Logical models of DNA damage induced pathways to cancer

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Life Sciences

2013

Kun Tian

Faculty of Life Sciences

Table of Contents

Abstract

This thesis with the title:”Logical models of DNA damage induced pathways to cancer” was completed by Kun Tian for his PhD degree in the University of Manchester and submitted in October 2013. Chemotherapy is commonly used in cancer treatments, however only 25 % of cancers are responsive and a significant proportion develops resistance. The p53 tumour suppressor is crucial for cancer development and therapy, but has been less amenable to therapeutic applications due to the complexity of its action reflected in 67,000 papers describing its function. Here we provide a systematic approach to integrate this information by constructing large-scale logical models of the p53 interactome using extensive database and literature integration. Initially we generated models using manual curation to demonstrate the feasibility of the approach. This was followed by creation of the next generation models by automatic text mining results retrieval. Final model PKT205/G3 was generated by choosing the size of the interactome that could be analysed with current available computing power and by linking upstream nodes to input environmental signals such as DNA damage and downstream nodes to output signal such as apoptosis. This final version of the PKT205/G3 model contains 205 nodes representing genes or proteins, DNA damage input and apoptosis output, and 677 logical interactions. Predictions from *in silico* knock-outs and steady state model analysis were validated using literature searches and *in vitro* experiments. We identify an up regulation of Chk1, ATM and ATR pathways in p53 negative cells and 58 other predictions obtained by knockout tests mimicking mutations. The comparison of model simulations with microarray data demonstrated a significant rate of successful predictions ranging between 52 % and 71 % depending on the cancer type. Growth factors and receptors FGF2, IGF1R, PDGFRB and TGFA were identified as factors contributing selectively to the control of U2OS osteosarcoma and HCT116 colon cancer cell growth. In summary, we provide the proof of principle that this versatile and predictive model has vast potential for use in cancer treatment by identifying pathways in individual patients that contribute to tumour growth, defining a sub population of “high” responders and identification of shifts in pathways leading to chemotherapy resistance.

Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of The University of Manchester or any other university or institute of learning.

Copyright statement

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and he has given the University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be mad **only** in accordance with the Copyright. Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has form time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks, and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written

permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http:// www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library. The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in the University's policy on presentation of Thesis.

Acknowledgements

First, I would like to thank my supervisors Dr. Marija Krstic-Demonacos and Dr. Jean-Marc Schwartz for their help and guidance in my project. I would like to thank Dr. Ramkumar Rajendran and Manjula Doddananjaiah for their help in my lab work. I would like to thank all my group members for their help in my project. I thank my parents for their moral and financial support for my PhD project in the last four years. I thank UMIP (University of Manchester Intellectual Property) for funding. I thank the Core Facility of the Faculty of Life Sciences in the University of Manchester for help with the microarray experiments.

Abbreviations

AATF	apoptosis antagonizing transcription factor
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
ABCC1	ATP-binding cassette, sub-family C (CFTR/MRP), member 1
ABM	agent-based modelling
ADP	adenosine diphosphate
APAF1	apoptotic peptidase activating factor 1
APE	AP endonuclease
API	activator protein 1
ATF3	activating transcription factor 3
AR	androgen receptor
ARID3A	AT rich interactive domain 3A (BRIGHT-like)
ATM	ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	ataxia telangiectasia and Rad3 related
AXIN1	axin 1
BAK1	BCL2-antagonist/killer 1
BAX	BCL2-associated X protein
BBC3	BCL2 binding component 3
BCL2	B-cell CLL/lymphoma 2
BCL3	B-cell CLL/lymphoma 3
BCL6	B-cell CLL/lymphoma 6
BDKRB1	bradykinin receptor B1
BioCreative initiative	Critical Assessment of Information Extraction Systems in Biology
BN	Boolean network
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like
BRCA1	breast cancer 1, early onset

BTG2	BTG family, member 2
C12orf5	chromosome 12 open reading frame 5
C13orf15	chromosome 13 open reading frame 15
CA	cellular automata
CALD1	caldesmon 1
CASP8	caspase 8, apoptosis-related cysteine peptidase
CCNA1	cyclin A1
CCNA2	cyclin A2
CCNB1	cyclin B1
CCND1	cyclin D1
CCNG1	cyclin G1
CDC20	cell division cycle 20 homolog (S. cerevisiae)
CDC25A	cell division cycle 25 homolog A (S. pombe)
CDK4	cyclin-dependent kinase 4
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)
CDKN2A	cyclin-dependent kinase inhibitor 2A
CD4	CD4 molecule
CD44	CD44 molecule (Indian blood group)
CD58	CD58 molecule
CD59	CD59 molecule, complement regulatory protein
CD82	CD82 molecule
CDK2	cyclin-dependent kinase 2
CDK4	cyclin-dependent kinase 4
CDK5	cyclin-dependent kinase 5
CDK6	cyclin-dependent kinase 6
CDK9	cyclin-dependent kinase 9
CDKN2A	cyclin-dependent kinase inhibitor 2A

CHEK1	checkpoint kinase 1
CHEK2	checkpoint kinase 2
CIAPIN1	cytokine induced apoptosis inhibitor 1
CKB	creatine kinase, brain
CKM	creatine kinase, muscle
CKS2	CDC28 protein kinase regulatory subunit 2
COL18A1	collagen, type XVIII, alpha 1
CRE	cAMP response element
CSNK2A1	casein kinase 2, alpha 1 polypeptide
CSNK2A2	casein kinase 2, alpha prime polypeptide
CXCR4	chemokine (C-X-C motif) receptor 4
DAVID	The Database for Annotation, Visualization and Integrated Discovery
DCA	Dynamic Cellular Automata
DDB2	damage-specific DNA binding protein 2, 48kDa
DDI	domain-domain interaction
DDIT4	DNA-damage-inducible transcript 4
DDX5	DEAD (Asp-Glu-Ala-Asp) box helicase 5
DFNA5	deafness, autosomal dominant 5
DHFR	dihydrofolate reductase
DKK1	dickkopf 1 homolog (Xenopus laevis)
DMP1	dentin matrix acidic phosphoprotein 1
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRAM1	damage-regulated autophagy modulator 1
DSB	DNA double-strand breaks
DUSP2	dual specificity phosphatase 2
DUSP4	dual specificity phosphatase 4
DUSP5	dual specificity phosphatase 5

DYRK2	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2
EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase 2
ELAVL1	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)
ECT2	epithelial cell transforming sequence 2 oncogene
EDA2R	ectodysplasin A2 receptor
EGFR	epidermal growth factor receptor
EPHB4	EPH receptor B4.
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)
ESR1	estrogen receptor 1
EZH2	enhancer of zeste homolog 2 (Drosophila)
E2F1	E2F transcription factor 1
E4F1	E4F transcription factor 1
FA/BRCA	Fanconi Anemia/Breast Cancer
FADD	Fas(TNFRSF6)-associated via death domain
FAS	Fas (TNF receptor superfamily, member 6)
FDR	False Discovery Rate
FDXR	ferredoxin reductase
FEN1	flap structure-specific endonuclease 1
FGF2	fibroblast growth factor 2 (basic)
FHL2	four and a half LIM domains 2
FOS	FBJ murine osteosarcoma viral oncogene homolog
FOXM1	forkhead box M1
GADD45A	growth arrest and DNA-damage-inducible, alpha
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GGR	global genomic repair
GO	Gene Ontology

GOF	gain of function
GSTP1	glutathione S-transferase pi 1
GTSE1	G-2 and S-phase expressed 1
G1 phase	Gap 1
G2 phase	Gap 2
H2AFZ	H2A histone family, member Z
HDAC1	histone deacetylase 1
HIC1	hypermethylated in cancer 1
HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
HIPK2	homeodomain interacting protein kinase 2
HIPK4	homeodomain interacting protein kinase 4
HMMR	hyaluronan-mediated motility receptor (RHAMM)
HNF4A	hepatocyte nuclear factor 4, alpha
HOXA11	homeobox A11
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
HSPA4	heat shock 70kDa protein 4
HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1
HTATIP2	HIV-1 Tat interactive protein 2, 30kDa
IAP	inhibitors of apoptosis
KEGG	Kyoto Encyclopedia of Genes and Genomes
LATS2	LATS, large tumor suppressor, homolog 2 (Drosophila)
ICAM1	intercellular adhesion molecule 1
ICLs	interstrand cross-links
ID3	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein
IER3	immediate early response 3
IFI16	interferon, gamma-inducible protein 16
IFITM2	interferon induced transmembrane protein 2

IFNA1	interferon, alpha 1
IGFBP1	insulin-like growth factor binding protein 1
IGFBP7	insulin-like growth factor binding protein 7
IGF1R	insulin-like growth factor 1 receptor
IL6	interleukin 6 (interferon, beta 2)
IQCB1	IQ motif containing B1
IR	ionizing radiation
ISG15	ISG15 ubiquitin-like modifier
KAT2B	K(lysine) acetyltransferase 2B
KLF4	Kruppel-like factor 4 (gut)
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
KRT8	keratin 8
KRT19	keratin 19
LATS2	LATS, large tumor suppressor, homolog 2 (Drosophila)
LRDD	PIDD p53-induced death domain protein
LSSA	logical steady state analysis
LTF	lactotransferrin
M phase	mitosis
MAP4	microtubule-associated protein 4
MAPK1	mitogen-activated protein kinase 1
MAPK8	mitogen-activated protein kinase 8
MAPK9	mitogen-activated protein kinase 9
MAPK14	mitogen-activated protein kinase 14
MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4
MCL1	myeloid cell leukemia sequence 1 (BCL2-related)
MCTS1	malignant T cell amplified sequence 1
MDM2	Mdm2, p53 E3 ubiquitin protein ligase homolog (mouse)
MDM4	p53 binding protein homolog (mouse)

MGMT	O-6-methylguanine-DNA methyltransferase
MEDLINE	Medical Literature Analysis and Retrieval System Online
MIP	MDM2 Inhibitory Peptide
MMP1	matrix metallopeptidase 1 (interstitial collagenase)
MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)
MMP13	matrix metallopeptidase 13 (collagenase 3)
mTOR	mammalian target of rapamycin
MUC1	mucin 1, cell surface associated
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
NAD	nicotinamide adenine dinucleotide
NCL	nucleolin
NER	nucleotide excision repair
NFAT	Nuclear factor of activated T-cells
NIH	National Institutes of Health
NLRC4	NLR family, CARD domain containing 4
NME1	non-metastatic cells 1, protein (NM23A) expressed in
NOTCH1	notch 1' Neurogenic locus notch homolog protein 1
NOV	nephroblastoma overexpressed
NRAS	neuroblastoma RAS viral (v-ras) oncogene homolog
NR2C1	nuclear receptor subfamily 2, group C, member 1
MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)
NTN1	netrin 1
ODE	ordinary differential equation
OMIM	Online Mendelian Inheritance in Man
OXPHOS	oxidative phosphorylation

P53	tumor protein p53
P53AIP1	tumor protein p53 regulated apoptosis inducing protein 1
P53BP2	tumour suppressor p53-binding protein 2
PADI4	peptidyl arginine deiminase, type IV
PARK2	parkinson protein 2, E3 ubiquitin protein ligase (parkin)
PCBP4	poly(rC) binding protein 4
PCNA	proliferating cell nuclear antigen
PDEs	partial differential equations
PDGFRB	platelet-derived growth factor receptor, beta polypeptide
PDRG1	p53 and DNA-damage regulated 1
PERP	PERP, TP53 apoptosis effector
PEG3	paternally expressed 3
PI	protease
PLAUR	plasminogen activator, urokinase receptor
PMID	PubMed identifier or PubMed unique identifier
PML	promyelocytic leukaemia
POU4F1	POU class 4 homeobox 1
PPM1A	protein phosphatase, Mg2+/Mn2+ dependent, 1A
PPM1D	protein phosphatase, Mg2+/Mn2+ dependent, 1D
PRC1	protein regulator of cytokinesis 1
PRKCA	protein kinase C, alpha
PRKDC	protein kinase, DNA-activated, catalytic polypeptide
PRKD1	protein kinase D1
PRKG1	protein kinase, cGMP-dependent, type I
PRSS50	protease, serine, 50
PSEN1	presenilin 1
PSMD10	proteasome (prosome, macropain) 26S subunit, non-ATPase, 10
PTEN	phosphatase and tensin homolog

PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
PTTG1	pituitary tumor-transforming 1
PVDF	polyvinylidene fluoride
Pu	purine
Py	pyrimidine
RAD51	RAD51 homolog (S. cerevisiae).
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1
RASD1	RAS, dexamethasone-induced 1
Rb	retinoblastoma protein
RECQL4	RecQ protein-like 4
REG	regulatory
RGS16	regulator of G-protein signaling 16
ROS	reactive oxygen species
rpm	revolutions per minute
RPRM	reprimo, TP53 dependent G2 arrest mediator candidate
RREB1	ras responsive element binding protein 1
RRM2B	ribonucleotide reductase M2 B (TP53 inducible)
S phase	DNA synthesis
SDS-PAGE	SDS-poly acrylamide gel electrophoresis
SEMA3B	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B
SERPINF1	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5
SESN2	sestrin 2
SFN	stratifin
SGK	serum/glucocorticoid regulated kinase 1

SH3	Src homology 3-like (SH3) domain
SIAH1	siah E3 ubiquitin protein ligase 1
SIPS	stress-induced premature senescence
siRNA	small interfering RNA
Sir2	silencing information regulator 2
SIVA1	SIVA1, apoptosis-inducing factor
SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1
SLC2A4	solute carrier family 2 (facilitated glucose transporter), member 4
SLC6A6	solute carrier family 6 (neurotransmitter transporter, taurine), member 6
SLOS	Smith-Lemli-Opitz syndrome
SOX4	SRY (sex determining region Y)-box 4
SP7	Sp7 transcription factor
SSB	DNA single strand breaks
SSH	Secure Shell
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SV40	simian virus 40
S100A2	S100 calcium binding protein A2
S100A6	S100 calcium binding protein A6
TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
TCR	T-cell receptor
TD	tetramerization domain
TET	tetramerization
TFDP1	transcription factor Dp-1
TGFA:	transforming growth factor, alpha
TGFB1	transforming growth factor, beta 1
THBS1	thrombospondin 1
TIAF1	TGFB1-induced anti-apoptotic factor 1
TK	thymidine kinase

TLR3	toll-like receptor 3
TNFA	tumour necrosis factor
TNFRSF10A	tumor necrosis factor receptor superfamily, member 10a
TNFRSF10B	tumor necrosis factor receptor superfamily, member 10b
TP53I13	tumor protein p53 inducible protein 13
TP53INP1	tumor protein p53 inducible nuclear protein 1
TRADD	TNFRSF1A-assocaited via death domain
TRN	transcriptional regulatory networks
TRS	transcriptional regulatory system
UV	ultraviolet radiation
VEGFA	vascular endothelial growth factor A
VRK1	vaccinia related kinase 1
WWP1	WW domain containing E3 ubiquitin protein ligase 1
XAF1	XIAP associated factor 1
XIAP	X-linked inhibitor of apoptosis
XPC	xeroderma pigmentosum, complementation group C
YBX1	Y box binding protein 1
ZMAT3	zinc finger, matrin-type 3
5-FU	5-fluorouracil

Chapter 1 Introduction

1.1 Systems biology and cancer modelling

1.1.1 General systems biology approaches

With the development of research techniques, biological research has encountered more and more challenges. The traditional approaches mainly based on experimentation, which focus on empirical and isolated descriptions of biological phenomena without the help of mathematics, could not cope with increasing challenges in biological research (Wang, 2010). For instance, these approaches could not capture the dynamic mechanisms of complex molecular pathways which result in cancer. Therefore, the improvement of high-throughput data collection and interpretation is necessary to understand better complex diseases such as cancer (Wang, 2010). The traditional biology approaches were not able to cope with this vast amount of high-throughput data. However, the system biology approach which relies on the combination of experimentation with mathematical modelling and the aid of computational tools, enabled researchers to deal with those complex biological networks and high-throughput datasets. Moreover, with the assistance of computational technologies, risk in clinical trials for the development of new drugs for cancer treatments could be decreased by system biology approaches (Wang, 2010). Because of those promising advantages for cancer research, the system biology approach was used in my project.

In general, available approaches for pathway modelling rely on the systems biology theory. There are two traditional approaches to investigate biological systems: the top-down and the bottom-up approaches. Figure 1.1 shows the

mechanism of these general approaches. Based on large-scale observations of correlated molecular behaviour observed in lab work studies, top-down systems biology identifies molecular interaction networks. On the other hand, the mechanisms through which functional properties arise in the interactions of known components are examined by bottom-up systems biology (Figure 1.1).

The top-down approach generates abstract properties of the system from detailed experimental data that were measured at the genome-scale. An iterative cycle is utilized to generate new molecular mechanisms from experimental data. These data will be analysed and an integration of the data will be obtained from the results of analysis. For example, Guda et al established a probabilistic model to predict domain-domain interaction (DDI) by a top-down approach (Guda et al, 2009). This model considered 5 scoring features and was generated on the basis of a protein-protein interaction data set comprising 2,735 species. The high confidence DDI dataset prediction produced by this model furthered their knowledge of DDI for biomedical research (Guda et al, 2009).

In contrast, the bottom-up systems biology approach starts from equations modelling individual reactions, and assembles them to predict systemic properties. The details of the functional properties can be characterized and each biological process will be modelled using abstract equations in various conditions. For example, Chang et al utilized the bottom-up approach for the research of breast cancer survival (Chang et al, 2005). They modelled the wound-response signature at first, then validated this signature using clinical data of 295 breast cancer patients and confirmed the useful role of this wound-response gene expression signature in the early stage of breast cancer treatment (Chang et al, 2005) .

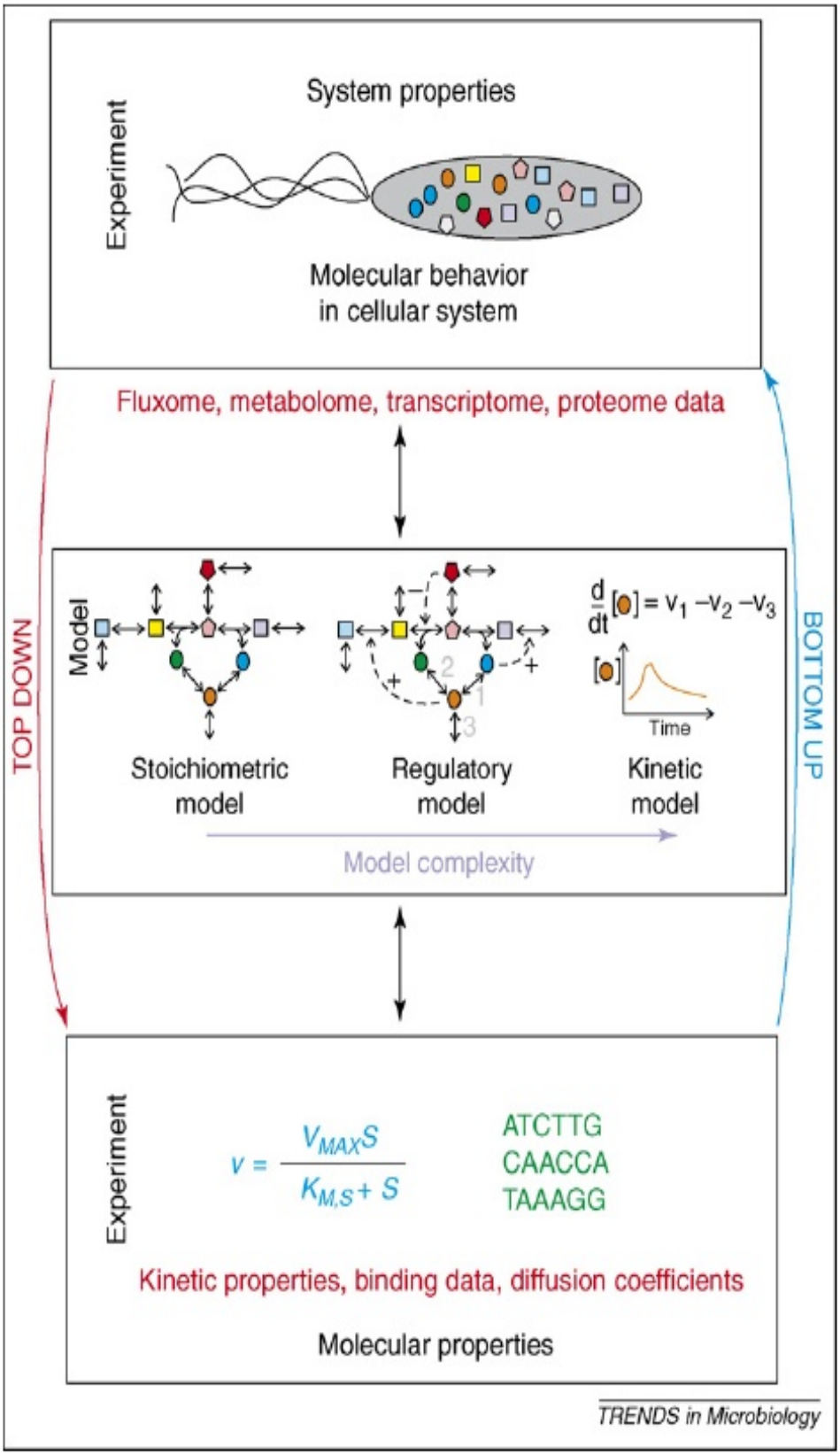

System properties
Experiment
Molecular behavior
in cellular system
Fluxome, metabolome, transcriptome, proteome data
TOP DOWN
BOTTOM UP
Model
Stoichiometric
model
Regulatory
model
Kinetic
model
Time
Model complexity
Experiment
ATCTTG
CAACCA
TAAAGG
Kinetic properties, binding data, diffusion coefficients
Molecular properties
TRENDS in Microbiology

Figure 1.1: The top-down and bottom-up approach to systems biology

In bottom-up systems biology, models are constructed according to general rules which are represented by equations describing the molecular properties. From the models constructed, predictions for the experimental results will be made for the biological system. In contrast, the top-down approach will construct general models according to the experimental data measured. By the use of abstract models, general rules for the biological systems will be represented as equations, which describe molecular properties for the biological systems. Molecular species such as enzymes transcription factors or metabolites are represented as coloured shapes, while reactions are displayed as full arrows and dashed arrows depict regulatory influences (e.g. inhibitory allosteric feedback interactions) (Bruggeman & Westerhoff, 2007).

1.1.2 Overview of modelling techniques

A model usually represents objects or processes in an abstract form, through which their features can be illustrated (Klipp et al, 2009). An object or a process could be represented by different types of models. For instance, the cellular process through which human cells produce proteins could be represented in various forms. This process could be represented by a mental model(Klipp et al, 2009), which conceptually describes how different enzymes work in the researchers' brain. The interactions between proteins produced inside the human cell could be represented by a network model (e.g. protein A activates protein B). The same process could also be represented by a process model in which all chemical processes are represented by equations and listed individually. Moreover, the dynamic mechanisms inside the process could be represented by a dynamic model. As a result, the same biological process or object could be investigated by different modelling techniques. For

example, the effects and relationships between small molecules by chemical reactions in a cell could be investigated at a microscopic level. For a biochemical network, the network structures, dynamics and function can be investigated to analyze the features of the network.

Those biochemical reactions can be represented and illustrated mathematically by stochastic or deterministic modelling approaches. In a stochastic model, the states of the system depend on a probability distribution. Stochastic models treat biochemical reaction systems as random processes, and analyze them by calculating mean values, fluctuations, and correlations of system states (Klipp et al, 2009). Stochastic modelling approaches are more computationally expensive than deterministic approaches, especially for nonlinear systems, in which the steady states of system switched randomly and are determined by probability distributions. These techniques have been widely utilized for biochemical systems. For instance, the complex pathway simulator (COPASI) is implements stochastic modelling approaches that allow simulations and analysis of biochemical networks. Another application of stochastic model was realized by Twycross et al (2010) to investigate the role of auxin, which was important for the plant growth (Twycross et al, 2010). A stochastic model was constructed and simulations were performed by a multi-compartment stochastic P system framework. Auxin movements at the molecular scale were analyzed by the stochastic model simulations. Variability of the auxin-transport system was investigated and the potential extreme behaviours of auxin were detected by simulation results (Twycross et al, 2010).

In contrast to stochastic models, previous or current state values determine the states of variables in deterministic models. Parameters and variables are

utilized to describe the quantities in the system. The difference between parameters and variables is that parameters have a given value but values of variable are changeable over time. In the next sections we describe some of the deterministic models utilized in the research of cancer, such as ordinary differential equation (ODE) models and Boolean network models. In this thesis, deterministic models were utilized to analyze the p53 pathways induced by DNA damage. These models were realized using Boolean networks described in the next sections.

