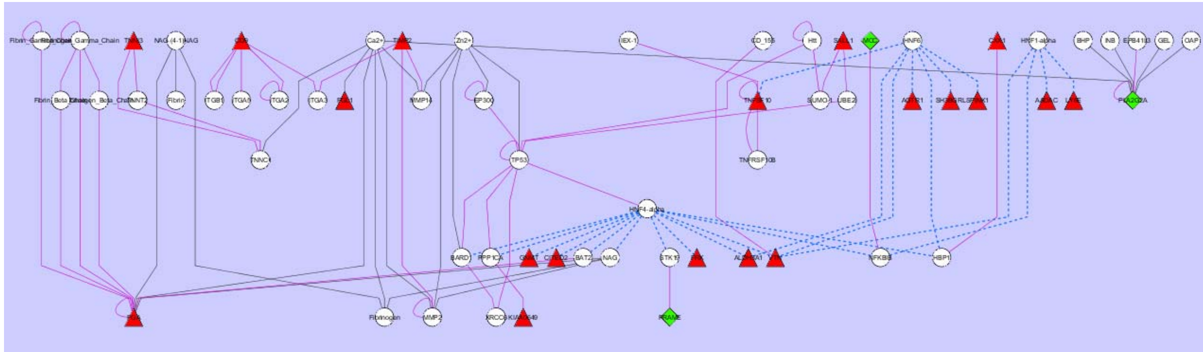


Figure 4.15: Hierarchical layout of the network.

CHAPTER 5: DISCUSSION

5.1. Discussion of the Results from Microarray Data Analysis

Microarray data analysis revealed a molecular signature consisting of 84 differentially regulated genes that could be segregated into two clusters of transcripts induced (n=63) and repressed (n=21) by mutant p53 expression [Table 4.1-4.2].

Since quantitative value sets gained from microarray data don't necessarily answer the research question, translation of this expression data to biologically meaningful information, known as data mining, is achieved using functional annotation tools which enable further analysis of significant genes for biological significance in the light of all the existing knowledge. Functional annotation tools helped us to accurately interpret our microarray data by analyzing our microarray results in the context of other biological information. Annotation showed that the translated proteins of our significant genes possessed diverse properties and were involved in various processes [Fig. 4.3-4.10].

To our surprise, for the most of our significant genes, there was no solid evidence in literature and databases about their functional involvement in cancer. Since molecular basis of mutant p53(R249S) oncogenicity is a poorly-known, these genes may constitute to a novel mechanism responsible for this process, and thus contribute to tumorigenesis.

Classification of our significant genes according to molecular function demonstrated a noticeable high prevalence of "binding proteins" among them [Fig. 4.5 and 4.8]. This finding may reflect the possible role of these proteins in a signal transduction resulting from mutant p53 expression.

It was remarkable that 78 out of 84 significant genes were specific to liver which could be due to the anticipated tissue specificity of p53 R249S mutation to HCC [Fig. 4.3].

Hierarchical clustering of significant genes according to their gene expressions revealed that the expression of our significant genes was significantly correlated with p53 status [Fig. 4.11].

Interestingly, CXX1 (CAAX box protein 1) was the most extraordinary and prominent differentially expressed transcript among our significant genes. Besides being the most overexpressed one, very little is known about its translated protein, which makes this gene one of our target candidate genes for further study. This putatively prenylated protein of

unknown function is ubiquitously expressed in tissues and localized to cell membrane in cells [69]. In our interaction network, CXX1 was found to be in interaction with HBP1 (HMG-box transcription factor 1), which is a transcription factor playing a role in the regulation of the cell cycle [69] [71] [Fig. 4.12]. One significant finding is that HBP1 maintains a proliferation barrier in differentiated liver tissue [72]. Noticeably, HBP1 is also a common transcriptional target of HNF4A and HNF6 transcription factors, which are highly functional in liver [73] [Fig. 4.12]. Taken together, this intriguing background makes CXX1, together with HBP1, worth further investigation for elucidation of their potential role in HCC.