1.1.2.1 Ordinary differential equation (ODE) and partial differential equation (PDE) techniques

Several modelling techniques have already been used for cancer research using the systems biology approach, for example ordinary differential equation (ODE) modelling techniques, partial differential equations (PDEs), Petri nets, cellular automata (CA), agent-based modelling (ABM) techniques, hybrid approaches (Materi & Wishart, 2007), π-calculus (Klipp et al, 2009), and Boolean modelling techniques.

The ODE modelling technique has been the most widely utilized to describe biochemical system kinetics (de Jong, 2002). All states depend on a single variable, and time is usually the dynamic variable in ODE models. The advantage of ODE models is that they are able to provide detailed dynamics of the molecular processes and they can be tested by quantitative experimental measurements (Albeck et al, 2008; Rehm et al, 2006). On the other hand, ODE models necessitate a large number of parameters, many of which are not accessible experimentally. In addition, ODE models become very complex and

difficult to run when there are large numbers of compounds and interactions. The exploration of systems properties is difficult with ODE models because the model space, kinetic parameter space and the initial condition space cannot be systematically spanned (Mai & Liu, 2009).

In contrast to ODEs, PDEs depend on more than one independent variable and are more feasible than ODE modelling techniques in the case of spatial models. For instance, Marciniak-Czochra and Kimmel used PDEs to establish an early tumour development process, which considered a tumour cell population in a linear or tubular structure (Marciniak-Czochra & Kimmel, 2007). These models provided simulation of tumour evolution at an early stage and the cell production was examined by linear or tubular structure models. Two PDE equations were utilized to represent cell growth and growth factor production. The receptor particle production was treated as a Markov process. Possible reasons were inferred to explain why cancer cells grow in an exponential manner and their invasion was shown to be temporarily stable by model simulations (Marciniak-Czochra & Kimmel, 2007).

1.1.2.2 Petri nets

Petri nets represent parallel and discrete systems using a mathematical graphic formalism. They usually consist of places, transitions and arcs in the graph (Klipp et al, 2009). Petri nets have been frequently utilized to simulate metabolic networks. A metabolic network consists of metabolites and reactions (Klipp et al, 2009), which are represented by different components in the Petri net. For example, metabolites in the metabolic network can be represented by places in the Petri net, reactions between metabolites by transitions, and the stoichiometric coefficients by arc weight values. For instance,

Zevedei-Oancea et al (2011) performed a topological analysis of the Trypanosoma brucei metabolic network with the help of the Petri net technique (Zevedei-Oancea & Schuster, 2011). Trypanosoma brucei is a parasite whose cellular structure is similar to other eukaryotes and its metabolism must be understood to develop drugs against African trypanosomiasis (sleeping sickness). Topological analysis of Petri net models were performed to investigate the function of triose phosphate isomerase (TPI) in Trypanosoma brucei metabolism and Petri net techniques showed their efficiency in the simulation of metabolic networks (Zevedei-Oancea & Schuster, 2011).

1.2.2.3 Cellular automata (CA)

Cellular automata (CA) perform simulations for temporal or spatio-temporal processes. The objects in a CA are called cells and connected to their neighbour's cells. The state of a cell changes in a synchronous manner over time and its state at the current time point depend on the states of its neighbour cells at the previous time point (Hasty et al, 2001). Dynamic Cellular Automata (DCA) allow cell movements over time in molecules (Materi & Wishart, 2007). This modelling technique has been widely applied in biological research, for instance, an automata model of enzyme kinetics was established by Kier et al (1996) and this application of DCA modelling played an important role in enzyme kinetics research (Kier et al, 1996). This CA model focused on a reaction between an enzyme and a substrate in water. Predictions that enzymes behave less reactively in the presence of hydrophobic substrates were obtained by model simulations (Kier et al, 1996).

1.2.2.4 Agent-based models (ABMs)

Agent-based models (ABMs) are similar to DCA and agents are utilized to represent metabolites, genes, proteins or even entire cells. However, ABMs do not restrict the spatial grids or synchronize properties (Materi & Wishart, 2007). One application of this technique was implicated in the research for bacterial chemotaxis (Emonet et al, 2005). This model, named as AgentCell, was established to simulate the behaviour of Escherichia *coli* bacteria. Each single bacterium was defined as an agent and the simulations were performed to predict Escherichia *coli* cell behaviours. Experimental data for single cell population behaviour were utilized to validate simulation results and it was found that this agent-based modelling was a powerful technique to study cell-to-cell communications (Emonet et al, 2005).

1.1.2.4 Hybrid approach

The hybrid approach was described to be a mixture of discrete and continuous modelling (Materi & Wishart, 2007; Sorger, 2005). For instance, an application of hybrid approaches was utilized by Osborne et al (2010) for a healthy crypt in solid tumours (Osborne et al, 2010). Multi-scale models were utilized to investigate dynamic mechanisms for both healthy and mutant colorectal crypt invasion. These multi-scale models consist of both a continuous model for the cell-centre and a discrete model for the cell-vertex model. Both cell-centre and cell vertex models treated cells as discrete entities, but parameters represented in their motion equations were determined by continuous changes of time (Osborne et al, 2010). These multi-scale models furthered our understanding of crypts and played an important role in the detection of new cancer therapeutic targets and prediction of their impact (Osborne et al, 2010).

1.1.2.5 π-calculus technique

The π-calculus (pi-calculus) technique describes a process for channel communication. It was applied by Regev et al.(2001) in the research of the RTK-MAPK signal transduction pathway for molecular cancer research (Regev et al, 2001). This pi-calculus model was run by a computer simulation system named PiFCP to perform various bimolecular process simulations for the PTK-MAPK signal transduction pathway (Regev et al, 2001).

Although the above modelling techniques are powerful in certain research fields, the Boolean modelling approach has its own advantages. The most attractive one is that Boolean modelling does not need a large amount of detailed input information and can be established on the basis of available experimental data, which may be incomplete or less quantitative (Wang, 2010). Boolean modelling only requires the identification of interactions, the characterisation of relationships between nodes and the effects of input signals. Boolean modelling facilitates the exploration of the dynamics of a complex gene expression network and focuses on its global features. The details of Boolean modelling will be described in the section 1.1.3.1. As we aim to gain an insight into the interaction network of p53, the Boolean modelling approach is more feasible than other modelling approaches to obtain data on the basis of limited information.

1.1.3 Logical modelling techniques

Although computational simulation technologies have been recently developing at a remarkable speed, the modelling of biological systems with realistic size and complexity using a molecular dynamic approach is still a challenge. As a result, models are usually built at a higher level by removing many details of the biological system. As cancer is regarded as a complex disease, network approaches are often utilized as an abstraction of this complexity. A network approach could help us to investigate a biological system with little information known it (Klipp et al, 2009). Similar to an adage saying:"A picture is worth a thousand words", a network is able to provide a more comprehensive and direct description for the system than a long paragraph of text explanations (Klipp et al, 2009). Moreover, the mathematical relationships in the network could help us further our understanding of the significance of the network structure.

A mathematical graph is utilized to represent a complex network. In the graph, nodes represent genes and edges represent interactions in the cellular network. Analysis of the cellular network topologies is performed to discover the significant features of the network. For example, the degree of a node *i* corresponds to the number of edges connected to this node and the distribution of node degrees (the number of edges connected to this node) can reveal the structure of a network (Klipp et al, 2009). Based on their structure, there are several main types of networks: random, scale-free and small-world networks. The construction of a random network (the Erdös-Rényi model) is made by adding edges with a probability, which is randomly determined. In a random network, degrees of nodes follow a Poisson distribution which means

that nodes in the random network are not highly connected and most of them have the same degree (Barabási & Oltvai, 2004). As a result, the mean path length of a random network depends on the logarithm value of the network size (Barabási & Oltvai, 2004). In a scale-free network, the distribution of node degrees obeys a power law which indicates that a few nodes have high degree and the majority of other nodes have a low degree (Klipp et al, 2009). A small-word network has a clustered structure and small distance for the short path between any two nodes. A clustered structure means that neighbouring nodes are more likely to connect to each other rather than other arbitrarily selected nodes (Klipp et al, 2009). Biological networks often combine the scale free and small world properties. The PKT205 models described in Chapter 5 adhere to the small-word property.

Since various types of biochemical processes are active in a biological cell, different networks are generally utilized to represent different functions of the cellular network, such as metabolic networks, transcription networks and protein-protein interaction networks (Klipp et al, 2009).

Metabolic networks include metabolites and chemical reactions between them to represent metabolic pathways in a cell. They represent the process of transformation from metabolites to their products, where each node represents a metabolite and each edge represents a reaction. The network analysis of metabolic networks focused on the hub nodes and the exchange fluxes. For instance, Wagner and Fell (2001) constructed a metabolic network model for Escherichia coli (E.coli) metabolism (Wagner & Fell, 2001). They analyzed the properties of the metabolic network. They found that it was a small-world network and the metabolite degrees follow a power law (Wagner & Fell, 2001).

.

Transcription networks represent the transcription processes occurring between transcription factors and genes. Each node represents a protein and edges represent transcription processes by which transcription factors bind to a gene. Such networks frequently contain several network motifs such as auto-regulation and feed-forward loop. For instance, Lee et al in 2002 investigated the transcription regulation in the eukaryote Saccharomyces cerevisiae (Lee et al, 2002). 141 transcriptional regulators for yeast cells were selected from database and systematic genome-wide location analysis for transcriptional regulators was performed. 2343 promoter regions were found to bind 106 transcription factors and the yeast regulatory networks were constructed. The function of network motifs in the yeast network was analyzed through small transcription network models. Knowledge about eukaryote evolution was obtained during the network motif analysis. Moreover, it was found that transcription factors facilitated connections between eukaryotic cellular functions (Lee et al, 2002)..

Protein-protein interaction networks represent physical interactions or causal influences between proteins. For instance, Barabási and Oltvai (2004) utilized a protein-protein interaction network (Figure 1.2) to represent interactions in Saccharomyces cerevisiae (Barabási & Oltvai, 2004). This network was obtained by yeast two-hybrid measurements and it was found that a few highly connected nodes hold the network together. The PKT38 models mentioned in Chapter 3 was an example of this protein-protein interaction network.

Figure 1.2: An example of protein-protein interaction network.

It shows protein-protein interactions in Saccharomyces cerevisiae and different colours of nodes represent their phenotypic effect once they were deleted. Red colour means lethal. Green means non-lethal. Orange means slow growth and yellow means that the effect was unknown (Barabási & Oltvai, 2004; Jeong et al, 2001).

1.1.3.1 Boolean modelling

The Boolean network (BN) is a modelling technique used for biological molecular networks. The advantage of Boolean networks is that they are much easier to construct and faster to operate than ODE models. Furthermore, in contrast to ODE models the extensive exploration of the model space can be achieved with a BN. BNs allow the systematic determination of logical steady states or cycles of the biological system. In that, they are able to capture the essential dynamic properties of the biological system. Because of the

complexity of the p53 network and the lack of kinetic parameters necessary for the use of ODE models, these ODE modelling techniques would be extremely difficult to use for p53 pathways (Mai & Liu, 2009). Both problems will be accommodated by the use of Boolean networks and for those reasons we chose BN to model the p53 pathway.

The use of Boolean networks in cancer research has been reported in a few other studies. For example, Schlatter's group(2009) constructed a Boolean network based on literature searches and described the behaviour of both intrinsic and extrinsic apoptosis pathways in response to diverse stimuli (Schlatter et al, 2009). Their model revealed the importance of crosstalk and feedback loops in controlling apoptotic pathways (Schlatter et al, 2009). Rodríguez et al (2012) constructed a large Boolean network for the FA/BRCA (Fanconi Anemia/Breast Cancer) pathway and simulated the repair of DNA ICLs (interstrand cross-links). This model revealed the relationship between the activated DNA repair pathway and defects in the FA/BRCA pathway (Rodríguez et al, 2012).

1.1.3.2 Network formalisms for Boolean modelling

Graph theory is the basis of a network model. There are two different graphic formalisms to represent Boolean networks: interaction graphs and interaction hypergraphs. The basic difference between a graph and a hypergraph is the ability to deal with multiple connections. In the example shown in Figure 1.3 the interaction graph treats the connection between receptor (Rec) and ligand (Lig) as two independent processes (a), while the interaction hypergraph regards the connection as a synchronous process (b), which means that both Rec and Lig are simultaneously needed to activate the receptor-ligand

complex (RecLig*).

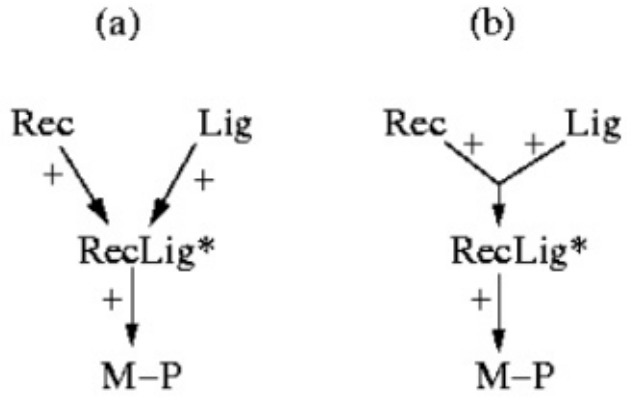


Figure 1.3: (a) Interaction graph and (b) logical interaction hypergraph representation of a simple interaction between receptor (Rec) and ligand (Lig).

This figure shows difference between interaction graph and logical interaction hypergraph (Klamt et al, 2006).

Although interaction graphs can represent feedback loops and pathways between any two molecular species, they cannot capture complex logical relationships in which species are connected by Boolean operators such as “AND”, “OR”,”NOT”. These Boolean operators obey three-valued logic rule and these logical relationships occur frequently in real networks. As a result, interaction hypergraphs are better suited to model signalling networks, since all the arcs whose targets are the same species and which need to interact synchronously can be concatenated. These logical interaction hypergraphs make logical steady state analysis (LSSA) possible (Figure 1.4), as indicated by Klamt et al. (Klamt et al, 2006).

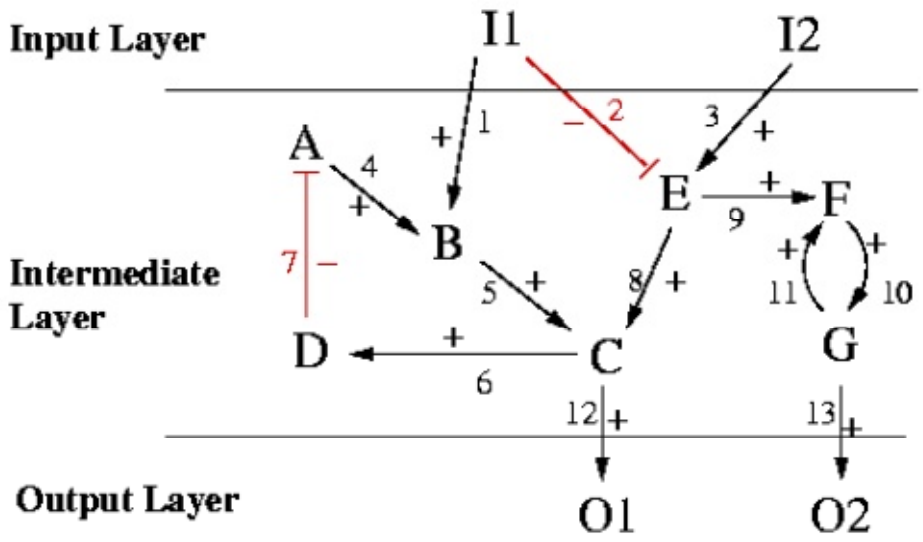


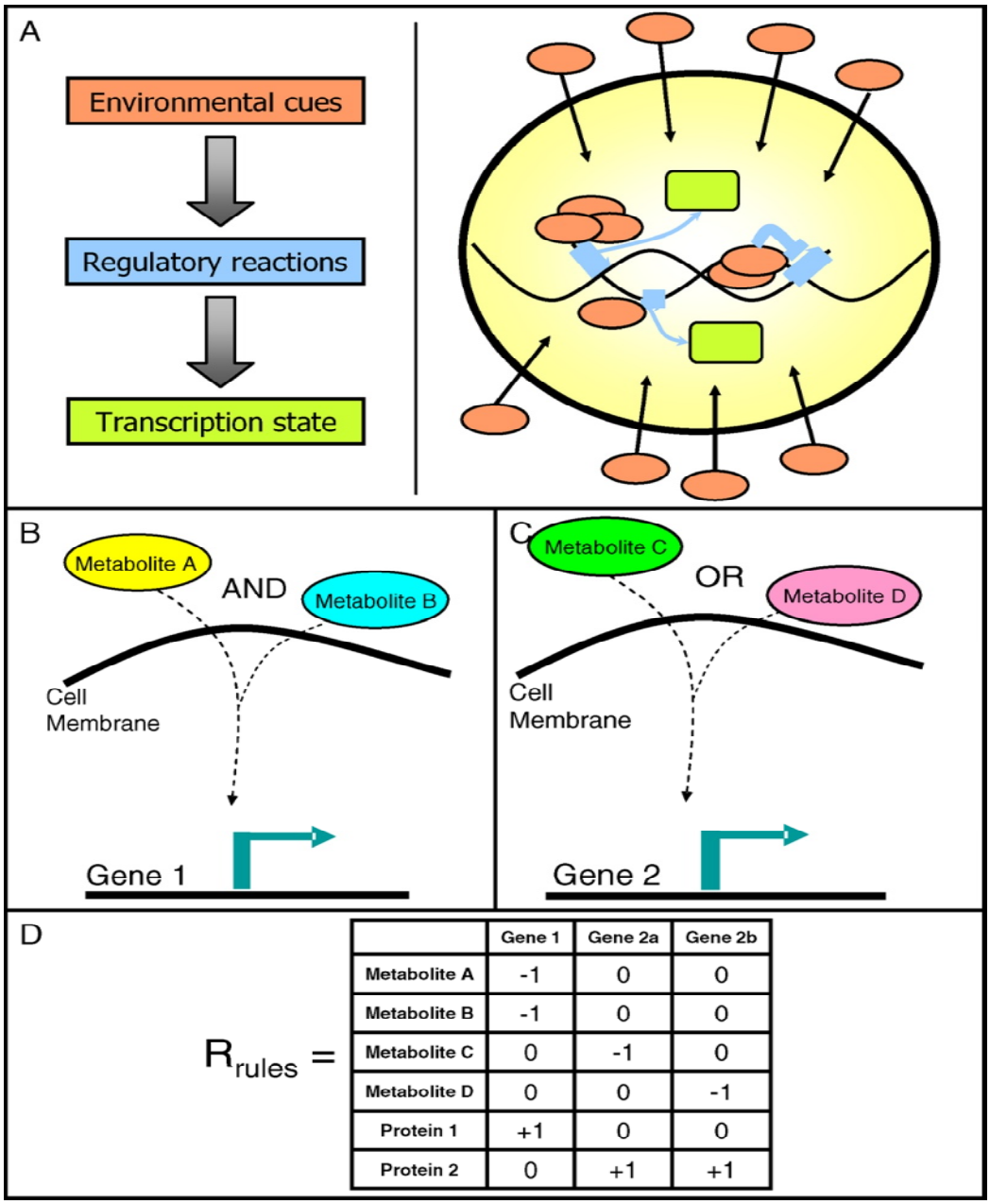
Figure 1.4: Example of interaction graph

Each number represents a reaction in the network. The red arcs represent inhibitions and the black arcs represent activations (Klamt et al, 2006).

An example of successful application of logical hypergraphs in BN was for the T-cell receptor signalling network (Klamt et al, 2006). The interaction hypergraph was utilized to facilitate a structural analysis for the T-cell receptor signalling network. This work provides a methodology to deal with complex mechanisms involving more than two proteins.

1.1.3.3 Matrix formalism for Boolean modelling

Boolean networks can be conveniently represented by matrix formalism. The functional state of genes that are controlled by transcriptional regulatory networks (TRN) can be represented by a Boolean network in which each vector of the Boolean network represents the state of a gene (1 if it is transcribed, 0 if not). External signals can furthermore be linked with a TRN, forming a transcriptional regulatory system (TRS). The transcription regulatory network matrix form was defined as **R** by Gianchandani et al (2006) and it can be used to investigate the functional states of the system extensively (Gianchandani et al, 2006). There are at least three states in the matrix formalism for TRS: -1, 0, and +1. For reactions that occur inside the system, 0 means that it is inactive, and 1 means that it is active. In other reactions which cross the system boundary, -1 means that the row of the reaction flows into the system, +1 indicates a flow out of the system, and 0 means that the production of the component is neither increased nor decreased. Consequently, the relationships between components of a network can be described by a network map. Although a network map cannot calculate functional states, the TRS can achieve this goal. TRS can be represented by in silico expressions so that all the possible environments and consequent systemic interpretations can be enumerated (Figure 1.5).



	Gene 1	Gene 2a	Gene 2b
Metabolite A	-1	0	0
Metabolite B	-1	0	0
Metabolite C	0	-1	0
Metabolite D	0	0	-1
Protein 1	+1	0	0
Protein 2	0	+1	+1

Figure 1.5: Formation of the TRN Matrix

(A) The environmental cues are defined as the inputs, in which the presence and absence of metabolites, reaction fluxes, and specific conditions are included. The output is the transcription state. These outputs are obtained according to the combination of the input with certain rules. (B) shows a situation with the logical relationship AND: Metabolite A **AND** Metabolite B are bound to Gene 1. (C) illustrates a situation where Metabolite C **OR** Metabolite D can be bound to Gene 2, by which, the inputs have a logical relationship OR. (D) depicts the rule matrix for the situation in 7(B) and 7(C) (Gianchandani et al, 2006).

1.1.3.4 Modelling approach using CellNetAnalyzer

CellNetAnalyzer (Klamt et al, 2006) is a computational tool to construct signalling and regulatory logical models and perform functional analysis. This computation tool was utilized by Klamt et al. (Klamt et al, 2006) in the analysis of different types of interaction networks, which may consist of mass flow or signal flow. A metabolic network is represented by a mass-flow (stoichiometric) network. On the other hand, a signalling or regulatory network is better represented by a signal-flow network. Interaction graphs and hypergraphs are supported by CellNetAnalyzer to represent the signal-flow interaction network. An interaction matrix is produced to represent the participating species and reactions. Then feedback loops, crosstalk, signalling pathways of interest, and minimal cut sets can be identified graphically. Moreover, the dependency relationships between any pair of molecular species can be explored using the dependency matrix. Those dependency matrixes are calculated on the basis of the shortest positive or negative pathway distance between nodes (see Methods). The internal changes of the interaction network in response to perturbations can be investigated and predicted by *in silico* knock-out or knock-in tests. The logical behaviours of all nodes of the network can be analyzed by logical steady state analysis (Klamt et al, 2006).

CellNetAnalyzer has been used in several model applications. For instance, a logical model of T-cell receptor signalling was constructed by Klamt et al (2006) (Klamt et al, 2006). This model consisted of 40 nodes and 49 hyper arcs. Two input signals were added in the model, one for TCR (T-cell receptor) and the other for CD4 (CD4 molecule), as well as four output nodes defined by CRE (cAMP response element), API (activator protein 1), NFAT (Nuclear factor of

activated T-cells) and NF-κB. The state of negative feedback loops was predefined to be either activated or inactivated and the dependency matrix was calculated in those two scenarios. According to the difference between these dependency matrices, changes in dependency relationships were revealed. The minimal cut sets which determined all possible pathways from input to output were explored and essential nodes for the whole T-cell activation were highlighted. Logical steady state analysis was also performed for three scenarios at different time scales. CellNetAnalyzer was found to be applicable for the complex T-cell signalling pathway and the analysis results matched current knowledge for the T-cell signalling cascade. A more recent application of CellNetAnalyzer in the research of p53 pathways was a dynamic model of the p53 and NF-κB pathways in response to DNA damage (Plotz & Naumann, 2012). This dynamic model utilized a Boolean network to represent the apoptosis pathway in which p53 and NF-κB were involved. *In silico* knock-out tests of p53 were performed and novel predictions for certain gene functions such as cell cycle arrest, apoptosis and other gene expression levels were obtained. The dependency matrix was also calculated and core sub networks were generated from the full network so as to identify potential target genes for carcinogenesis therapies. Moreover, simulations about particular gene disorders were performed and events which contributed to carcinogenesis caused by activation or inactivation of certain genes were explored. Another application of CellNetAnalyzer was a metabolic model of the Smith-Lemli-Opitz syndrome (SLOS), which consisted of 44 metabolites and 40 reactions. Flux balance analysis (FBA) was performed by CellNetAnalyzer to investigate drugs which exerted effects on the biosynthetic pathway of photosensitivity in SLOS (Eapen, 2007). Through the FBA results, zaragozic acid with cholesterol supplementation was predicted to be a potential treatment target for photosensitivity in SLOS.