Among our significant genes, Insulin-like growth factor-binding protein-2 (IGFBP2) and glypican-3 (GPC3) were proposed by earlier studies to be valuable as potential diagnostic biomarkers of HCC [74] [75]. Consistent with that, both of these genes were found to be significantly overexpressed in our microarray experiment.

Finally, it was remarkable that two components of fibrinogen complex (FGL1 and FGA) mapped p53 network as the significantly overexpressed genes. Interestingly, they were both possessing molecular function annotated as "cell communication". This could be representing a novel type of signalling mediated by fibrinogen components and contributing to mutant p53(R249S) oncogenicity in HCC.

Validation of significant genes by both semi-quantitative and real-time RT-PCR is among our future goals. Furthermore, we aim to use a list of different cell lines in order to demonstrate that differential expression of our significant genes is not cell line specific but rather a common feature.

5.2. Discussion of the Results from Gene Network Analysis

Comprehensive network analysis of significant genes using Cytoscape and additional plug-ins provided a further insight into the investigated molecular mechanism. To our surprise, significant genes had no interaction between them, but were in close relation with direct neighbours of p53. This resulted in accumulation of significant genes around p53 molecule in our network, clearly demonstrating their relationship to p53 [Fig. 4.12]. This was highly in concordance with our expectations since we were anticipating this relationship between our significant genes and p53. GO annotation of the obtained interaction network showed that the network was highly enriched with genes involved in cancer-related biological processes such as apoptosis, cell cycle, cell communication and angiogenesis [Fig.

4.13]. Noticeably, interacting genes were usually sharing the same GO category, which clearly demonstrates the significance of interactions in understanding gene function [Fig. 4.13]. On the other hand, an alternative layout generated by Cerebral plug-in demonstrated the distribution of molecules involved in the network according to their subcellular localization. Our significant genes, together with neighbor molecules, were found to be equally distributed throughout the cell [Fig. 4.14].

Taking in account that most of our significant genes in the network were the direct transcriptional downstream targets of HNF transcription factors, the functional relationship between HNF factors and mutant p53 oncogenicity in HCC becomes apparent [Fig. 4.12]. Hierarchical layout of Cytoscape interface, used to obtain an alternative image of the network, clearly demonstrated the significance and centrality of both p53 and HNF4A in the network [Fig. 4.15].

It's widely recognized by the scientific community that the transcription factors HNF1A, HNF4A and HNF6, which function coordinately in a connected network in hepatocytes, regulate the development and function of liver, [73] [76] [Fig. 5.1].