1.1.4 Text mining

Data mining plays an important role in biological research (Harmston et al, 2010). For a long time, manually curated protein-protein interactions were regarded as a golden standard to determine effects between proteins and increase the knowledge of signaling pathways. However, manual curation has poor performance in dealing with the rapidly increasing number of publications (Harmston et al, 2010). As a result, there is an increasing requirement to find automated methods to extract information from the substantial numbers of scientific publications in systems biology.

Text mining, which is also known as data mining, automatically retrieves high quality information from text and makes it easier for biologists to retrieve such information from literature. Biomedical text mining is focused on the following areas: named entity recognition, text classification, terminology extraction, relationship extraction and hypothesis generation (Cohen & Hersh, 2005). The aim of named entity recognition is to identify the name of a specific entity from a collection of text. For example, the tumour suppressor p53 protein should be identified as such by text mining, but page numbers (“page 53”), which are shortened as “p53” in the text, should not be misinterpreted as proteins. The recognition of named entities and their normalization are the basis of text mining algorithms, which are able to recognise the labels of target entities and relationships between them. Text classifications determine whether the candidate documents match the topic of interest for the user. Terminology extraction, also known as synonym and abbreviation extraction, solves the problems caused by multiple names of the same entity. For example, PTGS2 (prostaglandin-endoperoxide synthase 2) is also known as COX-2, GRIPGHS,

PGHS-2, PHS-2 and hCox-2. Relationship extraction aims to detect the relationship between a pair of entities according to their co-occurrence in the text. Hypothesis generation aims to predict the relationship between a pair of entities (Cohen & Hersh, 2005). Currently, the most important techniques used in text mining are natural language processing (NLP), information retrieval (IR), machine learning (ML), and statistic and computational linguistics technology (Harmston et al, 2010).

1.1.4.1 Natural language processing technique

Natural language processing (NLP) belongs to the computer science research field and focuses on interactions between human language (natural language) and computers. Although modern natural language processing uses machine learning techniques, machine learning does not cover all fields of natural language processing. For example, EAGi (http://eagl.unige.ch/EAGLi), is a biomedical question answering system that determines a short precise answer for a question by natural language processing techniques (Bauer & Berleant, 2012). For example, if the question “what activates p53” is asked by user, EAGi analyzes it semantically and provides a list of target answers from abstracts in PubMed.

1.1.4.2 Information retrieval technique

Information retrieval (IR) focuses on searching and retrieving information and relies on natural language processing methods. For example, document retrieval is the main task of IR. The retrieval results are limited by unstructured text where information may be hidden in unstructured sentences. However, natural language processing methods can help to solve those problems by stemming word, syntax analysis, semantic parsing, sentence boundary detection and other functions in both the sentence level and whole text (Konovalov et al, 2010; Nadkarni et al, 2011; Thessen et al, 2012).

1.1.4.3 Machine learning technique

Machine learning means that a computer program learns from training data and uses this experience to make predictions for new data. Machine learning applications are widely utilized in system biology text mining. Although some machine learning applications utilize natural language processing for text mining, there are also statistical machine learning applications that do not employ natural language processing techniques. For example, protein-protein relationships can be inferred by detecting co-occurrence of named entities in text, as done by Marcotte et al (2001) who developed a Bayesian model to determine interactions between yeast proteins (Marcotte et al, 2001). They determined the protein-protein interactions between yeast proteins in 2000 Medline abstracts using 80 selected words from 260 trained abstract examples and a likelihood score that could evaluate the probability of selected word co-occurrence in abstracts. Another machine learning application was developed by Stapley et al (2000) (Stapley & Benoit, 2000). They established

a co-occurrence data matrix to measure the dissimilarity of gene pairs that co-occurred in the trusted Medline literature, and extracted them to generate an interaction graph.

1.1.4.4Statistic and computational linguistics technique

Statistic and computational linguistics techniques were utilized to explore information in unstructured text by syntactic or semantic parsing using computational tools (Zhou & He, 2008). Here parsing refers to determining the structure of a sentence or analysing its meaning. Syntactic parsing is to analyse sentences in the text according to grammatical constituents in the sentences and identify syntactic relations. In contrast to syntactic parsing, semantic parsing maps a sentence into a formal representation in term of its meaning. In terms of linguistics complexity, there are two types of parsing methods: partial parsing (shallow parsing) and full parsing (deep parsing). Full parsing focuses on the structure of whole sentences whereas partial parsing focuses on the reliability and efficiency of the syntactic information analysis regardless of the completeness of the sentences. However, those two methods can be combined together for text mining. For example, Santos et al (2005) combined partial and full parsing methods to extract knowledge for the Wnt pathway (Santos et al, 2005). They developed a natural language processing system to analyze Wnt signaling pathway according to 3369 PubMed and 1230 full text papers. Full parsing was utilized to explore protein-protein interactions and partial parsing was utilized to identify name entities.

1.1.4.5Text mining and protein-protein interaction retrieval for p53

Compared with 67,000 papers about p53 published in PubMed and the frequency of new reports about p53 published, our knowledge about p53 was far away from fully understood. The population of published papers about p53 is increasing every year (Table 1.1). It was found that thousands of papers about p53 have been published every year in the last 15 years. There was an increasing strand in the last 15 years (Table 1.1).

Since the information retrieval is time consuming and the conclusion of interaction type is limited by current experimental conditions, there is an urgent requirement for automated retrieval of p53 interaction information by text mining tools or other approaches.

Since there were many types of text mining tools and databases for protein interactions available, it was found that 17 candidate protein-protein interaction databases and provided information about interactions with p53 (Table 1.2).

Year	Publications about p53 per year
1998	2959 (publications)
1999	3200 (publications)
2000	3416 (publications)
2001	3390 (publications)
2002	3462 (publications)
2003	3414 (publications)
2004	3469 (publications)
2005	3435 (publications)
2006	3506 (publications)
2007	3508 (publications)
2008	3663 (publications)
2009	3809 (publications)
2010	3923 (publications)
2011	4210 (publications)
2012	4339 (publications)

Table 1.1: Number of publications about p53 published per year for the last 15 years

Column 1 shows the year. Column 2 shows the publication population about p53 in PubMed in that year.

The name of database	The count of interactions or pathway for p53	The nature of interaction type provided or not?	Data resources	Web address
STRING v9.0	1894	Yes	IntAct, MINT, BIND	http://string.embl.de/newstring_cgi/show_input_page.pl?UserId=jgxXS1js0JHY&sessionId=usaReX3GDvfn
MINT	161	No		http://mint.bio.uniroma2.it/mint/Welcome.do
BIND	2051	No		http://bond.unleashedinformatics.com/index.jsp?pg=0
BioGRID	1162 physical interactions 9 genetic interactions	No		http:// thebiogrid.org/

APID	362	No	IntAct, MINT, DIP, BioGRID, HPRD, BIND	http://bioinfow.dep.usal.es/apid/index.html
HPRD	272	No		http://www.hprd.org/index_html
IntAct	588	No		http://www.ebi.ac.uk/intact/
EMBL-EBI				http://www.ebi.ac.uk/embl/
GeneGo	18 maps	Yes		http://www.genego.com/
KEGG	30 pathway maps	Yes		http://www.genome.jp/kegg/pathway.html
PubGene	2094 relevant term interacts with p53			http://www.pubgene.org/index.cgi

visANT	543	Not all, only 2 transcriptional up-regulation 5 transcriptional down- regulation	COG,KEGG, GeneQuiz	http://visant.bu.edu/
The p53 DATABASE	247	No		http://www.ibibiobase.com/projects/p53/p53/interaction.htm
iHOP	250	Yes, but through literature sentences		http://www.ihop-net.org
PIPs	56	No		http://www.compbio.dundee.ac.uk/www-pips/index.jsps
IARC TP53 Database	272	No	HPRD	http://www-p53.iarc.fr/
Reactome	699 reactions and 501 pathways	Yes	NCBI Gene, Ensembl, UniPort, UCSC, HapMap, KEGG, ChEBI, PubMed, Gene Ontology	http://www.reactome.org

Table 1.2: A survey of candidate text mining tools and databases describing p53 interactions

Table 1.2 lists all potential interaction databases we found on line.

By comparing the performance of these 17 candidate ones, we found that only 5 showed the nature of interaction: STRING, GeneGO, KEGG, iHOP and Reactome. KEGG and GeneGO provided pathway maps about p53. However, maps in KEGG were not updated frequently. Considering the cost of the fee for GeneGO license, we excluded it. The search engine, iHOP is a free tool for text mining of interactions from publications and all sentences included the target gene name listed in the web page; p53 was found to interact with 1558 human proteins in 44277 sentences (12768 abstracts). However, some of these protein-protein interactions were not represented by sentences in iHOP. For example, iHOP found an interaction between p53 and TBP, but this interaction was not clearly justified. Although iHOP highlights the possible verbs representing the interaction with p53, such as enhance, suppress, promote, repress, associate, etc, some of those protein-protein interaction conclusions were still ambiguous due to the variety of language. For example, iHOP found a sentence about the interaction between XIAP (X-linked inhibitor of apoptosis) and p53 reported (Carter et al, 2010), but the context was extremely complex and it was difficult to draw a conclusion for the nature of this interaction.

Although Reactome contained 699 reactions and 501 pathways about p53, some of them are not directly related to p53. For example, “Cell Cycle arrest” is contained in the pathway results, but there were only three terms shown in the pathway browser: “cell cycle/mitotic”, “cell cycle checkpoints”, and

“chromosome/maintenance”. These terms do not reveal any interaction with p53. Another example is that Netrin-4 binds to DCC/UNC5A, but this reaction does not have any direct relationship with p53. For these reasons, Reactome was not considered to be a suitable database to retrieve protein-protein interactions with p53.

On the other hand, STRING illustrated the type of interaction in a list of protein actions and all interactions were represented by a fixed format, which allowed automatic extraction of interaction information. Moreover, STRING has a clearer schema than iHOP to represent the confidence level of protein-protein interactions. The confidence score schema of STRING facilitated the estimation of the confidence quality of these interactions recorded by STRING (Jensen et al, 2009). Another advantage of the STRING database was that STRING provided protein-protein interaction information from three main resources: natural language processing from MEDLINE and other databases, high throughput experimental data and protein-protein interaction predictions. Those three resources included the majority of available approaches to retrieve protein-protein interaction information. The frequency of data update is twice a year, which guarantees that the interaction information is not out of date.

1.2 Cancer

Cancer is a disease which leads to the dysregulation of cell growth. It is caused by environmental factors and gene defects, with the environmental factors contributing to 90% -95% of cancer cases (Anand et al, 2008). As a result, the population of cancer patients increased globally with the changes of life style in the modern world and it was regarded as one of the top causes of death in the human population (Jemal et al, 2011). Currently there are more than 200 different types of cancer (Ainhoa et al, 2007). It was estimated that there were about 12.7 million cancer cases and 7.6 million cancer caused fatalities worldwide in 2008 (Jemal et al, 2011). There are several different treatments depending on the type of cancer but surgery, radiation and chemotherapy have been widely used. However, only 25% of cancer is responsive to treatment based on chemotherapy. As a result, there is an increasing interest in the research that would improve the efficiency of cancer treatment.

In cancer, uncontrolled growth and proliferation of cells leads to the formation of a tumour. There are six functional hallmarks of cancer which are required for the malignancy: growth signal independent of growth factor, evasion of growth suppressor signal, evasion of apoptosis, unlimited replication, stimulation of angiogenesis, induced invasion and metastasis (Hanahan & Weinberg, 2011; Pecorino, 2005). Moreover, there are two emerging hallmarks, deregulating cellular energetics and avoiding immune destruction, that were recently identified by Hanahan and Weinberg (Hanahan & Weinberg, 2011). It was found that thousands of mutated genes were involved in the development of cancers through different genetic pathways (Wang, 2010). The accumulated

DNA mutations and other genetic alternations promoted the transformation from normal cells to the cancer cells (Jaramillo & Tibiche, 2010).

Most of somatic gene mutations occur in two different types of genes. The first type of mutation is found in oncogenes and causes GOF (gain of function) to enhance cell growth, like in the case of Myc (Felsher & Bishop, 1999) and RAS (Bos, 1989). Oncogenes are those genes which when incorrectly activated can cause development of cancer cells. This is usually due to mutations or expression at high levels in cancer cells and because these genes mediate the regulation of cell growth, for instance, ARF (CDKN2A) and MYC. ARF is a protein that can bind MDM2 directly and affects the p53 protein indirectly. The ubiquitination of p53 will be inhibited by the interaction between ARF and MDM2 and p53 protein will accumulate as a result of the binding (Vousden, 2000). MYC is involved in the cell proliferation process and apoptosis (Li & Hann, 2009).

The second type of mutation occurs on tumour suppressor genes and in this case both copies of the gene must be inactivated for the full effect to take place. There are several examples of tumour suppressor genes, the most studied are retinoblastoma mutation causing a tumour of the eye and p53 tumour suppressor loss of function in cancer cells. The p53 protein has been regarded as “a Guardian of the Genome”(Lane, 1992; Levine, 1997).

1.3The p53 tumour suppressor

Since the discovery of p53, mutations of p53 were found in more than 50% of human tumours (Soussi, 2000). The p53 has a crucial role in cancer development and therapy. It is considered in the development of potential strategies utilized in clinical treatment of cancer, such as the isolation of p53 target proteins, genes which regulate the expression of p53 or certain modules in the p53 pathway (Gudkov, 2005). The first chemotherapy based on recombinant adenovirus encoding the p53 target gene, rAd-p53 was approved in the treatment of head and neck cancer in China in 2004 (Peng, 2005). There are two homologues of p53: p63 and p73, which also act as tumour suppressor. Mutation and lack of p63 and p73 were found in many tumour cells. They induce apoptosis in a similar way as p53 (Jung et al, 2001). TAp63 and TAp73 are two full length transactivation isoforms of p63 and p73, which were found to transactivate BAX, PUMA and NOXA as well as p53 (Melino et al, 2004; Mundt et al, 2010).

In response to a variety of cellular stresses, activated p53 leads to cell cycle arrest and cell death (Khoury & Bourdon, 2011). Those pathways were stimulated by several cellular stress signals, for instance, the phosphorylation of p53 by ATM in response to DNA damage. These post-translational modifications, protein-protein interactions and protein stabilization are found to be crucial levels of control of the activity of p53. Moreover, p53 is involved in various pathways either caused by normal physiological response to growth factor, or abnormal oncogenic stimuli; for example, Myc oncogene activation can affect p53 function by the basic-helix-loop-helix recognition motif to induce the p53 promoter (Reisman et al, 1993). In the early stage of p53 studies, p53 was suspected to be an oncogene (Lane & Benchimol, 1990), but now it was confirmed that p53 is a tumour suppressor (Levine & Oren, 2009).

1.3.1 Introduction to the p53 protein

1.3.1.1 P53 structure

The tumour suppressor protein p53 is a tetramer consisting of four polypeptide chains (Joerger & Fersht, 2010) and has a short half life (Giaccia & Kastan, 1998). There are 393 amino acids in human p53 (Figure 1.6). All of its amino acids are assembled into five structurally and functionally different domains: Transactivation domain (1-42), SH3 (Src homology 3-like (SH3) domain) (63-97), DNA binding domain (102-292), TET (tetramerization) domain (323-356) and REG (regulatory) domain (363-393) (Anderson & Appella, 2002) (Figure 1.6). The transactivation domain is involved in transcriptional regulation, mediates interactions with numerous proteins involved in transcriptional control and is intensively modified by phosphorylation (Figure 1.7). The interaction between p53 and other proteins with the SH3 domain may be regulated through protein ligand which contains the SH3 domain. For instance, the c-terminus of p53BP2 (tumour suppressor p53-binding protein 2), which contains the SH3 domain, bind to p53 to promote apoptosis (Samuels-Lev et al, 2001). The DNA binding domain is located in the centre and mediates specific DNA binding to p53 target genes. It was reported that 95% of mutations in p53 occurred in the DNA binding domain and those mutations lead to the loss of p53 wild type function (Sigal & Rotter, 2005). The tetramerization domain (TD) of p53 is important for p53 binding and the oligomerization of p53 is allowed by it. It was reported that TD exerted effects on the strength of interaction and enhanced DNA binding (Chène, 2001). The presence of residues from 334 to 354 in TD was essential for the phosphorylation of p53 by Chk1 (Shieh et al, 2000). Multiple post translational modifications occurred in the regulatory region of p53, which was also called

the C-terminal Basic Region. For instance, Sir2 (silencing information regulator 2) deacetylated p53 at position 382 in the regulatory region of p53 in vivo and yeast Sir2 was involved in double strand DNA repair (Langley et al, 2002).

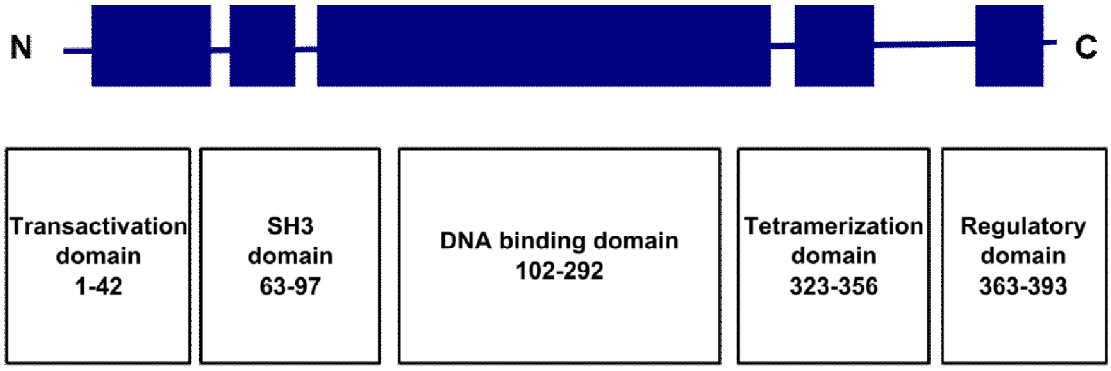


Figure 1.6: Representation of the domain structure of the p53 protein

This figure shows protein domains in p53 according to their known function.

1.3.1.2 Levels of control of p53 activity

Multiple protein complexes interact with p53 at different regions (Figure 1.7). The activation and stabilization of p53 includes several modifications, such as phosphorylation, acetylation, and ubiquitylation.

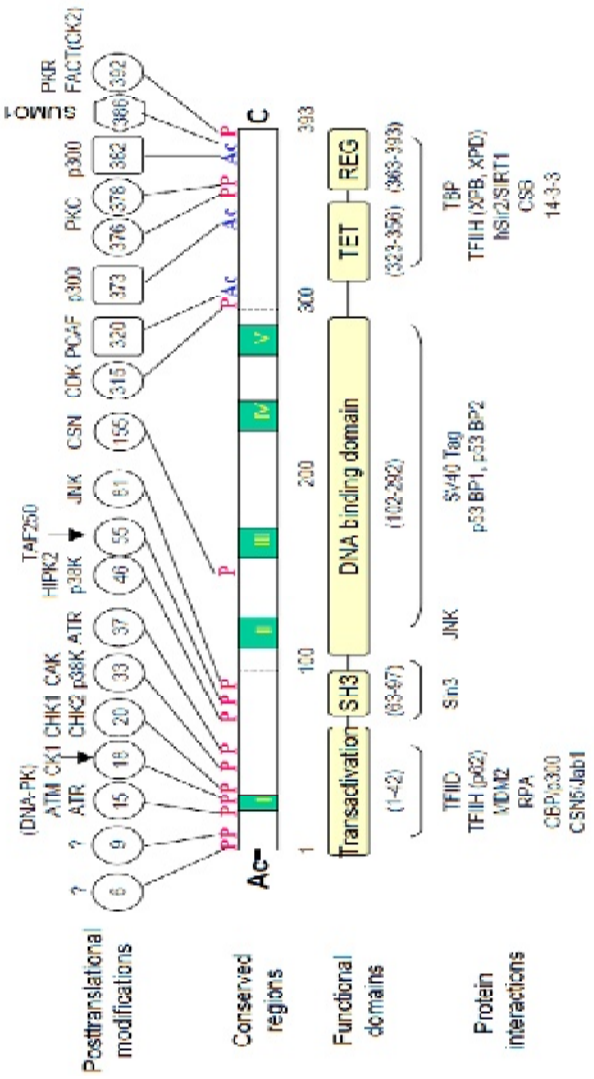


Figure 1.7: Domains and posttranslational modifications of the p53 protein

Different posttranslational modifications of p53 occur on different sites of p53. Five regions of p53 are shown: Transactivation domain (1-42), SH3 (Src homology 3-like (SH3) domain (63-97), DNA binding domain (102-292), TET (tetramerization) domain (323-356) and REG (regulatory) domain (363-393) (Anderson & Appella, 2002).

Those modifications are a major level of control and occur at different serine, threonine, and lysine residues of p53 (Levine et al, 2005). The phosphorylation of p53 occurs at the N-terminal domain, which is mostly induced by DNA damage and mediated by several kinases such as p38, JNK

(Jun NH2-terminal kinase), ATM (ataxia telangiectasia mutated), and ATR (AT and Rad3-related.). For instance, ATM phosphorylates p53 at serine 15 (Banin et al, 1998), CHEK1 (Chk1) phosphorylates p53 at serine 20 (Zhao & Piwnica-Worms, 2001). However, some phosphorylations also occur at other amino acid residues of the C-terminal and the mechanism of control is still unknown.

The post-translational modifications of p53 play a critical role in both p53 stabilization and p53 activation. For example, the phosphorylation of serine 20 by CHK1 and CHK2, which is a response to ionizing radiation, exerts effects on the interaction between p53 and Mdm2. This phosphorylation on serine 20 of p53 plays an important role in the stabilization of p53 (Unger et al, 1999). It has been reported that ATM and ATR phosphorylation on serine 15 of p53 and the phosphorylation on serine 20 of p53 by Chk1 and Chk2 inhibits the MDM2-p53 interaction and the p53 degradation caused by MDM2, resulting in increased p53 stability in response to DNA damage (Shieh et al, 1997).

The acetylation of p53 is another important modification for p53 which was found to stimulate the transactivation of the target genes which are downstream of p53. What is more, acetylation plays an important role in the regulation of p53. Those target genes mostly acetylated p53 at its C terminal, for example, p300/CBP mediates the acetylation of p53 at lysine 370, 372, 373, 381 and 382, and PCAF acetylates p53 at lysine 320 (Liu et al, 1999; Sakaguchi et al, 1998; Wang et al, 2003).

The phosphorylation and acetylation of p53 are interdependent and these modifications are able to activate p53 and induce several DNA events, such as cell cycle arrest, apoptosis, DNA repair, the inhibition of angiogenesis and

other signalling pathways.

P53 activity also depends and is regulated through DNA binding. The p53 binds to a consensus DNA sequence, in which two half-sites of sequence 5'-Pu-Pu-Pu-C-(A/T)-(T/A)-G-Py-Py-Py-3' are included. This sequence is separated by 0 to 13 base pairs (Pu in this sequence represents purine. A represents adenine, T represents thymine, and Py represents pyrimidine). The loss of p53 affinity is caused by delicate changes in those motifs (el-Deiry et al, 1992). The p53 protein acts as a transcription factor and is able to both activate and inhibit transcriptions. For example, p53 acts as transcription activator of p53AIP1 once p53 was phosphorylated at serine 46 in response to DNA damage. The p53AIP1 gene functions as a pro-apoptotic gene and induces apoptosis (Oda et al, 2000). Meanwhile, p53 represses the transcription of Bcl2, which prevents the process of apoptosis.

It should be recognised that the p53 modifications are inter-dependent and regulation in one domain can profoundly influence other domains. Posttranslational modifications can also control interactions with other macromolecules that are of great significance. In addition, p53 interacts with a wide range of proteins, facilitating both the regulation of p53 activity and the control of its concentration by the control of its degradation (Prives & Hall 1999).

1.3.2 Structure of the p53 network

As we mentioned before, the mutation of oncogenes and tumour suppressor genes result in cancer and those genes constitute a complex network through cellular signalling pathways. The p53 tumour suppressor plays an important role in the control of DNA damage in response to various cellular stress signals. These signals include DNA damage, oncogene activation, hypoxia, nucleotide depletion, and depending on the strength of the damage could lead to p53-mediated cell cycle arrest, apoptosis, angiogenesis, DNA repair, senescence and other cellular responses (Moll & Concin, 2005). These p53 pathways are very complex and can be divided into several layers (Figure 1.8): the upstream layer of p53, in which DNA damage, aberrant growth signals oncogene activation and cell stress are included; the central layer where p53 is regulated by MDM2, and the downstream layer where cell cycle arrest, apoptosis, DNA repair and inhibition of angiogenesis are included. The central protein-protein interaction is the p53 interaction with Mdm2. This p53-MDM2 feedback loop is a key signalling pathway to p53. Moreover, p53 receives and integrates these numerous upstream signals and translates them into a cellular response by regulating downstream target genes, resulting in growth arrest and apoptosis.

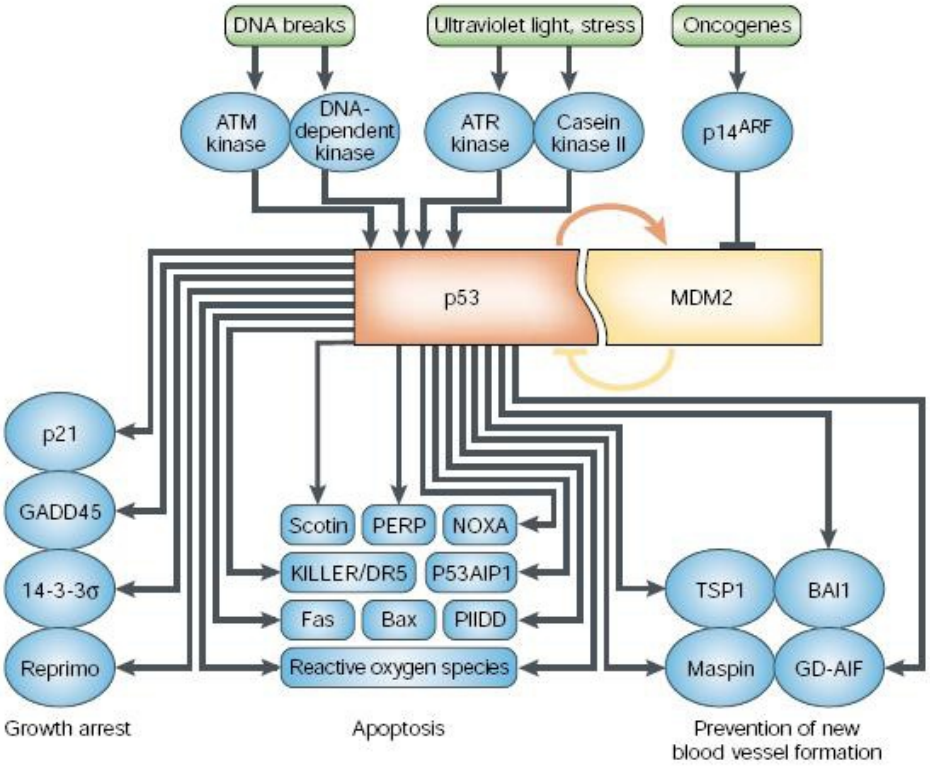


Figure 1.8: Structure of a simplified p53 network

This figure shows a simplified representation of the p53 network, which is composed of three layers: the upstream of p53 layer contains for example extracellular signals such as ultraviolet light that triggers ATR activation. Then, these signals affect p53 which is the core of the network, with the regulation of p53 by MDM2 at the centre. Then the downstream layer where the target genes p21, GADD45, 14-3-3sigma and processes like apoptosis and growth arrest are included (Hasty et al, 2001; Vogelstein et al, 2000).

1.3.3 Upstream molecular pathways of p53 activation

The half life of the tumour suppressor protein p53 is 6 to 20 minutes in cells(Levine et al, 2005). However, once p53 was induced by cellular stress signals, such as DNA damage, hypoxia, nitric oxide signalling or oncogenes, the half life of p53 increases and the p53 protein is accumulated (Levine et al, 2005). The p53 protein is activated by different types of stress signals such as

UV (ultraviolet radiation) and IR (ionizing radiation), which cause DNA damage. The DNA damage is detected by stress responsive kinases, which deliver the stress signals to p53. There are three independent molecular pathways utilized to signal cellular distress: DNA damage, oncogene activation, and other types of cell stress like hypoxia (Figure 1.8). As previously mentioned, phosphorylation, acetylation and sumoylation occur on more than 20 sites of p53 and those modifications lead to a cellular response, for instance, apoptosis, cell cycle arrest and other cellular outcomes (Anderson & Appela, 2005). Globally, these stress signals inducing p53 activation can be classified into two categories: genotoxic stress and non-genotoxic stress. Genotoxic stress includes UV and IR, while oncogenes induction, hypoxia, osmotic shock and microtubule disruption constitute the non-genotoxic stress signals (Anderson & Appela, 2005).

1.3.3.1 DNA damage

There are several types of DNA damage, for instance DNA double-strand breaks (DSBs) caused by ionising radiation (IR) and DNA single-strand breaks (SSBs) caused by ultraviolet radiation (UV). DBSs activate specific stress response kinases such as ATM. This activation results in several phosphorylation events on p53, for instance ATM phosphorylates p53 directly at serine 15 in response to DSBs and increases p53 protein stability. There are several kinases that are members of the same family that respond to DNA damage including ATM, ATR (Figure 1.9) and DNA-PK. After DNA damage, the p53 phosphorylation will lead to its activation and increase in protein stability. The regulation of cell growth and cell death also involves p53 (Prives & Hall 1999). It was found that ATM was activated with exposure to DNA damage and ATM dependent phosphorylation of p53 occurred on serine 9, serine 20, serine 46 and threonine 18 (Anderson & Appela, 2005). ATR is

another important kinase, which functions upstream of p53. ATR is found to be induced in response to UV light or other chemical agents. The activated ATR phosphorylates p53 directly at serine 15 in vitro as well as ATM (Anderson & Appela, 2005). ATM is a Ser/Thr protein kinase (Banin et al., 1998; Canman et al., 1998; Khannak et al., 1998) and plays a key role in DNA DSB (double-stand break) repair. The difference between ATM and ATR is that ATM is mostly responsive to IR whereas ATR responds to UV. Both ATM and ATR kinases phosphorylate p53 as well as other kinases such as Chk2 and Chk1 respectively and these kinases then target p53 as described above.

Moreover, in addition to the regulation of p53 protein stability and p53/MDM2 interaction, posttranslational modifications affect the target gene selectivity of p53. For example, DYRK2 phosphorylates p53 at serine 46 and this phosphorylation leads to the activation of p53AIP1, which promotes apoptosis (Taira et al, 2007). Furthermore, the pro-apoptotic gene BAX is induced once p53 was acetylated by PCAF at lysine 320 and by p300 at lysine 373. Either of those two acetylation result in the activation of CDKN1A (p21) (Roy & Tenniswood, 2007). The ubiquitylation of E4F1(E4F transcription factor 1) at lysine 320 of p53 induced CDKN1A, CCNG1 (cyclin G1) and GADD45A to block the cell cycle (Murray-Zmijewski et al, 2008). Taken together, these reports indicate that posttranslational modifications on different sites of p53 lead to the induction of different p53 target genes and therefore cause different biological processes.

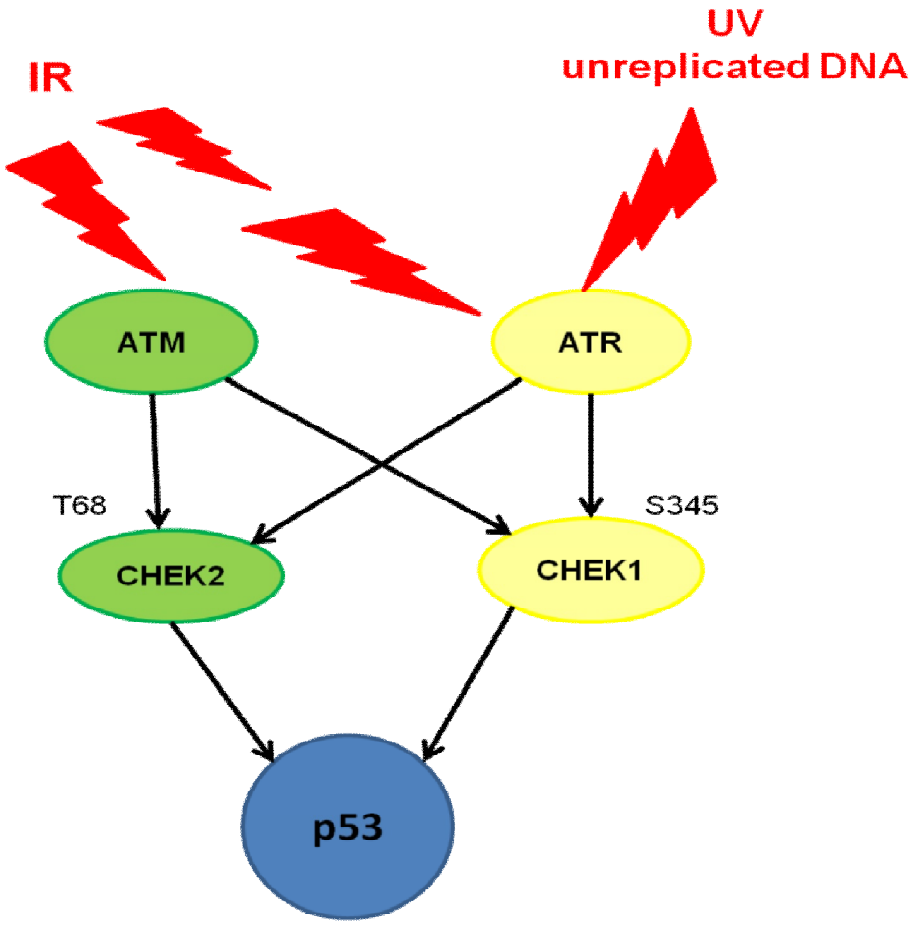


Figure 1.9: DNA damage induced pathway.

ATM and ATR are induced by DNA damage stress and stimulate p53 phosphorylation directly and through CHEK1 (Chk1) and CHEK2 (Chk2). This figure was adapted from Donzelli's paper (Donzelli & Draetta, 2003)

Moreover, there exist other genes which function as upstream regulators of p53, such as AURKA (Aurora kinase A), PRKDC (DNA-PK) and MAPK1 (p38MAPK) in response of DNA damage. AURKA was reported to phosphorylate p53 at serine 315 and this phosphorylation promoted MDM2 ubiquitination on p53, causing the degradation of p53. The down-regulation of AURKA results in G2/M cell cycle arrest through its inhibition on p53 through MDM2 (Katayama et al, 2004). PRKDC functions as an upstream regulator of

p53 and phosphorylates p53 at serine 15 and serine 37 in vitro, repairing the degradation of p53 caused by MDM2 inhibition (Shieh et al, 1997). The p53 protein was also phosphorylated by p38MAPK at serine 38 and serine 46 in response to UV light (Bulavin et al, 1999). If the p53 mutation occurred on these sites with the exposure of UV light, the apoptosis process regulated by p53 were prevented (Anderson & Appela, 2005)

1.3.3.2 Oncogene activation

The stabilization of p53 can be induced by several oncogenes through ARF. ARF is a protein that can bind MDM2 directly and affect the p53 protein indirectly. The ubiquitination of p53 will be inhibited by the interaction between ARF and MDM2, and the p53 protein will accumulate as a result of the binding. The ARF level and p53 level is very low in normal cells. There are diverse ways to activate ARF, for example by dentin matrix acidic phosphoprotein 1(DMP1) and E2F1, which are transcription factors that regulate ARF gene expression (Vousden, 2000). Moreover, ARF stabilizes p53 by repressing MDM2 and promotes c-Myc triggered apoptosis in both p53 dependent and p53 independent manner (Gregory et al, 2005). Another oncogene is the RAS family, which contains HRAS (v-Ha-ras Harvey rat sarcoma viral oncogene homolog), KRAS (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), and NRAS (neuroblastoma RAS viral (v-ras) oncogene homolog). Activated RAS was found to promote cell proliferation through the RAF-MEK-ERK signalling pathway. This pathway mediated MDM2 regulation in the absence of ARF to suppress p53 and played an important role in p53 dependent apoptosis (Ries et al, 2000). Moreover, it was found that p53 could be activated by oncogenic RAS through the MAP kinase signal pathway and induced cellular senescence (Ferbeyre et al, 2002).

1.2.3.3 Other types of stress

Other cellular stresses such as hypoxia and total lack of oxygen (anoxia), as well as oncogene activation, induce p53 activation by less known mechanisms (Gallagher et al, 2006; Hubert et al, 2006). Hypoxia results in CDKN1A (p21) (Koshiji et al, 2004) and CDKN1B (p27) (Gardner et al, 2001) regulated cell cycle arrest and the expression of CDKN1A and CDKN1B is mediated by HIF1A (Goda et al, 2003). Moreover, p53 mediates cell senescence through interaction with HIF1A (Welford & Giaccia, 2011). In addition, p53 induces different signalling pathways in response to different levels of oxidative stress. In low levels of oxidative stress, p53 promotes C12orf5 (TIGAR) and results in the accumulation of NAPDH to repress ROS (reactive oxygen species) levels whereas p53 induces BAX and PUMA to promote apoptosis in response to a high dose of oxidative stress (Liu & Xu, 2011). There are other types of stress mediated p53 activation, such as oxidative stress (Liu & Xu, 2011), nitric oxide ribonucleotide depletion (Linke et al, 1996), DNA replication stress (Burhans & Weinberger, 2007), mitotic apparatus dysfunction (Aylon et al, 2006), telomere erosion (Preto et al, 2004) and nutrition depletion (Wanka et al, 2012).

1.3.4 Regulation of p53 by MDM2

In normal cells, p53 can inhibit the growth of damaged cells and its activity is tightly controlled. According to this phenomenon, the potential cancer cells that will mutate can be controlled through the activity and expression level of the p53 protein. An important finding is that the level of p53 protein is regulated by the interaction with a particular protein, MDM2 (Dimitriadi et al, 2008). MDM2 is an E3 ligase whose targets are p53 and itself for ubiquitination. The function of p53 is regulated by MDM2 and vice versa

forming a negative feedback loop. There are two promoters on the MDM2 gene: P1 and P2, and the transcription from P2 is under the control of p53 (Dimitriadi et al, 2008). Depending whether DNA damage is present or not, MDM2 is able to regulate the response of p53. If DNA damage is absent, MDM2 will bind to the N-terminal of p53 to promote p53 degradation by complex mechanisms. In the presence of DNA damage, both MDM2 and p53 will be phosphorylated and p53 will up-regulate MDM2 transcription.

As a result of this mechanism, the p53 protein level is suppressed by MDM2. If p53 responds to DNA damage, hypoxia, oncogene activation and other cellular stresses, the p53 protein will be phosphorylated and stabilization of p53 will occur. The interaction between p53 and MDM2 will be blocked as well. If the levels of p53 become higher, p53 will initiate a cell death, cell-cycle arrest or senescence program. It was found that p53 can be exported from the nucleus to the cytoplasm, and this will lead to the ubiquitination of p53 (Momand et al, 2005). The p53 proteins that are ubiquitinated by MDM2 will be targeted by the proteasome and degraded (Momand et al, 2005)

The p53-MDM2 feedback loop is a key signalling pathway to p53. In addition to being activated by p53, the MDM2 gene is also transactivated independently by the Ras-driven RAF-MEK-MAP kinase pathway. However, the RAF-MEK-MAPK pathway influences the level of p53 by activating expression of the MDM2 inhibitor ARF. Therefore, it is not surprising that Ras activates the ARF-p53 pathway to suppress epithelial cell transformation. The oncogenic protein Myc also activates the ARF-p53 pathway (Lin & Lowe, 2001). In addition, Ras can also activate PML(promyelocytic leukaemia), which cross-talks with p53 to increase transactivation of specific genes and recruits p53 to nuclear bodies (Bargonetti & Manfredi, 2002; Vogelstein et al,

2000). Since the protein-protein interaction can be disrupted by the peptide, biologists focused on the peptide acting on the interaction between MDM2 and p53 in order to find potential anticancer targets. A D-peptide inhibitor of the p53-MDM2 interaction was identified by Liu et al. and this D-peptide functioned as an activator of p53 (Liu et al, 2010). Another peptide, named as MIP (MDM2 Inhibitory Peptide), which binds to MDM2, was found by Shiheido et al. It represses the MDM2-p53 and MDMX-p53 interactions to activate p53 efficiently (Shiheido et al, 2011).

1.3.5 Molecular pathways downstream of p53 leading to diverse cellular effects

The p53 protein exerts its tumour suppressor effects by inducing the expression of its target genes to prevent carcinogenesis by two main processes in response to DNA damage: p53 promotes apoptosis through the regulation of its target genes or stimulates DNA repair during the cell cycle(Smith et al, 2000). However, p53 is also involved in other cellular processes such as antioxidant defense, energy metabolism, stem cell renewal, neurodegenerative diseases, reproduction endosome/exosome production and so on(Feng & Levine, 2010). In this section, we mainly focus on those effects induced by target genes of p53: the inhibition of cell cycle, which is also known as cell cycle arrest, DNA repair, apoptosis, repression of angiogenesis and cellular senescence and will describe them one by one in the following paragraphs.

1.3.5.1 Control of cell cycle arrest

The cell cycle consists of four phases: G1 phase (Gap 1), S phase (DNA synthesis), G2 phase (Gap 2) and M phase (mitosis) (Figure 1.10). The cell starts to increase in size in the G1 phase. The S phase contains DNA synthesis, during which DNA replication occurs. Once the S phase has terminated, the cell enters into the G2 phase and increases its size. Then in the mitosis phase cell growth terminates and division starts. There exist positive feedback loops among the retinoblastoma protein (Rb), E2F1 and Cyclin E-CDK2. Rb is inhibited by the complex CDK4/6 and Cyclin D1 through the phosphorylation of Cyclin D1. Rb phosphorylates E2F1 to disturb its accumulation. E2F1 activates the Cyclin E and CDK2 complex, which represses Rb. Meanwhile E2F1 induces G1/S genes such as TK (thymidine kinase) and DHFR (dihydrofolate reductase) to trigger the G1-S cell cycle transition (Levine et al, 2005). In order to repair the DNA damage before the next round of replication, cell cycle arrest in response to DNA damage is caused by activated p53. If the DNA lesion was repaired successfully, the damaged DNA will not be replicated and passed on to daughter cells (Pecorino, 2005). The p53 protein mediates the growth arrest through its target genes, for instance, CDKN1A (p21), 14-3-3σ, and GADD45 (growth arrest and DNA damage-inducible protein 45). CDKN1A (p21) regulates the cell cycle in the G1 phase through multiple mechanisms, for instance p21 regulates cell cycle arrest through inhibition of cyclin-CDK4 and cyclin-CDK6. But it also inhibits PCNA to prevent DNA synthesis (Rousseau et al, 1999).

The DNA damage activates p53 through activation of ATM/ATR kinases and DNA-dependent protein kinases, and then p53 increases the gene expression of p21, and p21 binds to cyclin-CDK complexes to induce cell cycle arrest

(Coqueret, 2003). $p21^{WAF1/CIP1}$ is a protein regulated by p53 that can inhibit the activity of CDK. It was demonstrated that p21 is needed for the p53-dependent G1 checkpoint following genomic damage and p21 can mediate the action of p53 in the G1 phase (Walaman, 1995). In addition, 14-3-3σ is induced by p53 as a response to DNA damage and results in blocking the G2 phase. Gadd45 is another p53-regulated protein and it was identified by Kastan and colleagues that wild type p53 binds to a conserved element in the Gadd45 gene so as to control cell cycle arrest that follows DNA damage (Kastan, 1992). The p16 protein represses the complex formed by Cyclin D1, CDK4 (cyclin-dependent kinase 4) and CDK6 (cyclin-dependent kinase 6) as well as CDKN1B (p27). It was reported that p16 and p27 were induced by p53 and they resulted in blocking cell cycle progression (McConnell et al, 1999)

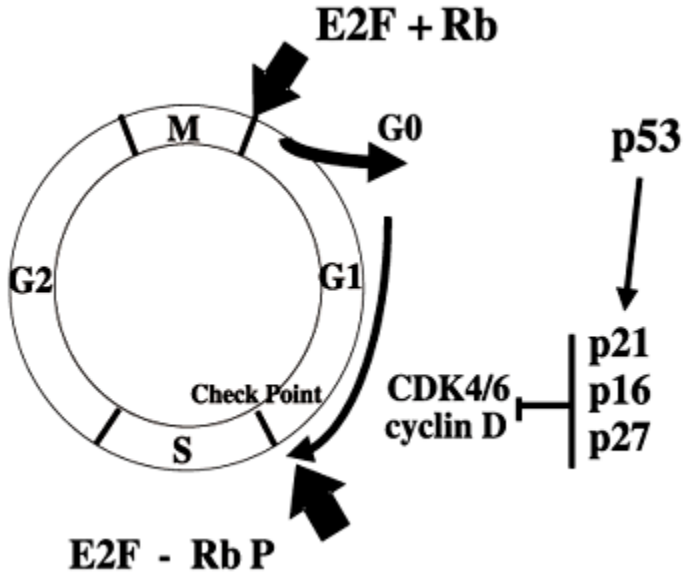


Figure 1.10: Cell cycle phases and p53.

This figure shows the process of cell cycle and p53 regulating the cell cycle arrest through its target gene p21 (Nita et al, 2002).

1.3.5.2 Control of DNA repair

DNA damage lesions are strongly implicated in carcinogenesis if they are not repaired before the division of a cell with mutated DNA. There are five different types of DNA repair: one-step repair, nucleotide excision repair, base excision repair, mismatch repair and recombination repair (Pecorino, 2005).

The p53 gene plays an important role in the process of nucleotide excision repair (Smith et al, 2000). The gene XPC (xeroderma pigmentosum, complementation group C), which is involved in nucleotide excision repair, is a well studied case. XPC gene has a p53 response element in its promoter that renders this gene subject to p53 control. However, if the damage cannot be repaired, the cell will die. It was found that p53 can inhibit a cellular factor which replicates DNA (Dutta et al, 1993). This conclusion was reached according to the interaction between p53 and the simian virus 40 (SV40)-encoded protein T antigen (Levine et al, 2005). Many checkpoint functions for the cell cycle are executed by p53. Martinez et al. reported that wild type p53 protein can block the replication of cells at the G1/S border so that the cell can repair the damaged DNA (Martinez et al, 1991). XPC is involved in nucleotide excision repair (NER) and regulated by p53 through BRCA1, which is a breast cancer tumour suppressor. BRCA1 was found to be repressed by p53 in p53 wild type cells and induces XPC and DDB2 (damage-specific DNA binding protein 2, 48kDa), which are global genomic repair (GGR)-specific damage recognition genes (Adimoolam & Ford, 2003). Moreover, GADD45, which is a target gene of p53, was also found to contribute to the NER process (Smith & Kumar, 2010). It was also found that p53 induces the base excision repair (BER) process in vitro through the AP endonuclease (APE) and DNA polymerase β (Pol β) (Zhou et al, 2001).

1.3.5.3 Control of apoptosis

Apoptosis is a form of cell death distinct from necrosis and autophagy. It has been regarded as a tumour suppression process (Levine & Oren, 2009) and this process is important for the elimination of tumour cells with damaged DNA (Debridge et al, 2012).

There are two different types of signals that induce apoptosis: intrinsic and extrinsic signals (Klipp et al, 2009). DNA damage, oxygen stress, nutrient and other cellular stress signals are intracellular signals. The extrinsic signals are induced by the death receptor ligand, for example TNF (tumour necrosis factor), which is a member of the tumour necrosis factor protein family. As a result, there are two main pathways to commence the apoptosis process (Figure 1.11) (Debridge et al, 2012; Mai & Liu, 2009). The intrinsic pathway includes the regulation of Bcl2, which is repressed by p53 target genes like BBC3 (PUMA, p53 unregulated modulator of apoptosis). Those genes repress the expression of Bcl2 family members, for example Bcl2. The Bcl2 down regulate the pro-apoptotic genes BAX and BAK, which stimulate cytochrome c release to enhance the apoptosis outcome through the activation cascade of caspase 9 to caspase 3, caspase 6 and caspase 7. The extrinsic pathway is altered by extrinsic signals from the death receptors ligand and is also called death receptor pathway. In this pathway, TNF (tumour necrosis factor receptor) induces the expression of TRADD (TNFRSF1A-assocaited via death domain). The activated TRADD promotes the induction of caspase 8, which induces apoptosis through the activation of caspase 3, 6 and 7. Those two signalling pathways cross talk through caspase 3, 6, and 7 to enhance apoptosis. The inhibitors of apoptosis (IAP) proteins were found in both extrinsic and intrinsic apoptosis pathways, such as XIAP (X-linked inhibitor of apoptosis), Surviving

and others (Vucic & Fairbrother, 2007). They were found to prevent cellular death in response to various stimuli, sustain tumour cellular survival and block cell death. As a result, eliminating IAP protein function was regarded to be a therapeutic target to improve cancer treatment (Hunter et al, 2007; Vucic & Fairbrother, 2007). The apoptosome is a multisubunit protein complex and is induced by the interaction between Cytochrome c and Apaf-1 (apoptosis protease activating factor-1). The activation of apoptosome plays an important role in the activation of caspase to trigger apoptosis. As a result, the activation of apoptosome was regarded as a cancer treatment target to promote tumour cell death (Ledgerwood & Morison, 2009).