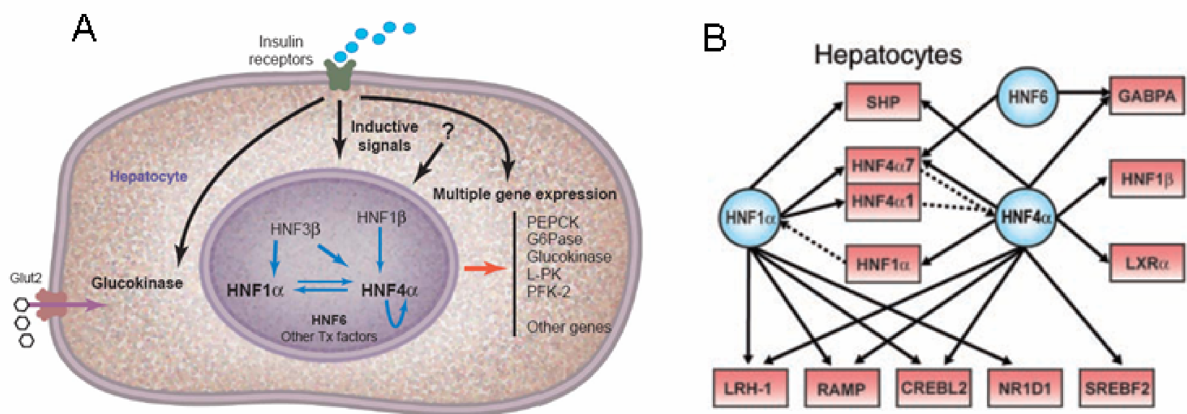


Figure 5.1: Control of liver gene expression by HNF transcription factors. A. Interactions among HNFs in a hepatocyte. B. HNF1A, HNF6, and HNF4A are at the center of tissue-specific transcriptional regulatory networks. In these examples selected for illustration, regulatory proteins and their gene targets are represented as blue circles and red boxes, respectively. Solid arrows indicate protein-DNA interactions, and genes encoding regulators are linked to their protein products by dashed lines. (Kulkarni RN and Kahn CR, 2004; Odom DT *et al*., 2004) [73] [76].

Interestingly, genome-scale chromatin immunoprecipitation (ChIP) assays performed by Odom DT and his colleagues revealed that the number of genes transcriptionally regulated by HNF4A in hepatocytes was much larger than observed with other transcription factors [73].

The observation that HNF4A is binding to an unusual large number (almost half) of active promoters suggests that HNF4A has a broad activity in liver and explains why HNF4A is so crucial in development and activity of this organ [73] [76]. This observation is also in consistence with our results obtained from network analysis, since most of our differentially regulated significant genes were found to be interacting partners and transcriptional targets of HNF4A [Fig. 4.12 and 4.15].

Evidence from the literature that wild-type p53 can bind to HNF4A protein and inhibit its transcriptional function is shedding some light on our findings [77] [Fig 4.12]. Since this repression of HNF4A has been shown with wild-type form of p53, it is difficult to speculate about the relation of this repression with mutant forms. But when this observation is interpreted in the light of our findings, especially those coming from the network analysis, it is possible to drive an appropriate conclusion about the role of HNF4A in mutant p53 oncogenicity in HCC. According to one of the models describing mutant p53 transcriptional activity, mutant p53 interacts with a specific transcription factor that drives its gene target specificity by recruiting it to target genes' promoters [Fig. 1.5]. Consistent with this, our findings strongly suggests that mutant p53 interacts with HNF4A in order to achieve transcriptional regulation of its target genes (which correspond to our significant genes) and promote its oncogenic effect in HCC. Specificity of both HNF4A transcription factors and p53 R249S mutant proteins to HCC, further increases significance of this hypothesis and emphasizes the tissue specificity of these molecular mechanisms to HCC. We aim to perform a series of biochemical analysis to test the proposed functional relationship between HNFA and p53 R249S mutant proteins.

5.3. Conclusion and Future Perspectives

The aim of the present study was to find out the list of differentially expressed genes and the associated gene network affected by the expression of p53(R249S) mutant proteins. Searching for genes that change expression in response to p53 mutation may provide a clue to the mechanism underlying mutant p53 oncogenicity in HCC. Thus, genome-wide gene expression profiling was used to discover a set of genes involved in this process. "Comparative genomic approach" using two isogenic HCC cell lines was exploited as a model for our microarray experiment.

Microarray data analysis revealed a molecular signature consisting of 84 differentially regulated genes (FDR<0.05) that could be segregated into two clusters of transcripts induced (n=63) and repressed (n=21) by mutant p53 expression, showing that the expression of mutant p53 proteins resulted in overall distinct expression profile.

Analyzing our microarray data in the light of the relevant biological data obtained from the curated databases (such as annotation and interaction data) provided a more reliable interpretation of our experimental findings, which led to more comprehensive understanding of the investigated molecular mechanisms. Functional annotation and network analysis resulted in a better elucidation of the interrelations among the discovered differentially expressed genes and aided comprehensive cross-validation of our findings with the existing knowledge about the related molecular mechanisms. We demonstrated that several Hepatocyte Nuclear Factors (HNF1A, HNF4A and HF6) could play an essential role in mediating mutant p53 oncogenic activity, as the key molecules of the gene network. Deregulation of the transcriptional control mediated by these transcription factors appears to be the major mechanism underlying mutant p53 oncogenicity in HCC. Remarkably, CXX1, which is a gene of unknown function, was prominent as the most upregulated transcript among our differentially expressed genes. Further functional analysis of these and other candidate genes of the gene network shall clarify their potential relation to mutant p53 and elucidate their presumptive contribution to the development of HCC.

BIBLIOGRAPHY

[1] Deppert W. Mutant p53: from guardian to fallen angel? *Oncogene*, 26(15):2142-4, 2007.

[2] Hussain SP, Harris CC. p53 biological network: at the crossroads of the cellular-stress response pathway and molecular carcinogenesis. *J Nippon Med Sch.*, 73(2):54-64, 2006.

[3] Cadwell C, Zambetti GP. The effects of wild-type p53 tumor suppressor activity and mutant p53 gain-of-function on cell growth. *Gene*, 277(1-2):15-30, Review, 2001.

[4] Pardee AB. Regulatory molecular biology. *Cell Cycle*, 5(8):846-52, 2006.

[5] Lane DP. Cancer. p53, guardian of the genome. *Nature*, 358(6381):15-6, 1992.

[6] Bykov VJ *et al*. Small molecules that reactivate mutant p53. *Eur J Cancer*, 39(13):1828-34, Review, 2003.

[7] Vousden KH, Lane DP. p53 in health and disease. *Nat Rev Mol Cell Biol.*, 8(4):275-83, Review, 2007.

[8] Entrez Gene *Web Site*. http://www.ncbi.nlm.nih.gov/sites/entrez.

[9] Gross L. Structural Insights into the Regulation of a Key Tumor Suppressor. *PLoS Biol.*, 4(2):e40, 2006.

[10] Sigal A, Rotter V. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. *Cancer Res*., 60(24):6788-93, Review, 2000.

[11] Weinberg RA. Biology of Cancer, Chapter 9:311-314, *Garland Pub*, 2006.

[12] Soussi T. p53 alterations in human cancer: more questions than answers. *Oncogene*, 26(15):2145-56, Review, 2007.

[13] Scian MJ *et al*. Modulation of gene expression by tumor-derived p53 mutants. *Cancer Res.*, 64(20):7447-54, 2004.

[14] Kastan MB, Berkovich E. p53: a two-faced cancer gene. *Nat Cell Biol.*, 9(5):489-91, 2007.

[15] Bossi G *et al*. Mutant p53 gain of function: reduction of tumor malignancy of human cancer cell lines through abrogation of mutant p53 expression. *Oncogene*, 25(2):304-9, 2006.

[16] Zalcenstein A *et al*. Mutant p53 gain of function: repression of CD95(Fas/APO-1)

gene expression by tumor-associated p53 mutants. *Oncogene*, 22(36):5667-76, 2003.

[17] Lang GA *et al*. Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell*, 119(6):861-72, 2004.

[18] Olive KP *et al*. Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell*, 119(6):847-60, 2004.

[19] Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*, 68(4):820-3, 1971.

[20] Levine AJ. p53, the cellular gatekeeper for growth and division. *Cell*, 88(3):323-31, 1997.

[21] Li Y, Prives C. Are interactions with p63 and p73 involved in mutant p53 gain of oncogenic function? *Oncogene*, 26(15):2220-5, Review, 2007.

[22] Blagosklonny MV. p53 from complexity to simplicity: mutant p53 stabilization, gain-of-function, and dominant-negative effect. *FASEB J.*, 14(13):1901-7, 2000.

[23] Dittmer D *et al*. Gain of function mutations in p53. *Nat Genet.*, 4(1):42-6, 1993.

[24] Weisz L *et al*. Transcription regulation by mutant p53. *Oncogene*, 26(15):2202-11, 2007.

[25] Strano S *et al*. Mutant p53: an oncogenic transcription factor. *Oncogene*, 26(15):2212-9, 2007.

[26] Kim E and Deppert W. Transcriptional activities of mutant p53: when mutations are more than a loss. *J Cell Biochem.*, 93(5):878-86, 2004.

[27] Kim E and Deppert W. Interactions of mutant p53 with DNA: guilt by association. *Oncogene*, 26(15):2185-90, 2007.