The tumour suppressor protein, p53 is crucial for the apoptosis process because it regulates numerous intracellular pathways. As p53 accumulates in normal cells in response to DNA damage and other cellular stress signals, p53 enhances the expression of numerous pro-apoptotic genes, like BAX, Noxa and PUMA to promote the apoptosis process. However, the loss of p53 function and the inactivation of p53 in tumour cells prevent the replication of DNA lesions, and therefore mutated p53 promotes cell proliferation and survival. The prevention of apoptosis leads to the development of cancer (Brown & Attardi, 2005). With the purpose of facilitating tumour cell death, the recovery of wild type p53 function to enhance cellular apoptosis was regarded as a potential target of cancer treatment (Gudkov, 2005).

Moreover, recent research found that the p53 target gene CDKN1A (p21) functioned as an anti-apoptotic gene in certain conditions. Figure 1.12 shows the different behaviours of p21 in p53 wild type cells (Figure 1.12 A) and solid tumour cells (Figure 1.12B). When p53 wild type cells were affected by ionizing radiation at the early state (less than 24 hours), ATM induced the

activation of p53, WRN (Werner's syndrome protein), DNA-PKcs (DNA dependent protein kinase catalyse) and other genes. Those proteins promoted DSB repair while p21 repressed apoptosis and inhibited the cell cycle transition. There exist positive feedback loops between p21 and ATM, p21 and p53 to enhance cell proliferation. However, if the solid tumour cells was exposed to the ionizing radiation for a long time, p21 repressed apoptosis, and then induced the SIPS (stress-induced premature senescence) programme (Mirzayans et al, 2012).

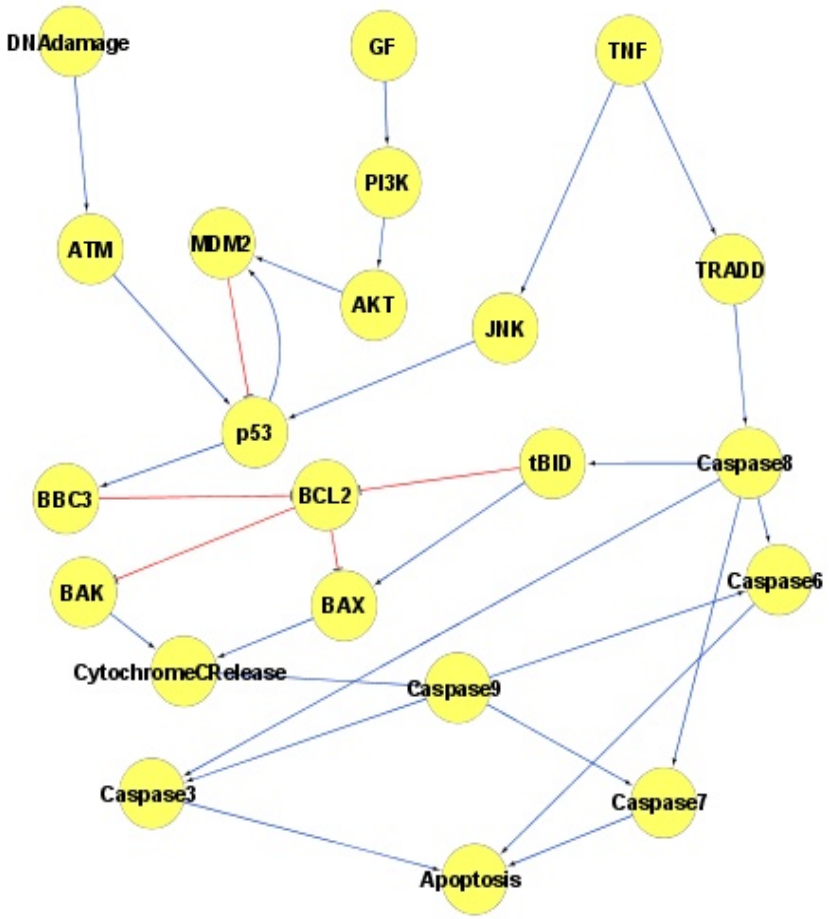


Figure 1.11: Two major signalling pathways that induce apoptosis.

Two crucial pathways start the apoptosis process: the pathway on the left is induced by intracellular signals and the one on the right is stimulated by extrinsic signals (Debridge et al, 2012; Mai & Liu, 2009).

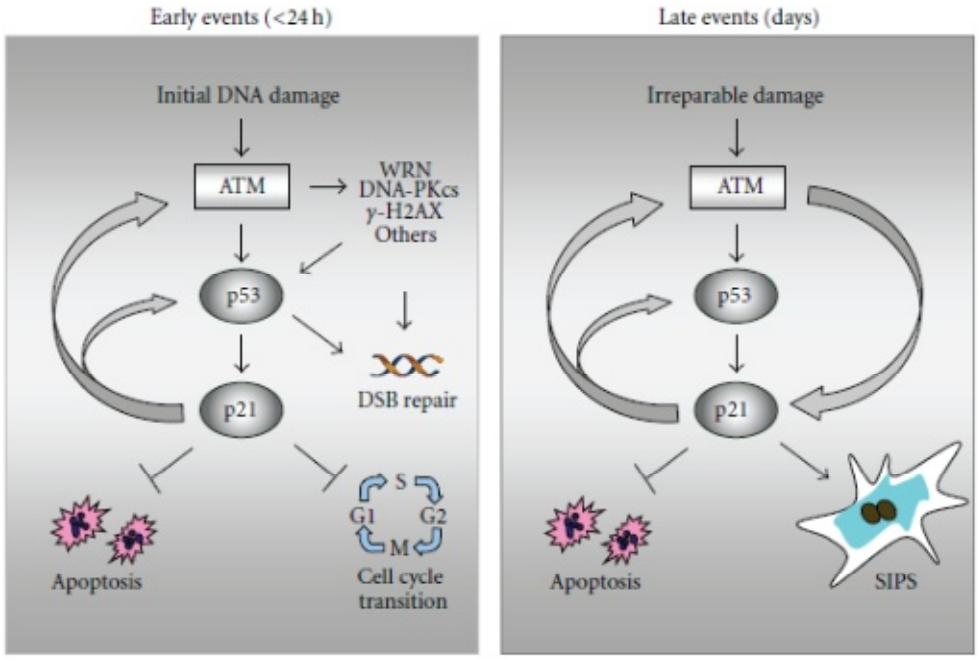


Figure 1.12: ATM-p53-p21 pathways in p53 wild type cells and solid tumour cells in response to DNA damage.

The left picture shows the short stage activity of ATM-p53-p21 pathways in response to DNA damage in p53 wild type cells. The right picture describes a long stage activity of ATM-p53-p21 signalling pathways in solid tumour cells (Mirzayans et al, 2012)

1.3.5.4 Control of angiogenesis

Angiogenesis is a process of the growth of new blood vessels from existing ones which plays an essential role in the growth and metastasis of tumours. As tumours need oxygen and nutrients to survive, which are provided by the network of blood vessels, new blood vessels promote the survival of the tumour. Previous studies of p53 found that the angiogenesis process was regulated by p53 through its control of p53 target genes, for example TSP1 (Thrombospondin-1), which would inhibit cancer angiogenesis and metastasis (Harris et al., 2005). There are three main mechanisms by which p53 regulate

angiogenesis. The first mechanism is to mediate the regulation of target genes which affect angiogenesis in response to hypoxia (Teodoro et al, 2007). For instance, HIF1A is a subunit of the hypoxia inducible factor and is repressed by the p53 protein in the presence of hypoxia(Teodoro et al, 2007). HIF1A mediates the angiogenesis process through the pathways in which the vascular endothelial growth factor takes part (Teodoro et al, 2007). The second mechanism is the repression of those genes that activate the process of angiogenesis (Teodoro et al, 2007). For example, p53 mediates the inhibition of angiogenesis by the regulation of VEGF (vascular endothelial growth factor) and PTGS2 (COX-2). VEGF and PTGS2 are down regulated by p53 and they are reported to enhance the angiogenesis process (Teodoro et al, 2007). The third mechanism is through the induction of angiogenesis repressors. For instance, TSP1 is up regulated by p53 and the migration and proliferation of endothelial cells is inhibited by TSP1 to inhibit angiogenesis (Ren et al, 2006). All those three mechanisms together allow p53 to prevent angiogenesis in normal cells. The loss of p53 function in p53 mutated cells will promote the angiogenesis and lead to the accumulation of tumours (Teodoro et al, 2007) .

1.3.5.5 Control of other cellular process

The tumour suppressor p53 is also involved in the regulation of other biological process, such as senescence, autophagy, metabolism, and inflammation. The tumour suppressor p53 was identified as a major regulator of cell senescence. It induces human cell senescence through CDKN1A (p21) and pRB in response to DNA damage. The p53-p21-pRB pathway could be promoted by ARF (Dimri, 2005). Autophagy is a catabolic mechanism that recycles cell organelles (Lin et al, 2012). It was reported that p53 both

activates and inhibits autophagy through different mechanisms, for instance p53 stimulates autophagy in both transcription dependent and independent manner. Moreover, some p53 target genes such as CDKN1A (p21) and BAX activate autophagy separately, whereas another p53 target gene, C12orf5 (TIGAR) prevents autophagy by the regulation of glycolysis and inhibition of ROS (reactive oxygen species) (Sui et al, 2011).

As well as autophagy, the cell metabolism was found to be regulated by p53. The p53 protein mediates energy metabolism, oxidative stress, and amino acid metabolism through complex mechanisms. Metabolic stress signals stimulate p53 and p53 regulates glycolysis and oxidative phosphorylation (OXPHOS). It was found that p53 mediates autophagy through the mTOR (mammalian target of rapamycin) and DRAM1 (damage-regulated autophagy modulator 1) pathways (Zhang et al, 2010).

The tumour suppressor protein p53 also represses the inflammation process. This inhibition is performed through an interaction with NF-κB using a complex mechanism (Gudkov et al, 2011). It was reported by Komarova et al. that inflammatory mice cells invaded more rapidly in response to ionizing radiation in p53-/- mice than in p53 wild type mice cells (Komarova et al, 2004).

1.4 Aims and objectives

We aimed that by using a systems biology approach to establish logical models of p53 action, our understanding of p53 function could be furthered and the unknown dynamic mechanisms of the p53 signalling pathway induced by DNA damage could be revealed. The overall objective of this project is to analyze DNA damage inducible p53 signalling pathways and the role of the

tumour suppressor p53 in the development of cancer. We aimed to reconstruct *in silico* protein-protein interaction networks of p53 on the basis of massive interaction information retrieved from text mining results. Predictions produced by the model simulations can be validated by experiments in vivo and vitro to estimate the performance of our models and improve them. By the analysis and simulation of this network, key genes and network features in the p53 network were expected to be identified.

In the first instance, our objective was to generate two models: one basic model with high level of certainty (five or more publications describing the interaction are shown in Additional Table 1) and one all inclusive model with low level of certainty (all interactions we found by text mining are included in Additional Table 3). Model construction would be initially based on literature curation. The analysis of those two models would provide a general overview of the p53 network.

Secondly, we aimed to extract massive interaction information from databases using a java programmed application interface and create a logical model in CellNetAnalyzer. Moreover, this interaction network would be connected with input signals and outcomes of biological processes on the basis of the literature survey, so as to explore the effects exerted by this interaction network and predict new findings on the dynamic mechanism of p53 pathways induced by DNA damage.

Finally, we aimed to validate those predictions through various approaches, such as western blotting experimental data (it occupied more than 25% of time spent in the validation work), genome wide experimental data (microarray and ChIP-seq data) (it occupied more than 50% of time spent in the validation

work), and literature survey (it occupied less than 25% of time spent in the validation work). The predictive strength of my model was aimed to be measured by those approaches. The results and predictions of this project are expected to be helpful for cancer therapy design and production of new drugs.

The combination of top-down and bottom-up systems biology approaches is the ultimate goal to establish comprehensive models. In this project, the top-down approach will be utilized first according to the information obtained from the literature to establish an abstract model of p53 pathways. Through the top-down approach, the global properties of the p53 pathway will be explored and predictions for the p53 pathway will be drawn from the model. Then the generic model for p53 will be tested and adjusted by the bottom-up approach. Moreover, by comparing the expected results with the experimental data obtained from lab work, the model will be adjusted according to the actual results. As a result, the combination of these steps constitutes an iterative loop to improve modelling and understanding of p53 pathways.

This thesis contains 7 chapters. Chapter 1 gives basic background knowledge about p53 and system biology. Chapter 2 presents the materials and methods used in this thesis. Chapter 3 describes the generation of p53 interactome models using manual literature survey. Chapter 4 describes early versions of the PKT205 model based on interaction information retrieved automatically from the STRING database. Two large models based on automatic retrieval of interaction information, the PKT1377 and PKT2275 models, are presented. Chapter 5 describes results for the final PKT205 model and its analysis by CellNetAnalyzer. In this chapter, three versions of the PKT205 model and their dependency matrix and logical steady state analysis are illustrated. The third version, the PKT205/G3 model was selected as the final version of our

PKT205 model. Chapter 6 shows various approaches to validate predictions from the PKT205/G3 model and lists the results obtained by comparing our predictions to literature, western blotting experiments, and microarray and ChIP-seq datasets. Chapter 7 discusses all results described in previous chapters and gives directions for model improvement in the future.

Chapter 2 Materials and Methods

2.1 Cell culture and maintenance

2.1.1 Cell lines and media

The p53 wild-type cell line U2OS (p53 +/+) and p53 mutant SAOS2 (p53 -/-) human osteosarcoma cell line were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich) supplemented with 10 % fetal bovine serum (FBS), 1 % penicillin and streptomycin (Lonza) and 2mM L-Glutamine. These two cell lines were used to provide suitable p53 positive and negative system suitable for validation of model prediction. These cells and media were utilized in western blotting and genome wide microarray experiments.

2.1.2 Cell maintenance

75 cm^2 vented tissue culture flasks were used to maintain those two cell lines. Cells were incubated in a 5 % CO_2 incubator at 37°C. They were cultured for 2 to 3 days until 80% confluent. Cell culture experiments were performed in class II safety cabinet to ensure sterile environment. To subculture the cells, media was aspirated, cells were rinsed with 3ml Trypsin-EDTA (Lonza) and then split by incubating in 1 ml Trypson-EDTA Trypsin at 37℃ for 3 min. Cells were detached by tapping and fresh media was added to dilute and stop trypsin action. Cells were passaged usually at 1:5 ratios. Cells were seeded in 100 mm plates until 80% confluent for western blot, microarrays and chromatin immunoprecipitation assays.

2.1.3 Freezing the cells

In order to freeze the cells, they were centrifuged at 12000 rpm (revolutions per minute) for 3 min (Hermle, Z300 centrifuge). Pellets were resuspended in 1ml of FBS and 1ml of 20% FBS in DMSO (dimethyl sulfoxide) to prevent cell death. Cryovails (Nalgene) were used and cell suspension stored at -80℃ overnight and then transferred to the liquid nitrogen containers (-196℃).

2.1.4 Thawing the cells

To obtain fresh cell stock, cells were thawed rapidly by addition of warm media to prevent formation of ice crystals (Ryan, 2004). Cells were transferred to a 25 cm^2 flask and incubated at 37℃ for 3-4 hours. After cell attachment, media was changed. When cells reached 70-80% confluence; they were transferred into the 75 cm^2 flask.

2.2 Antibodies

Following antibodies were used in this study. Antibodies used in the western blotting experiment were: β-Actin (Abcam, UK), Chk1 (DCS-300, sc56290, Santa Cruz Biotechnology, Santa Cruz, CA, USA), Phospho Chk1 (Ser 345, sc17922, Santa Cruz Biotechnology, Santa Cruz, CA, USA), ATM (ATM 11g12, monoclonal antibody, sc53173, Santa Cruz Biotechnology, Santa Cruz, CA, USA), ATR (2790 S, NEW ENGLAND BioLabs), ECL mouse IgG(NA931, Amersham Biosciences (UK)), ECL rabit IgG (NXA931, Amersham Biosciences (UK)) .

2.3 Immunoblotting

2.3.1 Protein extraction

U2OS cells and SAOS2 cells lines were seeded in 100mm culture plates and treated with 10μM etoposide for indicated times (Sigma, UK) that causes double strand DNA damage by inhibiting topoisomerase II (Zhou et al, 1999). Then the media was aspirated from the plates and cells were washed twice with 10 ml 1X cold PBS buffer. 250 μl of High Salt Lysis Buffer with freshly added protease (PI) and phosphatase inhibitors) (Table 2.1) was added to the cells and the cells were scrapped at 4℃ using cell scraper. Then the lysates were placed into eppendorf tubes and rotated for 20 minutes at 4℃. After the rotation, the tubes were centrifuged at 4℃ for 15 minutes at 13000 rpm. The supernatant was transferred into new tubes after the centrifugation and kept on ice.

Buffer	Composition
3X SDS sample buffer	187mM Tris, 30% Glycerol, 6%SDS, 15% 2-mercapto ethanol, 0.01% bromophenoblue
5% milk PBS buffer	5 g milk in 10 ml PBS
High Salt Lysis Buffer	45mM Hepes, 400mM NaCl, 1mMEDTA, 10% glycerol(VWR), 0.5% Igepal and added fresh: 1µg/ml protease inhibitor cocktail (aprotinin,leupeptin and pepstatin A),1M DTT, 1 mM PMSF, 2 mM sodium orthovanadate, 5mM sodium pyrophosphate and 20 mM β glycerophosphate,
PBS	125mM NaCl, 3mM NaH_2PO_4.H20, 7 mMNa_2HPO_4 anhydrous
2.5% milk PBS/Tween buffer	0.25 g milk, 10 ml PBS/Tween-20
Reference Cuevette	H_2O and Biorad
PBS /Tween	PBS+0.1%Tween-20
Stripping Buffer	100 mM 2-Mercaptoethanol, 2% SDS 62.5 mM Tris-HCL pH6.7
10X SDS running buffer	0.25M Tris, 1.9M glycine, 30mM SDS
Western transfer buffer	150mM Glycine, 25mM Tris-HCl pH8.3, 20% methanol.

Table 2.1: List of buffers and solutions for SDS PAGE and western blotting experiment

Column 1 shows the name of buffers or solutions. Column 2 lists the composition.

2.3.2 Protein concentration determination

Protein concentration in cellular extracts was determined using Bio-Rad Protein Assay reagent based on (Bradford, 1976). This essay is based on the change of colour after reagents dye binds to arginine and lysine residues in the protein resulting in the shift of absorbance from 465 to 595nm, which is determined by spectrophotometer.

800µl H_2O was mixed with 200µl Biorad reagent (1:5 dilution) in the reference cuvette and incubated for 5 minutes at room temperature. The absorbance of the samples was measured in spectrophotometer at 595nm optical density. Spectrophotometer was normalized using diluted reagent without the sample. Equal amount of protein was calculated from these values, mixed with 3XSDS sample buffer (Table 2.1) and loaded on SDS PAGE for analysis.

2.3.3 SDS-PAGE

The electrophoresis was performed by SDS-PAGE (SDS-poly acrylamide gel electrophoresis) protocol to separate proteins according to their molecular weight. This method is widely used to separate proteins according to their molecular weight. Polyacrylamide is used as a supporting material and SDS has a purpose of denaturing proteins. The gels were prepared as described in the Table 2.2. The resolving gel was prepared first using Biorad Mini Protean 2 apparatus and immediately overlayed with isopropanol. Once set, isopropanol was removed and the stacking gel was poured and combs inserted to for wells for loading. Gels were either used immediately or stored at 4℃ up to 3 days.

Gels were run using Biorad Mini Protean 2 running apparatus in the tanks filled with 1xSDS running buffer (Table 2.1). The samples were boiled at 95℃ for 5 minutes, centrifuged at 13000 rpm for 1 minute. 5 µl molecular weight markers was used to estimate the size of analysed proteins. The gels were run at 80 volts until samples reached the resolving gel. Then the voltage was increased to 110 volts for 1 hour or until samples were resolved.

Solutions	**7.5% Resolving gel**	**7.5% Stacking gel**
Distilled water	13.3 ml	6.73 ml
Acrylamide	7 ml	1.67 ml
1.5M Tris pH 8.95	7 ml	-
1M Tris pH 6.95	-	1.25 ml
0.2M EDTA	280µl	100µl
10% SDS	280µl	100µl
10% APS	157µl	100µl
TEMED	17µl	10µl

Table 2.2: The composition of polyacrylamide gels

Column 1 lists the name of solutions used. Column 2 shows volume of this solution added for resolving solution. Column 3 lists the volume of this solution added for stacking solution.

2.3.4 Western transfer and detection of proteins

The resolved proteins were transferred from the gel onto a polyvinylidene fluoride (PVDF) membrane (Millipore, UK) using following procedure. Western transfer cassettes (Biorad) were assembled and set up with sponges, blotting paper, gel and membranes in certain order. PVDF membranes were soaked in methanol for 30 seconds and 2 sponges were soaked in the transfer buffer (Table 2.1). Then these cassettes were placed in the tank on a stirrer and the transfer was performed using electric current at 0.4 amps for 90 minutes with ice packs, which was changed every 45 minutes. Once the protein was successfully transferred to the membranes, the membranes were blocked at normal temperature in 5% milk/PBS buffer (0.5 g powdered milk with 10 ml PBS) for 1 hour. Then the membranes were incubated with 2.5% milk/PBS Tween 20 buffers (0.25 g milk with 10 ml PBS/Tween 20) where the first antibodies were added. These membranes with first antibodies in 2.5% milk PBS buffer were incubated at 4℃ overnight.

Then, these membranes were washed by PBS/Tween 20 buffer for 10 minutes and the washing was repeated three times. After washing, they were incubated at room temperature for 1 hour with secondary antibody. The secondary antibodies were relevant anti-rabbit or anti-mouse IgG secondary antibodies in 2.5% milk/PBS/Tween buffer. When the incubation was finished, the membranes were washed 3 times again with PBS/ Tween buffer for 10 minutes. Then 500µl of One Step ECL (Chembio, UK) was added on the membranes. The membranes were incubated at room temperature for 5 minutes and then the proteins visualized by exposure to medical X-ray films (Fujifilm, UK) by the Compact X4, Xograph imaging system. This western

blotting approach was utilized to validate predictions of the PKT205/G3 models and these results are described in the Chapter 6.

2.3.5 Striping the membrane

Once the primary protein was detected, antibody was striped from the membrane, which facilitated the reuse of the membrane to detect the new primary protein. The membranes were incubated with stripping Buffer (Table 2.1) and incubated at 50℃ water bath for 30 minutes with occasional shaking Then the membranes were washed three times by 10 ml PBS/Tween buffer (PBS with 0.1% tween-20) to discard the stripping buffer. After washing, the membranes were blocked for 1 hour with 5% milk PBS (0.5 g milk powder in 10 ml PBS). This blocking was performed on a rocking platform at room temperature. Then the new primary antibody was added into a new 5% milk/PBS buffer and the membranes were incubated in this buffer overnight at 4℃ for new western transfer.

2.4 Microarray analysis

The samples for microarray analysis were prepared from U2OS and SAOS2 cells treated with vehicle or with 10 μM etoposide for 16 hours. Total RNA was extracted from U2OS and SAOS2 cells using RNeasy plus mini columns (Qiagen, UK) according to the manufacturer's recommendations. For each hybridization, 100 ng of total RNA was used in the Affymetrix GeneChip Two-Cycle Target Labeling kit and in the Ambion MEGAscript T7 kit before hybridizing to the GeneChip human genome U133 Plus 2.0 array (Affymetrix) according to manufacturer's instructions (Tian et al, 2013). The microarray

hybridization was performed by the Core Facility staff in the University of Manchester and data provided as excel files. Then analysis was performed by the software Genesis and DAVID (The Database for Annotation, Visualization and Integrated Discovery). The heat map of microarray data was created by use of Genesis software, which allows hierarchal clustering and other functional analysis (Sturn et al, 2002), and cluster analysis was performed according to this heat map. Functional annotation analysis for microarray experimental results was performed using DAVID, meanwhile, the validation of model predictions by microarray experiment result were performed by the Java-based interface as described in the 2.6 paragraph below. This analysis approach was utilized to validate predictions of the PKT205/G3 models and these results are described in the Chapter 6.

2.5 Chromatin immunoprecipitation followed by sequencing analysis (ChIP-Seq)

It has been described that different p53 modified isoforms may have altered interaction with a selected subset of proteins or genes by literature, in order to detect how these p53 isoforms selectively control downstream targets, we turned to the ChIP-seq analysis. The process of ChIP-sequencing includes two main steps: ChIP and Sequencing. First, p53 is cross-linked to DNA. Then chromatin is isolated and DNA sheared. The p53 protein is precipitated with antibody specific for total p53 or for modified p53. This step is followed by reverse cross-linking and digestion of the protein. Here we used ChIP-Seq data obtained for total p53 and published by Smeenk et al.(Smeenk et al, 2011) The analysis of the overlapping data from ChIP-seq results and the model was performed by the Java-based interface generated here. This analysis approach was utilized to validate predictions of the PKT205/G3 models and

results are described in Chapter 6.