[28] Menendez D *et al*. Changing the p53 master regulatory network: ELEMENTary, my dear Mr Watson. *Oncogene*, 26(15):2191-201, 2007.

[29] Joerger AC and Fersht AR. Structure-function-rescue: the diverse nature of common p53 cancer mutants. *Oncogene*, 26(15):2226-42, 2007.

[30] Chen X *et al*. Gene expression patterns in human liver cancers. *Mol Biol Cell.*, 13(6):1929-39, 2002.

[31] Lee JS and Thorgeirsson SS. Comparative and integrative functional genomics of HCC. *Oncogene*, 25(27):3801-9, 2006.

[32] Hussain SP *et al.* TP53 mutations and hepatocellular carcinoma: insights into the

etiology and pathogenesis of liver cancer. *Oncogene*, 26(15):2166-76, Review, 2007.

[33] Patil MA *et al*. An integrated data analysis approach to characterize genes highly expressed in hepatocellular carcinoma. *Oncogene*, 24(23):3737-47, 2005.

[34] Wang XW *et al*. Molecular pathogenesis of human hepatocellular carcinoma. *Toxicology*, 181-182:43-7, 2002.

[35] Laurent-Puig P and Zucman-Rossi J. Genetics of hepatocellular tumors. *Oncogene*, 25(27):3778-86. 2006.

[36] Hainaut P. Tumor-specific mutations in p53: the acid test. *Nat Med.*, 8(1):21-3, 2002.

[37] Quackenbush J. Microarray analysis and tumor classification. *N Engl J Med.*, 354(23):2463-72, 2006.

[38] Hornberg JJ *et al*. Cancer: a Systems Biology disease. *Biosystems*, 83(2-3):81-90, 2006.

[39] Hooper SD *et al*. Identification of tightly regulated groups of genes during Drosophila melanogaster embryogenesis. *Mol Syst Biol.*, 3:72, 2007.

[40] Rhodes DR *et al*. Mining for regulatory programs in the cancer transcriptome. *Nat Genet.*, 37(6):579-83, 2005.

[41] Brentani RR *et al*. Gene expression arrays in cancer research: methods and applications. *Crit Rev Oncol Hematol.*, 54(2):95-105, 2005.

[42] Schlitt T *et al*. From gene networks to gene function. *Genome Res.*, 13(12):2568-76, 2003.

[43] Brazhnik P *et al*. Gene networks: how to put the function in genomics. *Trends Biotechnol.*, 20(11):467-72, 2002.

[44] Rung J *et al*. Building and analysing genome-wide gene disruption networks. *Bioinformatics*, 18 Suppl 2:S202-10, 2002.

[45] Sharan R *et al*. Network-based prediction of protein function. *Mol Syst Biol.*, 3:88, 2007.

[46] Moran LB *et al*. The microglial gene regulatory network activated by interferon-gamma. *J Neuroimmunol.*, 183(1-2):1-6, 2007.

[47] Huang Y *et al*. Identification of novel transcriptional networks in response to treatment with the anticarcinogen 3H-1,2-dithiole-3-thione. *Physiol Genomics*,

24(2):144-53, 2006.

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

[49] Troyanskaya OG. Putting microarrays in a context: integrated analysis of diverse biological data. *Brief Bioinform.*, 6(1):34-43, 2005.

[50] Affymetrix *Web Site*. https://www.affymetrix.com.

[51] The R Project *Web Site*. http://www.r-project.org.

[52] Bolstad BM *et al*. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185-93, 2003.

[53] Tusher VG *et al*. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*, 98(9):5116-21, 2001.

[54] SAMR package. *Web Site*. http://cran.r-project.org/doc/packages/samr.pdf.

[55] Liu G *et al*. NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res.*, 31(1):82-6, 2003.

[56] Khatri P *et al*. Profiling gene expression using onto-express. *Genomics*, 79(2):266-70, 2002.

[57] Zhang B *et al*. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.*, 33(Web Server issue):W741-8, 2005.

[58] Al-Shahrour F *et al*. FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res.*, 35(Web Server issue):W91-6, 2007.

[59] Eisen MB *et al*. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A.*, 95(25):14863-8, 1998.

[60] Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics*, 20(17):3246-8, 2004.

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

[62] Gilbert D. Biomolecular interaction network database. *Brief Bioinform.*, 6(2):194-8, 2005.

[63] Chatr-aryamontri A *et al*. MINT: the Molecular INTeraction database. *Nucleic Acids*

Res., 35(Database issue):D572-4, 2007.

[64] von Mering C *et al*. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.*, 33(Database issue):D433-7, 2005.

[65] The Cytoscape Project. *Web Site*. http://www.cytoscape.org.

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

[67] Garcia O *et al*. GOlorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. *Bioinformatics*, 23(3):394-6, 2007.

[68] Barsky A *et al*. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics*, 23(8):1040-2, 2007.

[69] The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 35(Database issue):D193-7, 2007.

[70] The MEP2SL Project. *Web Site*. http://www.i-cancer.fen.bilkent.edu.tr/mep2sl.

[71] Lim J *et al*. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*, 125(4):801-14, 2006.

[72] Shih HH *et al*. HMG box transcriptional repressor HBP1 maintains a proliferation barrier in differentiated liver tissue. *Mol Cell Biol.*, 21(17):5723-32, 2001.

[73] Odom DT *et al*. Control of pancreas and liver gene expression by HNF transcription factors. *Science*, 303(5662):1378-81, 2004.

[74] Ranke MB *et al*. Pilot study of elevated levels of insulin-like growth factor-binding protein-2 as indicators of hepatocellular carcinoma. *Clin Cancer Res.*, 13(4):1133-9, 2007.

[75] Jia HL *et al*. Gene expression profiling reveals potential biomarkers of human hepatocellular carcinoma. *Clin Cancer Res.*, 13(4):1133-9, 2007.

[76] Kulkarni RN and Kahn CR. Molecular biology. HNFs--linking the liver and pancreatic islets in diabetes. *Science*, 303(5662):1311-2, 2004.

[77] Maeda Y *et al*. Repression of hepatocyte nuclear factor 4alpha tumor suppressor p53: involvement of the ligand-binding domain and histone deacetylase activity. *Mol Endocrinol.*, 16(2):402-10, 2002.

APPENDIX

Table A.1: Delta table.

delta	# med false pos	90th perc false pos	# called	median FDR	90th perc FDR
0	8911.524277	9046.506931	11957	0.74529767	0.75658668
0.000470238	8882.163644	9025.153743	11926	0.74477307	0.756762849
0.001880953	8815.053624	8962.619406	11864	0.74300857	0.75544668
0.004232143	8688.078416	8853.565624	11755	0.73909642	0.753174447
0.007523811	8518.778139	8714.388594	11603	0.73418755	0.75104616
0.011755954	8317.066772	8543.563089	11431	0.72758873	0.747402947
0.016928574	7703.162614	8065.404198	10853	0.70977265	0.743149746
0.02304167	6991.262574	7488.105505	10170	0.68743978	0.73629356
0.030095242	6132.559366	6790.695129	9305	0.65906065	0.729789912
0.038089291	5220.854495	6034.563485	8349	0.62532692	0.722788775
0.047023816	4419.347327	5333.34004	7508	0.58861845	0.710354294
0.056898817	3601.44396	4599.324198	6626	0.54353214	0.69413284
0.067714295	2719.862337	3738.333149	5519	0.49281796	0.677356976
0.079470249	1875.648792	2848.362772	4339	0.43227674	0.656456043
0.092166679	1348.301307	2236.746455	3503	0.38489903	0.63852311
0.105803585	920.4749307	1675.84396	2704	0.34041233	0.619764778
0.120380968	640.2143366	1257.550257	2117	0.30241584	0.594024685
0.135898827	415.2432475	885.3946931	1543	0.26911422	0.5738138
0.152357163	271.8718416	626.1059802	1141	0.23827506	0.548734426
0.169755975	162.4367525	416.0058614	775	0.20959581	0.536781757
0.188095263	113.2481584	301.9950891	601	0.18843288	0.502487669
0.207375027	64.44087129	196.7543762	414	0.15565428	0.475252117
0.227595268	38.512	142.990099	317	0.12148896	0.451072868
0.248755985	25.92887129	109.0537822	249	0.10413201	0.437966997
0.270857178	13.7270495	72.82962376	167	0.0821979	0.436105532
0.293898848	7.626138614	55.67081188	138	0.05526187	0.40341168
0.317880994	5.33829703	41.18114851	110	0.04852997	0.374374077
0.342803616	2.287841584	29.74194059	86	0.02660281	0.345836519
0.368666715	1.906534653	24.7849505	79	0.02413335	0.313733551
0.39547029	1.143920792	20.20926733	69	0.01657856	0.292887932
0.423214341	0.762613861	15.63358416	42	0.01815747	0.372228194
0.451898869	0	13.34574257	36	0	0.370715072
0.481523873	0	12.58312871	34	0	0.370092021
0.512089353	0	10.67659406	30	0	0.355886469
0.543595309	0	10.29528713	29	0	0.355009901
0.576041742	0	9.913980198	28	0	0.354070721
0.609428651	0	8.388752475	27	0	0.310694536
0.643756037	0	7.626138614	24	0	0.317755776
0.679023899	0	6.100910891	18	0	0.338939494
0.715232237	0	3.050455446	10	0	0.305045545
0.752381051	0	2.287841584	9	0	0.25420462