2.6 Computational methods

2.6.1 Text mining approaches for literature search of p53 interactions

With the purpose of establishing p53 network model based on literature survey, we used various text mining approaches and tools to perform the literature search for p53 pathway. Here we mainly used three different text mining tools for p53 interaction information retrieval in Chapter 3 and a trusted database in Chapter 4, and 5. For example, the first search engine one was MEDIE, which retrieved information from publications in PubMed. The second tool was to gather literature evidence by plugins provided for Cytoscape (Shannon et al, 2003; Smoot et al, 2011). The third one was KEGG (Kyoto Encyclopedia of Genes and Genomes), which was utilized to check p53 pathway maps (Kanehisa & Goto, 2000; Kanehisa et al, 2012). These three text mining tools were utilized to construct the PKT38 and the PKT62 models and results are described in Chapter 3. The trusted database was STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (Jensen et al, 2009; Szklarczyk et al, 2011). This database was utilized to construct the PKT1377 and the PKT2275 model and results are described in Chapter 4. It was also utilized to construct the PKT205 models and results are described in Chapter 5.

2.6.1.1 MEDIE

MEDIE (Tsujii Laboratory, 2013) is a search engine which retrieves biochemical information from the data base MEDLINE (Medical Literature Analysis and Retrieval System Online)(U.S.National Library of Medicine, 2013), where millions of biomedical journal citations and abstracts of publications are stored. MEDIE (http://www.nacterm.ac.uk/medie/) performs semantic search by identifying semantics of relevant words. We performed semantic search through the MEDIE engine to retrieve relevant information, for example, what was inhibited by MDM2 in Homo sapiens (Figure 2.1). There were three optional input areas provided by MEDIE to type in the key words: subject of a sentence, verb of the sentence and object of the sentence. Once the key word was typed in the target area, the search task was executed by clicking the “search” button. Then a summary result was displayed and the details were listed in the bottom of the screen. The text mining results were represented by three different types: sentences, articles, and tables. The results represented in the format of sentences listed sentences, in which words semantically correlated to the key word were filtered form abstracts in MEDLINE database. Moreover, those correlated words in sentences were highlighted by different colours. In the format of article, each result listed the title and an abstract of publications. The relevant gene names and other relevant verbs were highlighted with different colours. All the sentences and abstracts included in the results of MEDIE are from papers cited in PubMed(U.S.National Library of Medicine, 2013). PubMed (http://www.ncbi.nlm.gov/pmc/) was served by NCBI (National Center for Biotechnology Information) Entrez retrieval system. The Entrez system which is located at the U.S National Institutes of Health (NIH) is a text-based search

and retrieval system. It provides services to PubMed, Nucleotide and Protein Sequences, Protein Structures, Complete Genomes, Taxonomy, OMIM (Online Mendelian Inheritance in Man) and others. The results were also represented by format of tables which included four columns. The first columns listed the title of publications found by search engine by natural language process technologies. The second column listed the entities of the relevant subjects and the gene names indentified were marked by red colours. The third column showed the entities of the relevant verbs in a darker background, and the forth column listed the entities of the relevant objects. This database was accessed to construct the PKT38 model and the PKT62 model, whose results are described in Chapter 3.

The advantage of MEDIE is that the result displayed facilitated analysis of the relevant interaction information and publications in PubMed. The sorted order of the results could be adjusted by setting rank related or published date of journals.

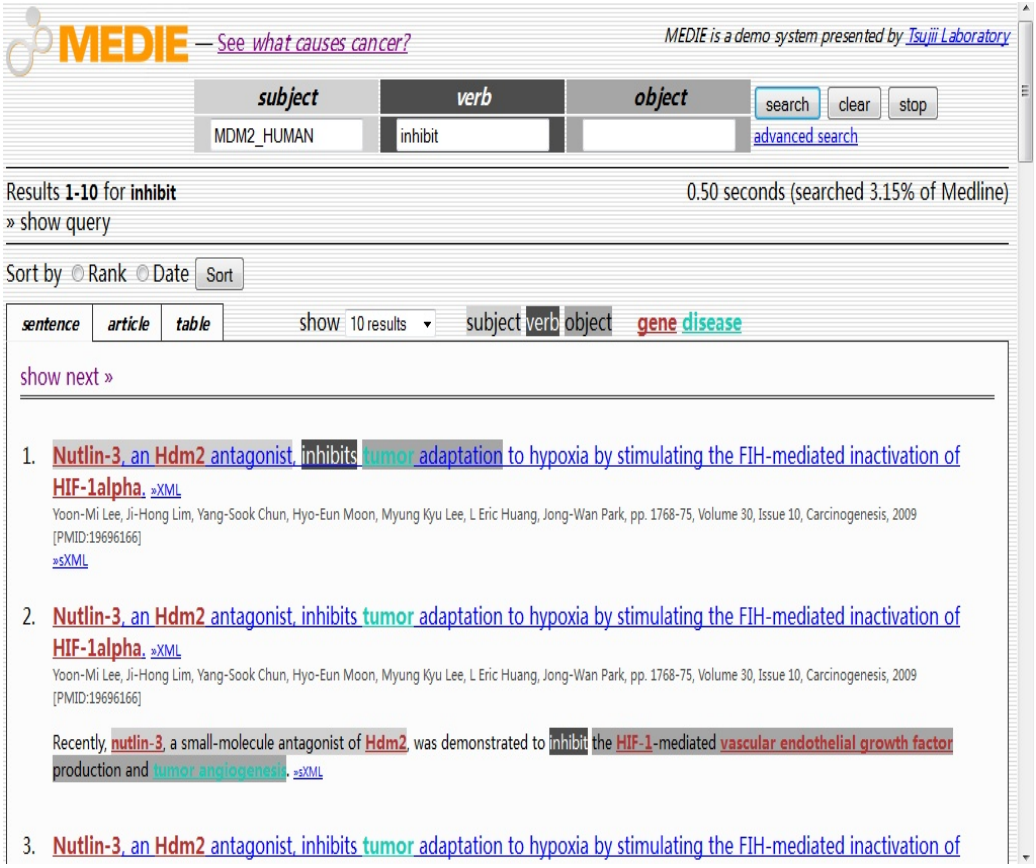


Figure 2.1: An example of MEDIE search result

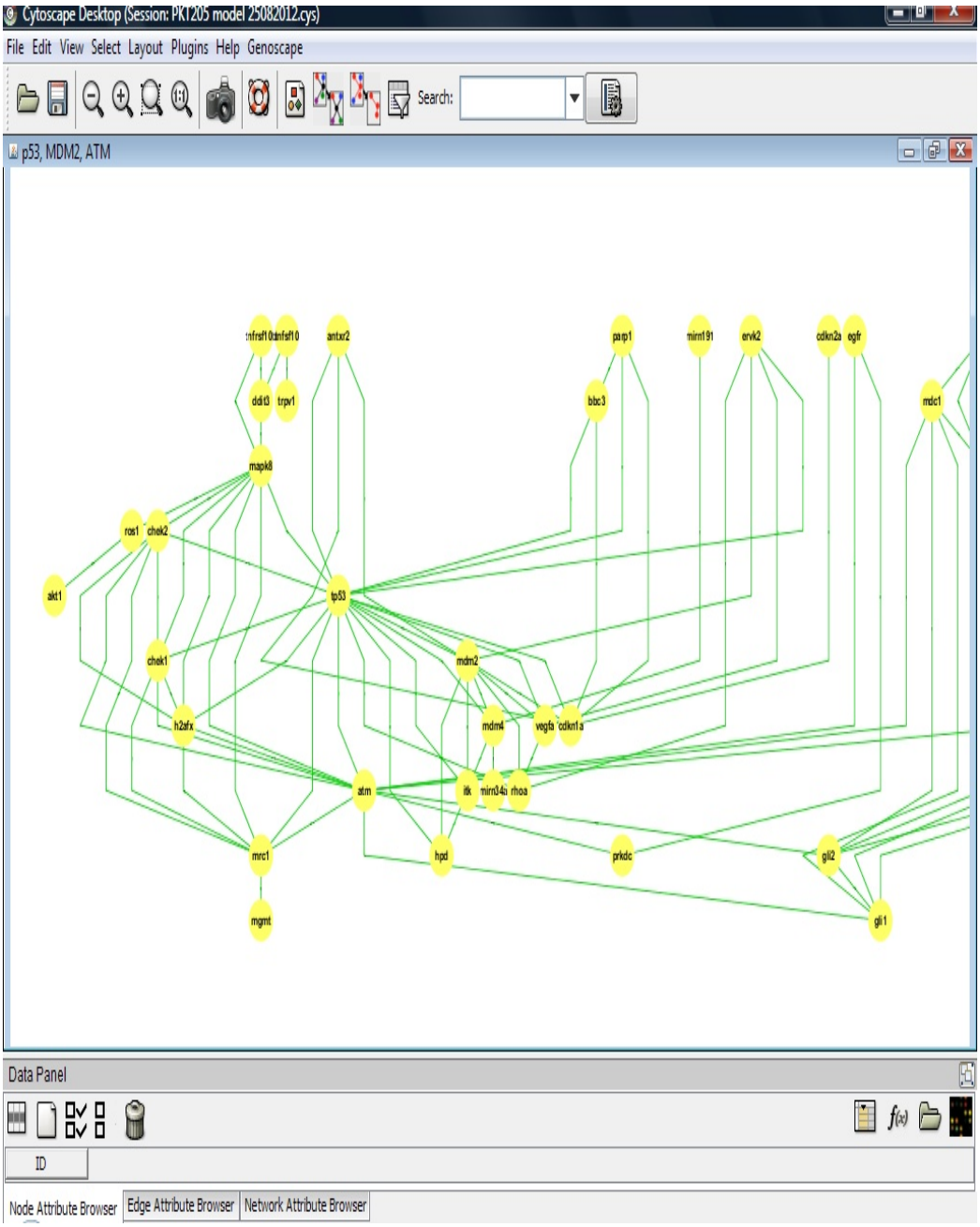
The upper area in the web page has three optional input areas: subject, verb and object and the bottom area lists sentences from abstract retrieved, which contained the key word “MDM2_Human” and “inhibit”. The gene name and the verb word which was synonyms of the key word “inhibit” were highlighted with different colours.

2.6.1.2 Cytoscape and AgilentLiteratureSearch Plugin

Cytoscape (Shannon et al, 2003) is open source software for biologists to provide an overview of visual molecular interaction networks. Complex analysis and visualization can be processed using the Cytoscape platform (Shannon et al, 2003). The AgilentLiteratureSearch is a plugin developed for Cytoscape to establish network model, which includes target genes which we

were interested in, on the basis of literature search. For example, in order to establish the network model that contained p53, MDM2 and ATM for human cells. This plug in could be started by clicking “Agilent Literature Search” in the menu “Plugins” of Cytoscape to run the Agilent Literature Search 2.77. After the terms “p53”, “mdm2” and “ATM” were typed in the area of “Terms” (Figure 2.2A), then this plugin performed a search for these three key words in PubMed and edited the query automatically. As a result, a network model with 41 nodes and 84 edges was established (Figure 2.2A), and literature search results were listed in the bottom area of “Query Matches” in this plug in interface (Figure 2.2B). Through the evidence gathered by the AgilentLiteratureSearch plug in, it is more convenient to check literature evidence about each interaction included in the network. However, the disadvantage of the plug in is that it is not able to provide the most updated literature evidence for the interactions for the p53 network constructed by Cytoscape. In addition, sometimes, there is no evidence for particular interactions. As a result, this plug in is a candidate tool to help cross check the network constructed by Cytoscape in Chapter 3. This plugin was utilized to construct the PKT38 and PKT62 models and Cytoscape was utilized to represent their network maps, whose results are described in Chapter 3. Besides, Cytoscape was also utilized to represent the PKT205/G1, PKT205/G2 and PKT205/G3 models which are described in the Chapter 5.

A


Cytoscape Desktop (Session: PK1205 model 25082012.cys)
File Edit View Select Layout Plugins Help Genoscape
Search:
p53, MDM2, ATM
tnfrsf10dtnfsf10
antxr2
pmp1
mirn191
ervk2
cdkn2a
egfr
ddit3
trpv1
bbc3
mdc1
mapk8
ros1
chek2
akt1
tp53
chek1
mdm2
h2afx
mdm4
vegfa
cdkn1a
atm
ilk
mirn34a
rhoa
mre1
hpd
prkdc
gli2
mgmt
gli1
Data Panel
ID
Node Attribute Browser
Edge Attribute Browser
Network Attribute Browser

B

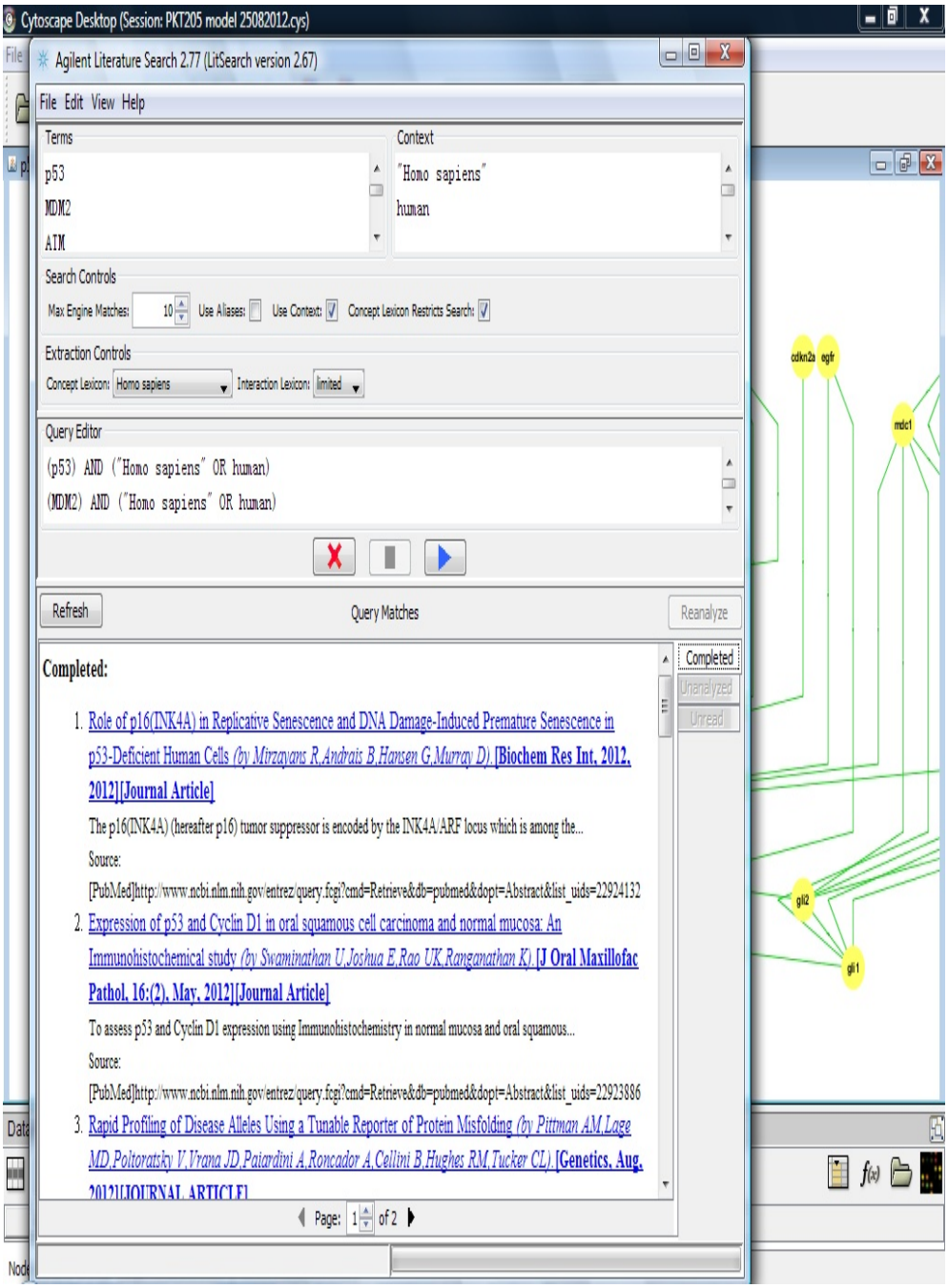


Figure 2.2: An example of AgilentLiteratureSearch plug in for p53 network

Figure 2.2A shows an example of p53 network constructed by AgilentLiteratureSearch plug in. Figure 2.2B shows the interface of AgilentLiteratureSearch and a list of sentences which contains the associations between p53, MDM2 and ATM.

2.6.1.3 Literature searches from other databases

The third text mining tool used to analyse p53 pathways was KEGG (Kanehisa & Goto, 2000; Kanehisa et al, 2012). KEGG is a bioinformatics resource for the biological research and we performed literature search through a unit of KEGG, KEGG PATHWAY(Kanehisa & Goto, 2000; Kanehisa et al, 2012), which provided pathway maps and pathway modules. Several network maps relevant to p53 were found by KEGG PATHWAY. The type of results were represented by pathway maps and the results contained the key words, “p53” and ”human”. This software was utilized for literature work to construct the PKT38 model and the PKT62 model, whose results were described in Chapter 3. For example, one pathway map showed on the summary result page: hsa04115, which was a map of p53 signalling pathway for Homo sapiens (Figure 2.3). 23 pathways about p53 were identified using KEGG, and the interactions were mainly from map04115 (p53 signalling pathway) and map04110 (cell cycle).

The advantage of KEGG was that these results about p53 pathways from databases provided a more global view to investigate p53 networks and contributed to improvement of the construction of the p53 network. However, the drawback of KEGG was that some literature evidence to support these data was out of date and not precise for interaction nature. Moreover, KEGG was not updated frequently.

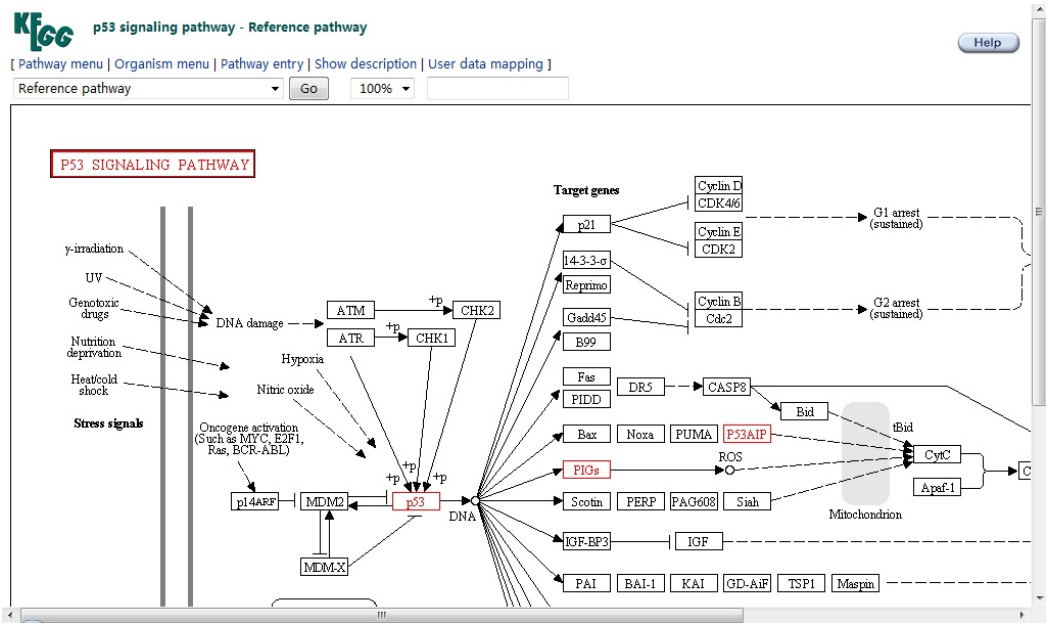


Figure 2.3: A screenshot of KEGG about p53 network maps in human cells

Figure 2.3 shows a p53 signalling pathway map for human provided by KEGG.

2.6.1.4 STRING database

The trusted database we utilized in Chapter 4 and 5 was the STRING database (Szklarczyk et al, 2011). STRING is a protein-protein interaction database which encompasses protein interactions from four sources: genomic context, high throughput experiments, conserved co-expression and previous knowledge by natural language processing (Szklarczyk et al, 2011). The protein-protein interaction results were web accessible by being visualized as network map with the requirement of users. For example, the results displayed should not exceed 10 interactors. The interface of STRING was user friendly and allowed adjustment of the type of interactions of interest. For example, we could select the interaction map for which certain a confidence score is assigned by STRING and also the nature of the interaction (activation or inhibition) is provided by STRING and included in the network map. STRING

established a confidence score schema to evaluate the confidence level of protein-protein interactions using four sources mentioned above (genomic context, high throughput experiments, conserved co-expression and previous knowledge by natural language processing). Data was updated frequently to replace out of date interactions with the new ones. Although, the drawback of STRING was the limitation of results displayed in the web page graphically (the network should include no more than 50 interactors), STRING provided protein action file which contained all protein-protein interaction records, to download. As a result, STRING was selected as the main source of data for model construction. The confidence score is a value between 0 and 1; a confidence score of more than 0.7 (the probability that the interaction was true was more than 70%), is regarded as a high confidence level. Using these criteria, we extracted all high confidence (human) protein interactions using a custom designed Java interface (described below). All interaction records were subsequently manually curated by surveying associated literature references and searching for additional evidence wherever necessary. For example, there were 677 interactions in the PKT205/G3 model (details were described in Chapter 5): (1) all direct interactions with p53 (225 interactions), (2) all interactions between genes/proteins that interact with p53. These interaction records were listed in a text file, which was further processed into a node transcript and a reaction transcript readable by CellNetAnalyzer. The node transcript includes gene names and the reaction transcript includes interaction types (activation or inhibition) and the names of the two genes participating in the interaction. This database was utilized to extract interaction information automatically to construct series of models, such as the PKT1377model and the PKT2275 model, whose results were described in Chapter 4, the PKT205/G1 model, the PKT205/G2 model and the PKT205/G3 model which were described in the Chapter 5.

2.6.2 Java programme interfaces to automatically import STRING records into the CellNetAnalyzer

During the course of this study, several Java based interfaces were constructed serving different purposes: the first one was used to extract interaction information from data in STRING database and convert them into input files for CellNetAnalyzer to calculate the dependency matrix; the second one was generated to compare two different dependency matrixes for *in silico* knock-out tests; the third interface was created to compare the simulation prediction with the microarray experimental data.

The protein action file from the STRING database was downloaded and then all interaction records were filtered by a Java based programme interface (as described below) to retrieve interactions involving p53 (Figure 2.4) (Tian et al, 2013). Since all interaction records were represented in a fixed form (Table 2.3), we designed a Java based programme interface to retrieve these interactions (Additional File 2). The mechanism of interaction retrieval was illustrated in previous section. Once the interaction data were downloaded from STRING database, the interaction records for Homo sapiens interactants were filtered by selecting the appropriate organism identifier using a Java programme interface.

item_id_a	item_id_b	mode	action	a_is_acting	score	sources	transferred_sources
9606.ENSP00000372415	9606.ENSP00000269305	expression	inhibition	1	800	NLP	NLP
9606.ENSP00000200453	9606.ENSP00000269305	ptmod		1	234	NLP	NLP

Table 2.3: An example of the protein-protein interaction record downloaded from the STRING database

Column 1 shows the id of species A in STRING. Column 2 shows the id of species B in STRING. Column 3 defines the type of the mode, the values of mode included binding, expression, activation, reaction. Column 4 defines the type of actions; the values included activation and inhibition. Column 5 defines the direction of the interaction. If the value is “1”, it means the interaction is from species A to species B. Column 6 defines the confidence score. Column 7 defines the text mining source where the STRING extracted the information. Column 8 defines the other resource transferred.

The filtered interactions were then further classified for the interaction type. All interaction records whose mode type was “ptmod” (posttranslational modification) or whose action type was “activation”, or “inhibition” were selected. The remaining interactions were imported into the interaction direction identifier unit to remove duplicate records. Since an activation or inhibition involves two interaction partners and can be represented in either direction, these duplicate records were eliminated by checking the interaction direction. These original interaction records were further filtered using the confidence scheme defined in STRING. They were classified into different confidence score categories through the score classifier unit. Those interaction results enabled us to create p53 networks with different confidence levels. Using the confidence score scheme defined by the STRING database, we created two different types of models: the first model included all interaction records with a confidence score higher than 0.700 (such as the PKT1377 model, the PKT205/G1 model, the PKT205/G2 model, and the PKT205/G3 model) and the second model included all interactions with a confidence score between 0 and 1(for example, the PKT2275 model). This confidence score was defined by the developer of STRING database and

showed a probability of the interaction prediction found in trusted evidence from literature or experiment data (Jensen et al, 2009; Szklarczyk et al, 2011).

This interface application was utilized to extract interaction information automatically to construct series of models, such as the PKT1377model and the PKT2275 model, whose results were described in Chapter 4, the PKT205/G1 model, the PKT205/G2 model and the PKT205/G3 model which were described in the Chapter 5.

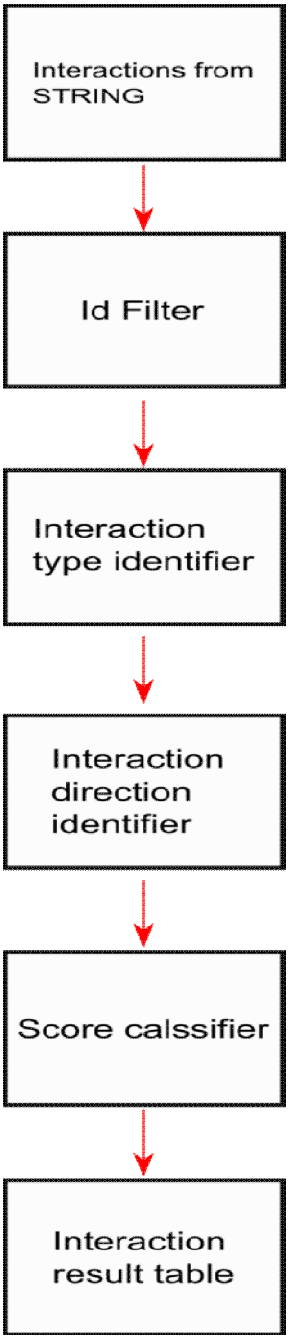


Figure 2.4: Process diagram of interaction information retrieval

It shows the process to extract interaction information automatically from STRING database by the interface. Interactions were filtered by certain criteria and finally listed in a text table which could be imported into CellNetAnalyzer.

2.6.3 Functional analysis by CellNetAnalyzer

CellNetAnalyzer is a powerful analysis tool for signal flow models, which accepts two types of interactions: activation and inhibition (Klamt et al, 2007). We used two techniques provided by CellNetAnalyzer (v. 9.8) to analyze our model. The first technique is the calculation of the dependency matrix, which represents the effects between pairs of nodes in the model. CellNetAnalyzer calculates positive and negative paths between two nodes *i* and *j* and identifies six types of effects in the dependency matrix: no effect, ambivalent factor, weak inhibitor, weak activator, strong inhibitor, strong activator defined below:

If there is neither a positive nor negative path from node *i* to node *j*, node *i* has no effect on node *j*;

If there is both a positive and negative path from node *i* to node *j*, node *i* is an ambivalent factor of node *j*;

If there are only negative paths from node *i* to node *j* and negative feedback loops are present in these negative paths, node *i* is a weak inhibitor of node *j*;

If there are only positive paths from node *i* to node *j* and negative feedback loops are present in these positive paths, node *i* is a weak activator of node *j*;

If there are only negative paths from node *i* to node *j* and negative feedback loops are absent in these negative paths, node *i* is a strong inhibitor of node *j*;

If there are only positive paths from node *i* to node *j* and negative feedback loops are absent in these positive paths, node *i* is a strong activator of node *j*.

The second approach used was logical steady state analysis. In logical steady state analysis, each scenario is set where input signals are initiated by different values (“0” means inactivated, “1” means activated and “NaN” means undetermined). Then CellNetAnalyzer calculates the state of each node and each interaction in the network model by logical operation and produces a list of node states and interaction states in the whole model (Klamt et al, 2006).

2.6.4 Java programme interfaces to automatically compare dependency matrixes obtained by knock-out tests

After the transcript files for nodes and reactions were imported into the empty signalling flow network project constructed by CellNetAnalyzer, the p53 Boolean network model was established. Then we performed *in silico* knock-out tests by deleting individual nodes (such as the p53 node) and new dependency matrix was compared with the one in the p53 wild type model(Tian et al, 2013). Those two dependency matrixes were compared by the second Java based interface (Additional File 2). This interface compared the remaining effect cells in the dependency matrix of the knock-out test with the default dependency matrix in the p53 wild type model. The changed effect cells were classified into different categories: for example, category 1 is for all of effect cells in the dependency matrix of p53 wild type model changed to no effect elements in the dependency matrix of the knock-out test model. In this way, effect cell changes were classified into 6 different categories (no effect, ambivalent factor, weak inhibitor, weak activator, strong inhibitor, and strong activator). This interface was utilized to estimate predictions produced by knock-out tests for the following models: the PKT38 model in the Chapter 3, the PKT205/G1 model, the PKT205/G2 model and the PKT205/G3 model mentioned in Chapter 5 and validate predictions of the PKT205/G3 model, which were described in the Chapter 6.

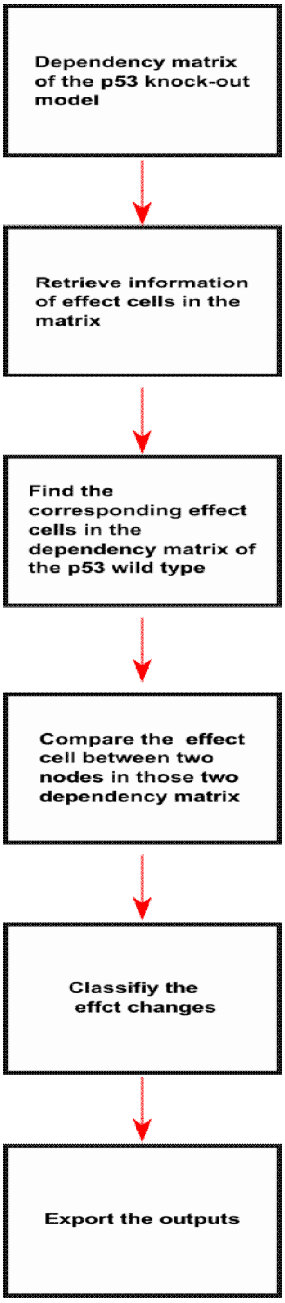


Figure 2.5: Process diagram of dependency matrix comparison

Figure2.5 shows an example of the process to compare dependency matrixes in the *in silico* p53 knock-out test. The dependency matrix of the knock-out tests was analyzed by the approach that each effect cells between two genes in the matrix were compared with the one in the dependency matrix of p53 wild type. Those two elements should correspond to the same gene pairs (the row for a gene and the column for a gene should be same).

2.6.5 Microarray processing and analysis

The samples for microarray analysis were extracted from U2OS and SAOS2 cells and treated with 10 μM etoposide for 16 hours. The microarray experimental results were obtained from the Core Facility and exported to excel files. Then analysis was performed by the software Genesis (Sturn et al, 2002) and the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al, 2009). The heat map of microarray data was created by use of Genesis software, which allows hierarchal clustering and other functional analysis (Sturn et al, 2002), and cluster analysis was obtained according to this heat map. Functional annotation analysis for microarray experimental results was performed using DAVID (Huang et al, 2009), meanwhile, the validation of model predictions by microarray experiment result were performed by the Java-based interface.

DAVID is a web accessible database for functional annotation analysis and facilitates understanding of biological meaning behind large gene lists (Huang et al, 2009). Functional annotation analysis for the gene lists obtained from genome microarray experimental data was performed to explore biological meaning behind these gene lists. Since DAVID is able to indentify enriched Gene Ontology terms, find enriched functional-related gene groups and cluster redundant annotation terms, it was used for functional annotation analysis for the gene lists from microarray data. The functional annotation tools provided gene-annotation enrichment analysis and functional annotation clustering analysis on the uploaded gene list. The detailed functional enrichment analysis of results relevant to Gene Ontology (GO) terms was investigated by selecting the annotation categories relevant to GO terms. For example, the GO terms relevant to apoptosis. We sorted all selected annotation report results by the Benjamini-Hochberg False Discovery Rate

(FDR) from smallest to largest. Then we exported and saved the annotation result by predefined criteria, for example, all annotation charts with GO terms should have the Benjamini-Hochberg FDR less than 1.0×10^{-4} (1.0E-4). The population of genes found relevant to the GO terms, the percentage of them out of the total genes and their modified Fisher Exact P-Value were also exported and saved. DAVID used Fisher Exact P Value to measure significance of association with the selected GO terms. This software was utilized to estimate predictions produced by the PKT205/G3 model which were mentioned in Chapter 5 and validate predictions of the PKT205/G3 model. Its validation results were described in the Chapter 6.

2.6.6 Model evaluation from microarray data

Here we compared our predictions obtained by *in silico* deletion of p53 to *in vitro* generated microarray data from p53 positive (for example, U2OS and HCT116 p53+/+) and p53 negative (for example, SAOS2 and HCT116 p53-/-) cell lines treated by the DNA damaging compound etoposide. Logical steady state analysis produces a steady state in each scenario, and changes of gene states can be compared between model predictions and experimental data (Christensen et al, 2009). For a node i, the predicted state of i in the p53 wild-type was defined as $S(i)_{wt}$, which could take values of 0, 1 or NaN. In the p53 mutant, the state of node i was defined as $S(i)_{mu}$, which could take the same values (Tian et al, 2013). The value of E_{mod} was defined to represent the predicted change of gene state from p53 wild-type to mutant in all 9 types of possible situations as indicated below:

$E_{\mathrm{mod}} = 0$, if $S(i)_{wt} = 1$ and $S(i)_{mu} = 1$;

$E_{\mathrm{mod}} = 0$, if $S(i)_{wt} = 0$ and $S(i)_{mu} = 0$;

$E_{\mathrm{mod}} = 0$, if $S(i)_{wt} = NaN$ and $S(i)_{mu} = NaN$;

$E_{\mathrm{mod}} = 1$, if $S(i)_{wt} = 0$ and $S(i)_{mu} = 1$;

$E_{\mathrm{mod}} = 1$, if $S(i)_{wt} = 0$ and $S(i)_{mu} = NaN$;

$E_{\mathrm{mod}} = 1$, if $S(i)_{wt} = NaN$ and $S(i)_{mu} = 1$;

$E_{\mathrm{mod}} = -1$, if $S(i)_{wt} = 1$ and $S(i)_{mu} = 0$;

$E_{\mathrm{mod}} = -1$, if $S(i)_{wt} = 1$ and $S(i)_{mu} = NaN$;

$E_{\mathrm{mod}} = -1$, if $S(i)_{wt} = NaN$ and $S(i)_{mu} = 0$.

Another parameter E_{exp} was defined to represent the change trend of expression levels from experimental validation(Tian et al, 2013). Those validations were from experimental results of literature survey or microarray analysis result. For experimental results from literature survey:

If the expression level of gene i *w*as considered as up-regulated, $E_{\mathrm{exp}} = 1$;

If the expression level of gene i *was* considered as down-regulated, $E_{\mathrm{exp}} = -1$;

If the expression level of gene i was considered as unchanged, $E_{\mathrm{exp}} = 0$.

For the microarray data, the gene fold change $FC(i)$ was determined by the following equation:

$$FC(i) = \frac{M1(i)}{M2(i)}$$

Where $M1(i)$ is the median of expression values in the target scenario and $M2(i)$ is the median of expression values in the source scenario.

In order to normalize the distributions of expression profiles for different types of cells, the $\log_{10}$ value of all fold changes $FC(i)$ were calculated and two thresholds ($q_{\max}$ and $q_{\min}$) were chosen to determine whether each gene was considered up-regulated, down-regulated or unchanged (Schwartz et al, 2007). The thresholds were determined using the mean value ($\bar{X}$) and the standard deviation (s) of the distribution of $\log_{10}(\mathrm{FC(i)})$ as follows:

$q_{\max} = \bar{X} + s$;

$q_{\min} = \bar{X} - s$

Next, we determined whether the gene was considered up-regulated, down-regulated or unchanged as follows:

If $\log_{10}(\mathrm{FC(i)}) > q_{\max}$, gene i was considered as up-regulated, $E_{\exp} = 1$;

If $\log_{10}(\mathrm{FC(i)}) < q_{\min}$, gene i was considered as down-regulated, $E_{\exp} = -1$;

If $q_{\min} < \log_{10}(\mathrm{FC(i)}) < q_{\max}$, gene i was considered as unchanged, $E_{\exp} = 0$.

The difference between E_{mod} and $E_{\exp}$ was evaluated by the expression $\left|E_{\mathrm{mod}} - E_{\exp}\right|$ (Tian et al, 2013). This difference can take three possible values: 0, 1 or 2. Here, a value of 0 meant that the simulation prediction matched the experimental result; 1 meant that there was a small error between the simulation prediction and the experimental result; 2 meant that there was a large error between the simulation prediction and the experimental result, the

model predicting an opposite direction of change than experimental results. This approach was utilized to validate logical steady state analysis predictions produced by the PKT205/G3 model which were mentioned in Chapter 5. The validation results were described in the Chapter 6.

2.6.7 Java programme interface to automatically validate the model prediction using microarray data as a source

The third interface was established with the purpose of model validations (Additional File 2). By this interface, the logical steady state analysis simulation result was translated into different lists with gene name and their states. For example, the state of ATM in the p53 wild type cells was activated, which was represented as ON, when the input signal, DNA damage was predefined as “ON” (activated). Then two lists for different conditions were compared and the change trend was classified into three values: “1” (which represented up-regulation),”0” (which represented no change), and “-1” (which represented down-regulation). For example, the comparison of logical steady state simulation from the p53 wild type model with DNA damage ON with the logical steady state simulation of p53 minus model with DNA damage ON, the state of ATM remained “ON” and the change trend value was “0”(unchanged). Then the Java based interface compared the change trend of model simulation with the change trend value of microarray experimental data, and performed the statistical calculation of the percentage distribution of correct prediction, small errors and large errors. This interface was utilized to validate predictions produced by logical steady state analysis for the PKT205/G3 model, which was mentioned in Chapter 5 and validation results were described in the Chapter 6.

2.6.8 High performance computation

The structure of the model was so complex that the simulation computation time consumed increased exponentially with the increase of the model size. During the computation process of the PKT1377 model, it was observed that the local desk-top computer could not afford the computation task of the dependency matrix and the computer indicated the state of busy and stopped responding. As a result, the high performance computation tools were used to solve this problem (Smolinski, 2010).

After the local computer was already connected to the server agent, CellNetAnalyzer API (application programming interface) commands were run in local computer to calculate the dependency matrix. The dependency matrix result was exported into a text file which was downloaded to local computer by the SSH Tectia software, and the time consumed by different algorithms was recorded for further comparison. Results of high performance computation were described in Chapter 4.

Chapter 3 Generation of p53 interactome using manual literature survey

3.1 Introduction

Since mutations of p53 are found in more than 50% of malignant tumours (Tang et al, 2007), there is an increasing interest to study p53 function and the dynamic mechanisms of the way it controls pathways to cancers and response to chemotherapy. The regulatory network of p53 is regarded as a transcriptional regulatory system and in order to increase our understanding of p53 in the development of cancer, systems biology approaches were adopted. We aimed to get an insight into p53 pathways involved in cancer development and treatment using a Boolean network models. In those Boolean networks, the signalling pathway was simplified and nodes in the network were representing genes or proteins which interact with p53. Those nodes were connected by Boolean functions to represent the nature of their relationships. This approach is expected to further our understanding of the mechanisms for the p53 pathways and make predictions for their functions. In this chapter, two p53 models constructed by manual literature survey and the results of their functional analysis are described. The first section is about the construction of the PKT38 model (PKT is short for p53 network model constructed by Kun Tian and the number indicates the number of protein or gene nodes included in the model) and its simulation results. The second section is about the PKT62 model and its simulation results. Only interaction graphs were utilized to represent the p53 regulatory network in my project because of the lack of quantitative experimental evidence about relationships involving more than two species in the p53 network.

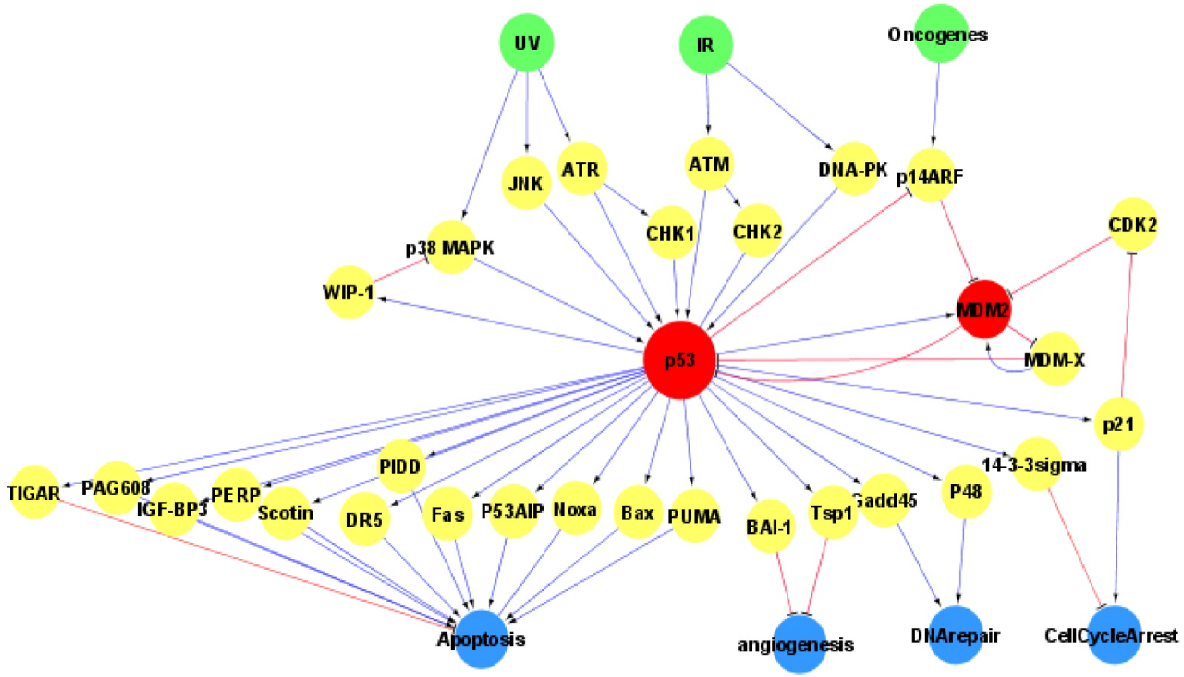
As described in the Method chapter, different approaches were utilized to construct these two initial models of p53. Using text mining tools mentioned in previous chapter, the p53 interactome information was retrieved manually from more than 500 papers published in PubMed (Additional Table 1 and Additional Table 3), where there are more than 21 million biomedical literature citations included in PubMed. All those literature citations were from the database MEDLINE and the population of citations were constantly increasing. The PKT38 model was constructed to demonstrate feasibility of model generation and provide preliminary data of its usefulness and predictive power. The model was generated based on interactions of p53 which are documented in five or more than five published papers in PubMed where the number five was set arbitrarily (Additional Table 1). On the other side, the PKT62 model includes p53 interactions that have been demonstrated in at least one paper published in PubMed (Additional Table 3). This criterion to define an established interaction was set as an arbitrary value in order to construct a simple initial model and demonstrate feasibility of modelling techniques applied to p53 network, given the fact that analysis of all 64,000 publications (accessed on Dec 6th, 2012) was not possible in a manual way. The PKT38 model also included interactions found from other database resources, such as MEDIE search engine and KEGG pathway introduced in the method chapter.

3.2 The PKT38 model

In this section, the construction process of the PKT38 model was described and the structure of this model was illustrated. The PKT38 model was established by CellNetAnalyzer for functional analysis and the network map was represented by Cytoscape. There are 38 nodes and 62 edges included in the PKT38 model (Figure.3.1; Additional Table 1). All interactions were represented by the interaction matrix (Figure 3.2A).The interaction matrix

represented the role of each node in the interaction. One column displayed an interaction in the network model and one row showed one node in the network. If the node did not participate in the interaction, it was marked by black colour. Otherwise, they were marked by three different colours to represent their role. The first example in the Figure 3.2A represented the inhibition of apoptosis by TIGAR. The source node of this inhibition, TIGAR was marked by red colour, and the target node of the inhibition, apoptosis was marked by blue colour. The second example of the interaction matrix in the Figure 3.2A represents the activation of p21 (CDKN1A) by p53. The source node of this activation, p53 was represented by green colour, and the target node, p21 was marked by blue colour. These four elements were marked by red circles (Figure 3.2A)

According to the function of nodes in the PKT38 model, they were divided into five layers: the input signals, the upstream of p53, p53 and MDM2, the downstream of p53 and the outputs. There are three input nodes linked to genes or proteins in this model to represent the stress signals onto the model: UV, IR and oncogenes and four output nodes: Apoptosis, Angiogenesis, DNA repair and Cell Cycle Arrest lined as the cellular response to stress.



Input

Other genes interact with p53 or MDM2

MDM2 and p53

Output

Figure 3.1: The PKT38 model network map

The input nodes of PKT38 were represented by green colours, the medium nodes of the model were represented by yellow colour, and the output nodes were marked as blue colour. Because of the central role of MDM2 and p53 in numerous feedback loops, it was marked by red colour. PKT38 model has two different interactions: activations which are represented by blue arrows, and inhibitions, which are represented by red arrows.

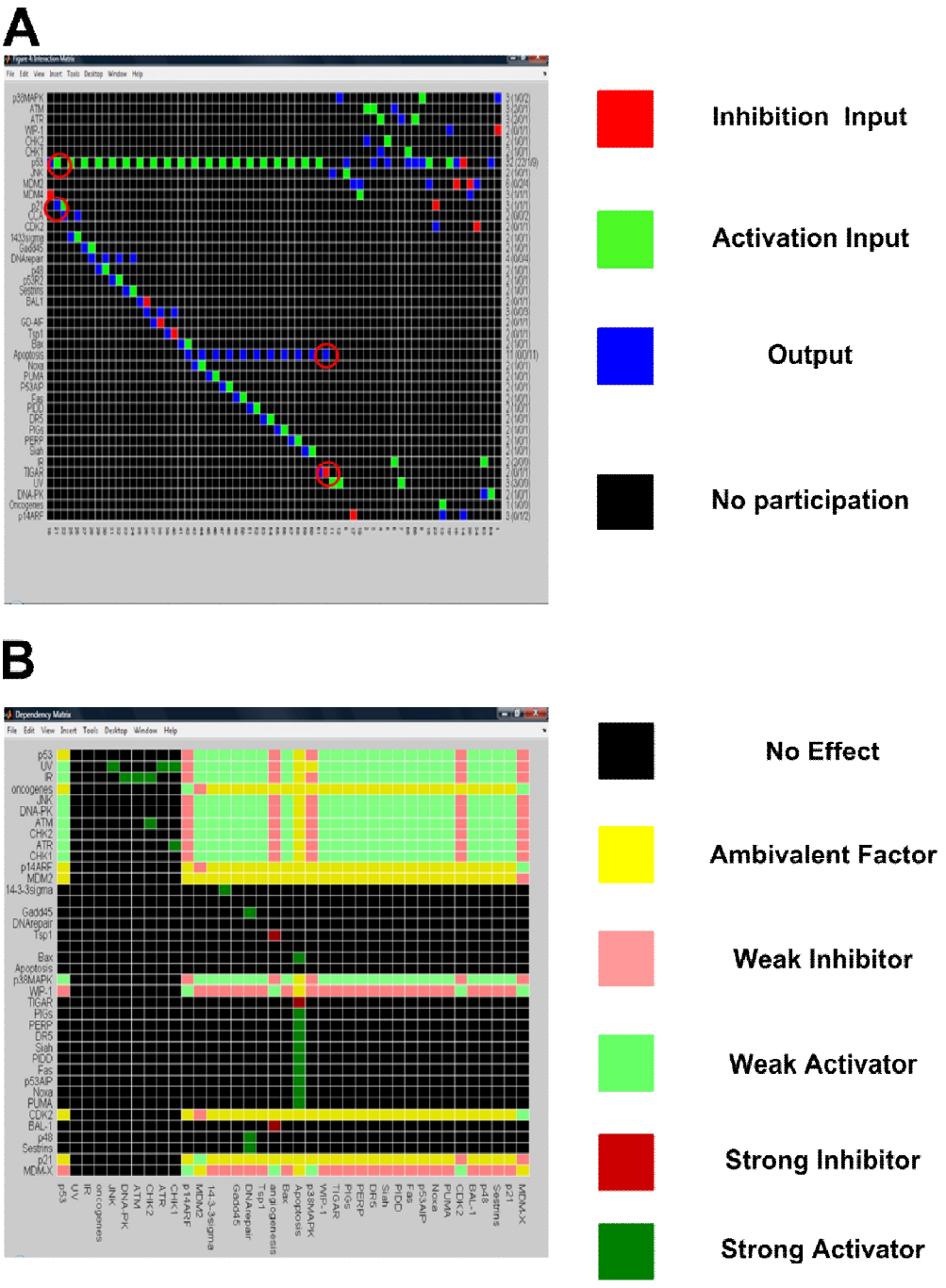


Figure 3.2: Analysis result of the PKT38 model by CellNetAnalyzer

In Figure 3.2 A, each row in the interaction matrix represents a node in the p53 network and each column represents an interaction edge in the p53 network and marked by a number. The green colour elements in the interaction matrix represent the activation input, while the red colour elements represent an inhibition input for the interaction displayed in the column; the blue colour elements in the interaction matrix represent the output of the interaction, and black colour indicates that the node does not participate in this interaction. In Figure 3.2 B, each row and each column represent a node in the p53 network. Each row of the dependency matrix shows how the corresponding species influences the other nodes and each column shows how the corresponding species is influenced by the others. The colour of the dependency matrix element e_{AB} indicates one of the 6 possible types of dependency: 1) the black means A has no influence on B; 2) yellow means A has activating and inhibiting effect on B; 3) pink means A is a weak inhibitor of B; 4) turquoise green means that A is a weak activator of B; 5) red means A is a strong inhibitor of B; 6) dark green means A is a strong activator of B.