0.790470342	0	2.287841584	8	0	0.285980198
0.829500109	0	0.381306931	4	0	0.095326733
0.869470352	0	0.381306931	3	0	0.12710231
0.910381072	0	0	2	0	0
0.952232268	0	0	2	0	0

Table A.2: Upregulated significant probesets.

Order	Affy ID	Fold Change	Order	Affy ID	Fold Change
1	33856_at	23.22800715	44	35540_at	3.6593011
2	39389_at	10.66870392	45	35599_at	2.8330203
3	717_at	10.73610878	46	912_s_at	2.6748451
4	38582_at	6.836368805	47	35703_at	4.3669426
5	40567_at	10.78087903	48	37360_at	2.8121463
6	41416_at	6.910175987	49	37538_at	3.8505759
7	37478_at	7.394791124	50	32139_at	3.3023671
8	39332_at	6.631740161	51	38454_g_at	3.2906377
9	1741_s_at	5.618615288	52	32771_at	3.4573222
10	38396_at	5.099639465	53	41352_at	2.5740928
11	2053_at	5.566360188	54	1089_i_at	11.995563
12	40113_at	8.830691762	55	39413_at	2.6728229
13	33387_at	7.459127617	56	39959_at	3.1956445
14	39350_at	4.385885356	57	40162_s_at	3.2447711
15	33232_at	5.06460286	58	33564_at	4.4038022
16	31477_at	5.646708948	59	444_g_at	2.4560301
17	31926_at	4.75273648	60	41628_at	3.699118
18	31886_at	4.821628984	61	37235_g_at	3.2801068
19	38826_at	3.892380553	62	35718_at	3.2369906
20	40422_at	4.975579655	63	33699_at	2.4102668
21	33377_at	3.497558863	64	443_at	2.9843342
22	35831_at	3.657102437	65	38645_at	2.6455748
23	32985_at	3.509610598	66	39714_at	3.1787233
24	37179_at	4.418474567	67	33113_at	2.4473609
25	36757_at	4.224963097	68	31814_i_at	4.5694804
26	35127_at	4.062052119	69	40031_at	2.9872145
27	2054_g_at	4.044187193	70	38408_at	2.3152697
28	36707_s_at	3.056986098	71	36512_at	2.531105
29	266_s_at	3.528537463	72	37203_at	2.7636325
30	39631_at	3.292796742	73	38576_at	3.2969438
31	33408_at	3.680506015	74	40717_at	3.4912466
32	1375_s_at	3.791424047	75	38063_at	2.5623775
33	37929_at	3.126117522	76	36477_at	3.1097768
34	38294_at	3.09622336	77	35829_at	2.3612818
35	39282_at	3.145138992	78	38825_at	2.4867213

36	1090_f_at	4.446741526	79	38634_at	3.4380836
37	32609_at	3.880073136	80	35023_at	3.6212811
38	33407_at	2.814455399	81	32724_at	2.5036978
39	36727_at	4.36872055	82	34202_at	2.3633705
40	37598_at	3.483605186	83	37983_at	2.9770639
41	36983_f_at	3.15185107	84	39580_at	2.9357271
42	32105_f_at	3.20015185	85	1715_at	2.8402515
43	35362_at	6.26521794			

Table A.3: Downregulated significant probesets.

Order	Affy ID	Fold Change
1	39145_at	0.069119421
2	37326_at	0.128723343
3	33541_s_at	0.136525709
4	157_at	0.15274071
5	37017_at	0.166988218
6	39550_at	0.167024973
7	35099_at	0.188269253
8	614_at	0.195493145
9	41522_at	0.200999525
10	32648_at	0.223333805
11	660_at	0.17863134
12	38323_at	0.246093797
13	39470_at	0.254135678
14	36671_at	0.244807658
15	1665_s_at	0.243437875
16	41123_s_at	0.271027122
17	34256_at	0.246145082
18	38522_s_at	0.273442262
19	39597_at	0.322833829
20	34476_r_at	0.260104239
21	1165_at	0.25988872
22	39506_at	0.307110964
23	35561_at	0.285608674
24	1832_at	0.340424877
25	32319_at	0.298730427

Table A.4: Input data of subcellular localization labels for Cerebral plug-in.

Localization

Fibrinogen_Beta_Chain = Secreted protein

Ca2+ = Secreted protein

OAP = Unknown

NFKBIB = Secreted protein

Fibrin_Gamma_Chain = Secreted protein

PRAME = Cytoplasm/Nucleus

HNF6 = Nucleus

BHP = Unknown

ITGA3 = Cell membrane

BARD1 = Cytoplasm/Nucleus

HNF4-alpha =Nucleus

PPP1CA = Cytoplasm

EP300 = Nucleus

SPINK1 = Secreted protein

Fibrin = Secreted protein

XRCC6 = Nucleus

ITGB1 = Cell membrane

PLA2G2A = Cell membrane

STK19 = Nucleus

SUMO-1 = Cytoplasm/Nucleus

TNNC1 = Cytoplasm

EPB41L3 = Cell membrane

CD9 = Cell membrane

Htt = Cell membrane

GEL = Cytoplasm/Nucleus

AGTR1 = Cell membrane

FRK = Cytoplasm

ITGA5 = Cell membrane

SALL1 = Nucleus

TNNT2 = Cytoplasm

AADAC = Cytoplasm

Fibrin_Beta_Chain = Secreted protein

CD_155 = Cell membrane

SH3BGRL = Cytoplasm/Nucleus

TIMP2 = Secreted protein

INB = Unknown

FGA = Secreted protein

CXX1 = Cell membrane

NAG-(4-1)NAG = Unknown

Fibrinogen = Secreted protein

MMP14 = Cell membrane

CITED2 = Nucleus

TNFRSF10B = Cell membrane

Fibrinogen_Gamma_Chain = Secreted protein

IEX-1 = Cell membrane

Zn2+ = Cytoplasm/Nucleus

HNF1-alpha = Nucleus

ITGA2 = Cell membrane

KIAA0649 = Cytoplasm/Nucleus

UBE2I = Nucleus

ALDH3A1 = Cytoplasm

TNFSF10 = Cell membrane

FGL1 = Secreted protein

BAT2 = Cytoplasm/Nucleus

MMP2 = Secreted protein

GNMT = Cytoplasm

NAG = Cytoplasm/Nucleus

LY6E = Cell membrane

HBP1 = Nucleus

MCC = Cytoplasm/Nucleus

VTN = Secreted protein

TP53 = Cytoplasm/Nucleus

TNNI3 = Cytoplasm