3.3 Analysis of dependencies in the PKT38 model

As described above, the PKT38 model was divided into five layers according to the function of the nodes. In this section, we performed the analysis of dependencies in the PKT38 model. We investigated the dependencies relationship between nodes in the PKT38 model at first. Then two types of *in silico* knock-out tests were performed. The dependency relationship between nodes was explored by the dependency matrix of the model, which was calculated by CellNetAnalyzer. Certain nodes or interactions were selected to be depleted from PKT38 model so as to obtain an insight into their role in the whole p53 network.

According to the *in silico* simulations, dependency matrix for the PKT38 model

in Figure 3.2B was calculated, the relationship between input nodes and output nodes was obtained and described below. Six different types of effect elements in the dependency matrix were observed, for example, IR was strong activator of ATM. The effect cell from IR onto ATR was marked by dark green colour in the row of IR and the column of ATM.

In order to investigate the dependencies between the input signals from environment and the outputs from the network onto the environment, the effect cells in the dependency matrix of the PKT38 model were investigated. The global dependency relationship between the input nodes and the output nodes in the PKT38 model were explained and classified into four groups depending on the type of target nodes:

The first group is that oncogenes is an ambivalent factor for Cell Cycle Arrest, UV is a weak activator of Cell Cycle Arrest, and IR is a weak activator of Cell Cycle Arrest;

The second group is that oncogenes is an ambivalent factor for DNA repair, UV is a weak activator of DNA repair, and IR is a weak activator of DNA repair;

The third group is that oncogene is an ambivalent factor for Apoptosis, UV is an ambivalent factor for Apoptosis, and IR is an ambivalent factor for Apoptosis;

The fourth group is that oncogene is an ambivalent factor for Angiogenesis, UV is a weak inhibitor of Angiogenesis, and IR is a weak inhibitor of Angiogenesis.

Those four groups of dependency relationships were obtained from the dependency matrix of the PKT38 model. It indicated that different stimuli exert different effects on the outcome of biological processes in the p53 network. With the purpose of exploring the internal relationship between genes or proteins in the PKT38 model, we performed *in silico* knock-out tests of selected genes or interactions. We performed two types of *in silico* knock-out tests. The first one was to remove selected node from the PKT38 model, calculate the new dependency matrix and compare it with the default dependency matrix shown in Figure 3.2B. The second one was to remove selected interactions and compare the new dependency matrix with the default one for the p53 wild type. The comparison results were described in the section 3.3.2.

3.3.1 *in silico* knock-out tests for selected node depletion

In the previous section, the internal relationship between input nodes and output nodes in the PKT38 model were explored. Here the focus is on role of other nodes in the PKT38 model. Since there were three layers in the medium nodes of the PKT38 model: the upstream of p53, p53 and MDM2 and the downstream of p53, certain nodes were selected from those three different layers to investigate their role in the whole model. As shown below (Table 3.2), selected nodes were removed from the PKT38 model and new dependency matrix was calculated for these new modified models. The result of the six different effect element populations was listed in Table 3.2.

Node removed from the PKT38 model	Total dependency effect elements	No Effect	Ambivalent Factor	Weak Inhibitor	Weak Activator	Strong Inhibitor	Strong Activator
Null	1444	926	150	101	242	3	22
P53	1369	1327	0	7	5	5	25
MDM2	1369	1017	13	85	226	4	24
ATM	1369	883	149	96	219	3	19
BAX	1369	869	145	99	232	3	21
P38MAPK	1369	924	142	64	214	3	22
TIGAR	1369	869	133	101	242	2	22

Table 3.2: Effect numbers of dependency matrix for the in silico knock-out tests of the PKT38 model

Column 1 lists the selected gene for depletion; column 2 displays the total population of effect elements in the dependency matrix; column 3 shows the population of no effect elements in the dependency matrix; column 4 displays the total population of ambivalent factor elements in the dependency matrix; column 5 displays the population of weak inhibitor elements in the dependency matrix; column 6 shows the population of weak activator elements in the dependency matrix; column 7 displays the population of strong inhibitor elements in the dependency matrix and column 8 shows the population of strong activator elements in the dependency matrix. The value “null” in the column 1 represent the unmodified PKT38 model. The value “Null” in the selected gene column represents the wild-type.

The population of effect elements in the p53-null model was compared to the p53 wild type PKT38 model and it was found that there was no ambivalent factors in the dependency matrix of p53-null model, and the majority of effect elements turned to no effect because p53 was connected with the majority of nodes in the model and most of pathways were through p53 node. These finding indicates the central role of p53 on the whole network since all pathway from input signal to output effect of the model must contain p53. MDM2 was the second most connected node in the PKT38 model and its depletion caused fewer changes than the depletion of p53. ATM and p38MAPK functioned upstream of the p53 (Roos & Kaina, 2013), while BAX and TIGAR functioned downstream of the p53 (Chiu et al, 2003). However, those four nodes had less connectivity than MDM2 and the change of effect elements was less than the removal of MDM2. This phenomenon indicated that the node at the boundary of the network model had less effect on the whole network than those nodes which were located in the centre of the network. The position of the node determined the effect it caused onto internal dependency relationships inside the whole network

3.3.2 The *in silico* knock-out test for selected edges

In the next series of experiments the *in silico* knock-out tests for individual interaction depletion in the PKT38 model were performed. There were three main negative feedback cycles in the PKT38 model: p38-Wip-1-p53, p53-MDM2, and p14ARF-MDM4-p53-p21-CDK2-MDM2. Two types of interactions were classified in the PKT38 model according to the relationship with the negative feedback loop: the interaction involved in the negative feedback loops or the interaction not involved in the negative feedback loops. We aimed to perform two different types of knock-out test for selected edges

with this interaction classification. The first one is to investigate the effect of removal of interactions that are not involved in feedback loops, for example, the interaction that ATM activates p53 (Fig 3.3B) (Prives & Hall 1999).

Figure name	Edge removed	Is this edge involved in negative feedback loops?	The negative feedback loop involved
Figure 3.3A	Null	No	
Figure 3.3B	ATM activates p53;	No	
Figure 3.4	P38MAPK activates p53; P53 activates Wip-1; Wip-1 inhibits P38MAPK;	Yes	P38MAPK-p53-Wip-1
Figure 3.5	p53 activates MDM2; MDM2 inhibits p53;	Yes	p53-MDM2
Figure 3.6	MDM4 inhibits p53; P53 activates MDM2; MDM2 inhibits p53; p53 inhibits p14ARF; MDM4 activates MDM2; MDM2 inhibits MDM4; CDK2 inhibits MDM2; P21 inhibits CDK2; P14ARF inhibits MDM2;	Yes	P14ARF-MDM4-p53-p21-CDK2-MDM2
Figure 3.7	P38MAPK activates p53; P53 activates Wip-1; Wip-1 inhibits P38MAPK; MDM4 inhibits p53; P53 activates MDM2; MDM2 inhibits p53; p53 inhibits p14ARF; MDM4 activates MDM2; MDM2 inhibits MDM4; CDK2 inhibits MDM2; P21 inhibits CDK2; P14ARF inhibits MDM2;	Yes	P38MAPK-p53-Wip-1; P14ARF-MDM4-p53-p21-CDK2-MDM2

Table 3.3: Knock-out tests for selected edges.

Table 3.3 lists 5 knock-out tests for selected edges. Column 1 lists the figure name corresponding to the knock-out test. Column 2 shows interactions removed from the PKT38 network. Column 3 shows the relationship between removed interactions and negative feedback loops. Column 4 shows negative feedback loops which contain removed interactions.

Those interactions, which were not involved in negative feedback cycles, were mostly connected with nodes whose connectivity degree was low, for example DNA damage activates ATM, ATM activates p53, p53 activates BAX and BAX activates apoptosis. Figure 3.3B was compared with Figure 3.3A either visually or by Java programme interfaces. There was no change in the effect elements because the row of ATM and the column of p53 were removed from the default dependency matrix (Figure 3.3A) and the other dependency relationships remained in Figure 3.3B for other genes were not affected by this depletion. As a result, these interactions not involved in feedback loops may not be the most important edges in the p53 network.

Then the second *in silico* knock-out tests were focused on interactions involved in these three negative feedback cycles mentioned above. In order to explore the role of negative feedback loops play in the p53 pathway, knock-out tests for these three feedback loops individually were performed at first, and then we removed all those three negative feedback cycles.

In the first knock-out test for negative feedback loop depletion, we removed the negative feedback loops, p38MAPK1-p53-Wip-1 from the PKT38 model. Table 3.4 lists all the interactions removed in order to knock out the negative feedback loop of p38MAPK-p53-Wip-1 and the major changes in effect cells were listed in Table 3.4 below. One major change was found that the effect cell from UV onto p38 MAPK changed from ambivalent factor in the p53 wild type to strong activator in the modified model when the negative feedback loop of p38MAPK-p53-Wip-1 was removed. The distribution of the effect changes was shown in Fig 3.4C.

Source node of deleted interaction	**Target node of deleted interaction**	**Deleted interaction type**
P38MAPK	P53	+1
P53	Wip-1	+1
Wip-1	P38MAPK	-1

Table 3.4: Edges removed from the PKT38 model for the model in Fig 3.4A

Table 3.4 lists three interactions removed from the PKT38 model so as to remove the negative feedback loop of p38MAPK-p53-Wip-1. Column 1 shows the name of the source node of the deleted interaction; column 2 shows the name of the target node of the deleted interaction and

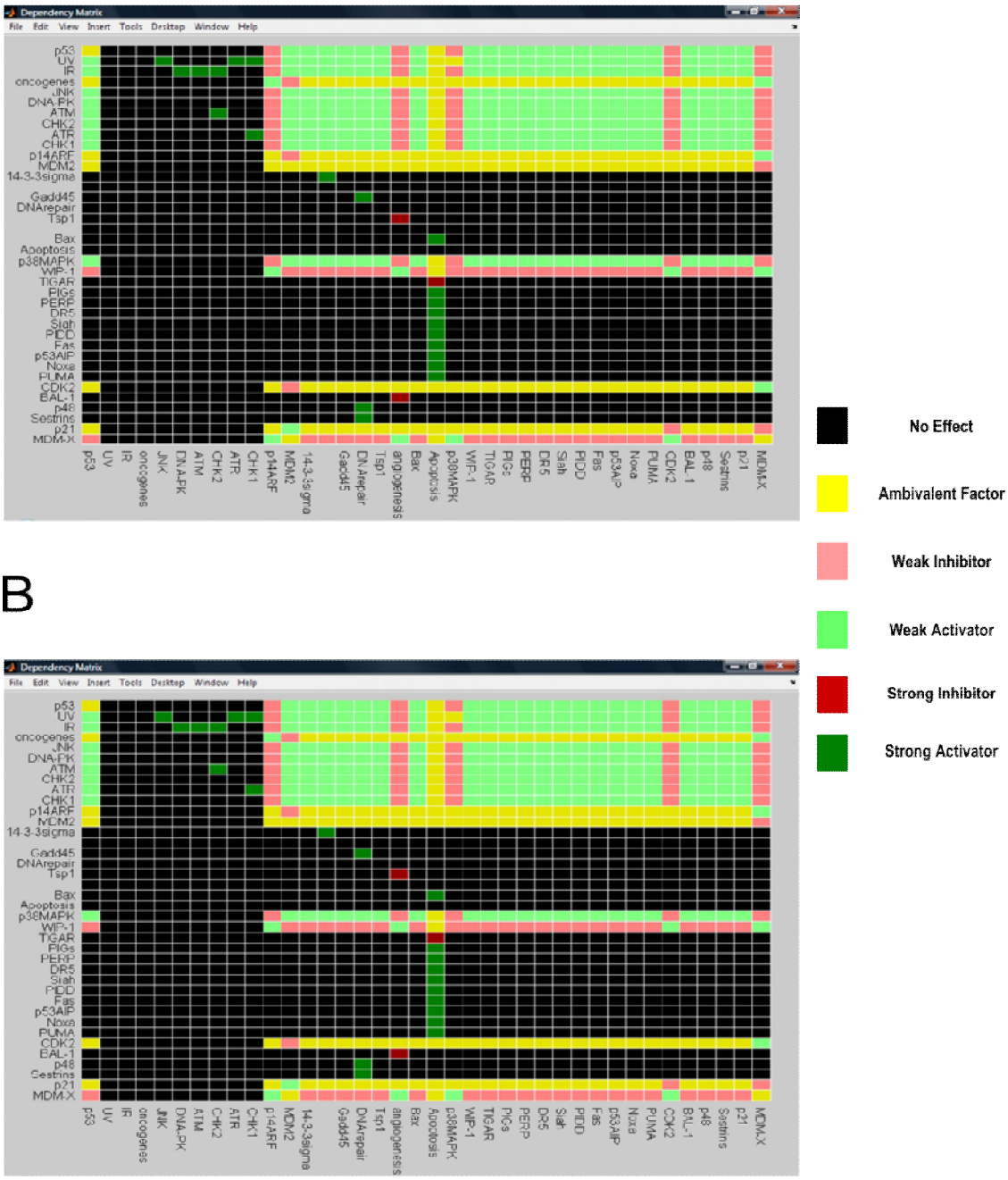
column 3 list the interaction type.

Source node	**Target node**	**Effect cell in the p53 wild type of the PKT38 model**	**Effect cell in the *in silico* knock-out test when the negative feedback loop of p38MAPK-p53-Wip-1 was removed**
UV	p38MAPK	Ambivalent Factor	Strong activator

Table 3.5: Effect elements changed from graphic view in Figure 3.4B when the p38-p53-WIP-1 negative feedback loop was removed.

This table lists the change of effect element observed for the *in silico* knock out test described in Figure 3.4B. Column 1 shows the name of the source node of the deleted interaction; column 2 shows the name of the target node of the deleted interaction; column 3 shows the effect cell in the p53 wild type of the PKT38 model and column 4 list the effect cell in the *in silico* knock-out test when the negative feedback loop of p38MAPK-p53-Wip-1 was removed.

A

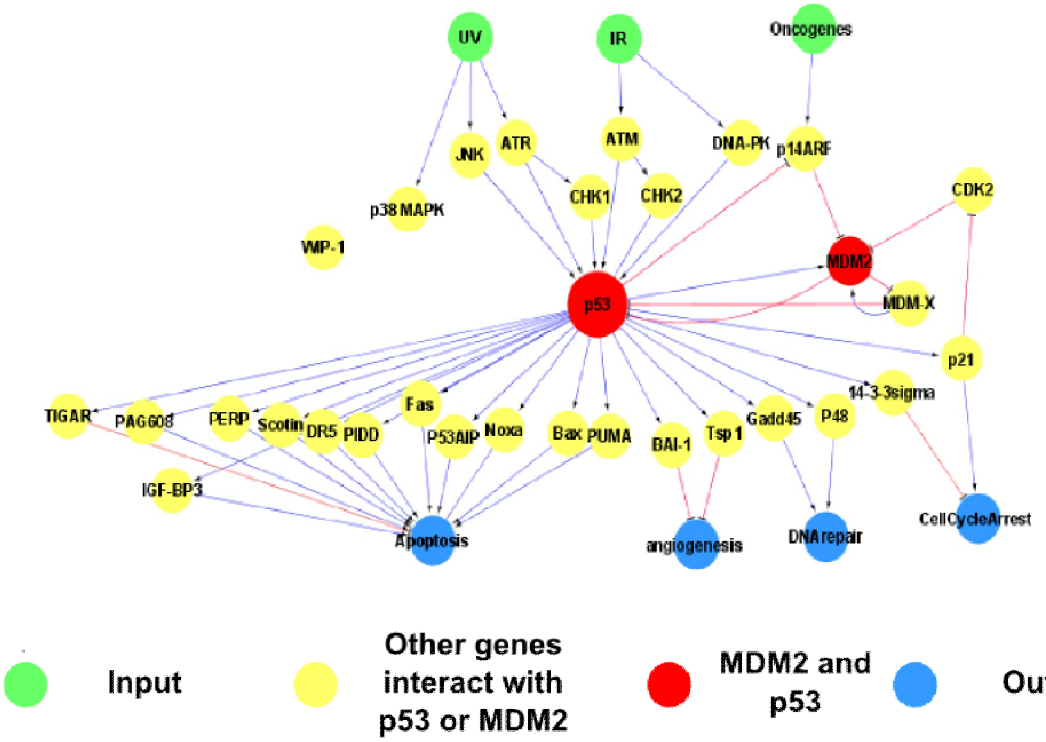


B

Figure 3.3: *in silico* knock-out of nodes not involved in negative feedback loops

Figure 3.3A shows the dependency matrix corresponding to the unmodified PKT38 model and Figure 3.3B shows the dependency matrix corresponding to the PKT38 model when the interaction that ATM activates p53 was removed. There were no major changes found in Figure3.3B.

A



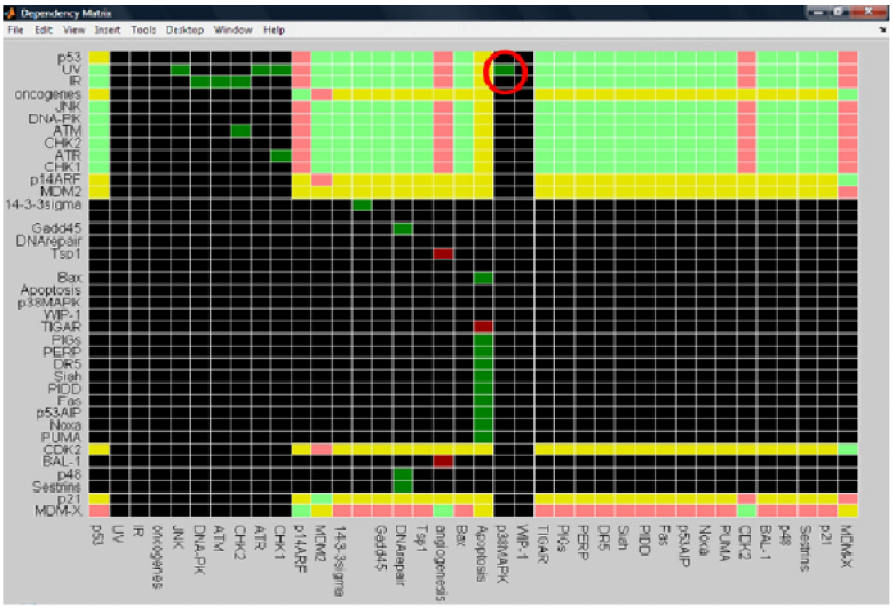
Input

Other genes interact with p53 or MDM2

MDM2 and p53

Output

B



No Effect

Ambivalent Factor

Weak Inhibitor

Weak Activator

Strong Inhibitor

Strong Activator

C

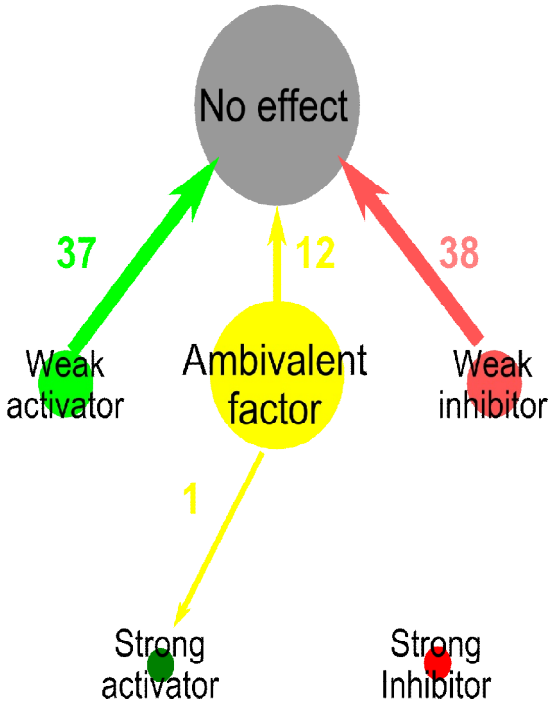


Figure 3.4: Knock-out tests for the p38MAPK-p53-Wip-1 negative feedback cycles

Figure 3.4A shows a network map showed the PKT38 model without the negative feedback loop of p38MAPK-p53-Wip-1. Figure 3.4B shows the dependency matrix corresponding to the modified PKT38 model when the negative feedback loop of p38MAPK-p53-Wip-1was deleted from the PKT38 model and the major change was marked by a red circle. Figure 3.4C shows the distribution of effect changes between the default dependency matrix and the new dependency matrix when the negative feedback loop p38MAPK-p53-Wip-1 was removed from the PKT38 model. The gray circle represents no effect elements, the yellow circle represents ambivalent factors, the turquoise green circle represents weak activators, the pink circle represent weak inhibitors, the red circle represents strong inhibitors, and the dark green circle represents strong activators; the direction of the arrow represents the direction of changes in the knock-out.

In the second type of knock-out test with individual negative feedback loop depletion, the edge between p53 and MDM2 was removed from the PKT38 model so that there was no negative feedback loops between MDM2 and p53. Table 3.6 lists all the interactions removed in order to knock out the negative feedback loop of p53-MDM2.

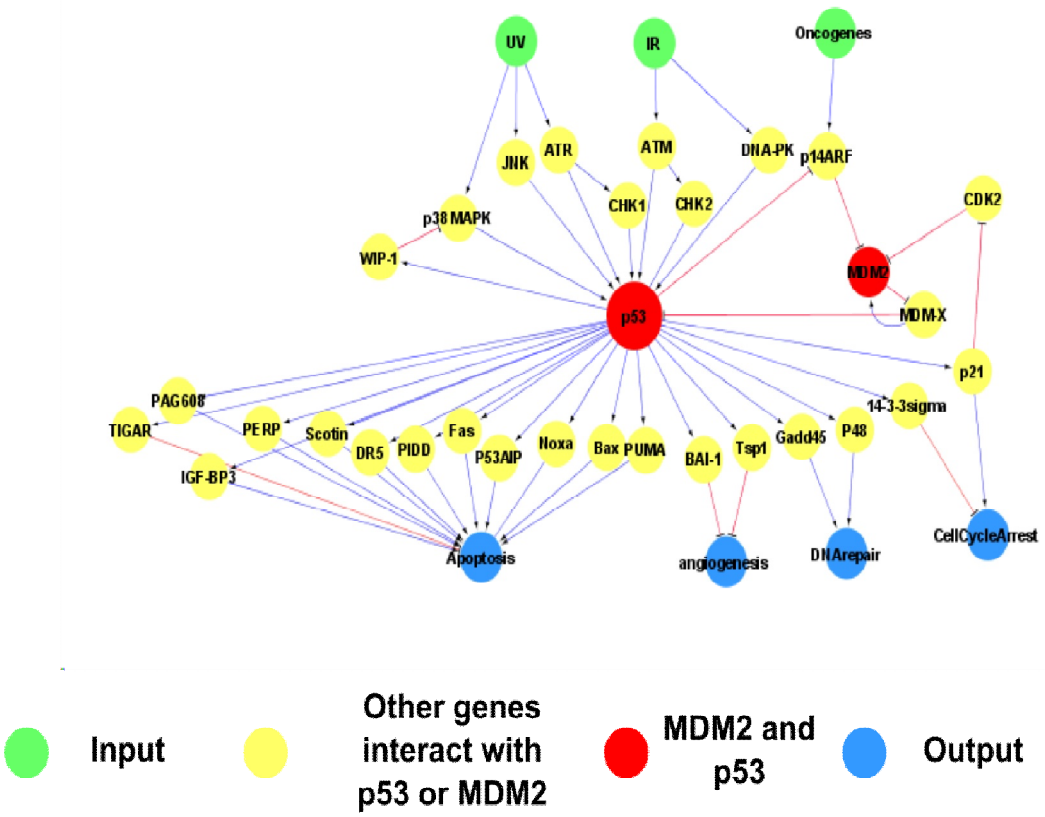
Source node of deleted interaction	Target node of deleted interaction	Deleted interaction type
MDM2	P53	-1
P53	MDM2	+1

Table 3.6: Edges removed from the PKT38 model for the model in Figure 3.5A

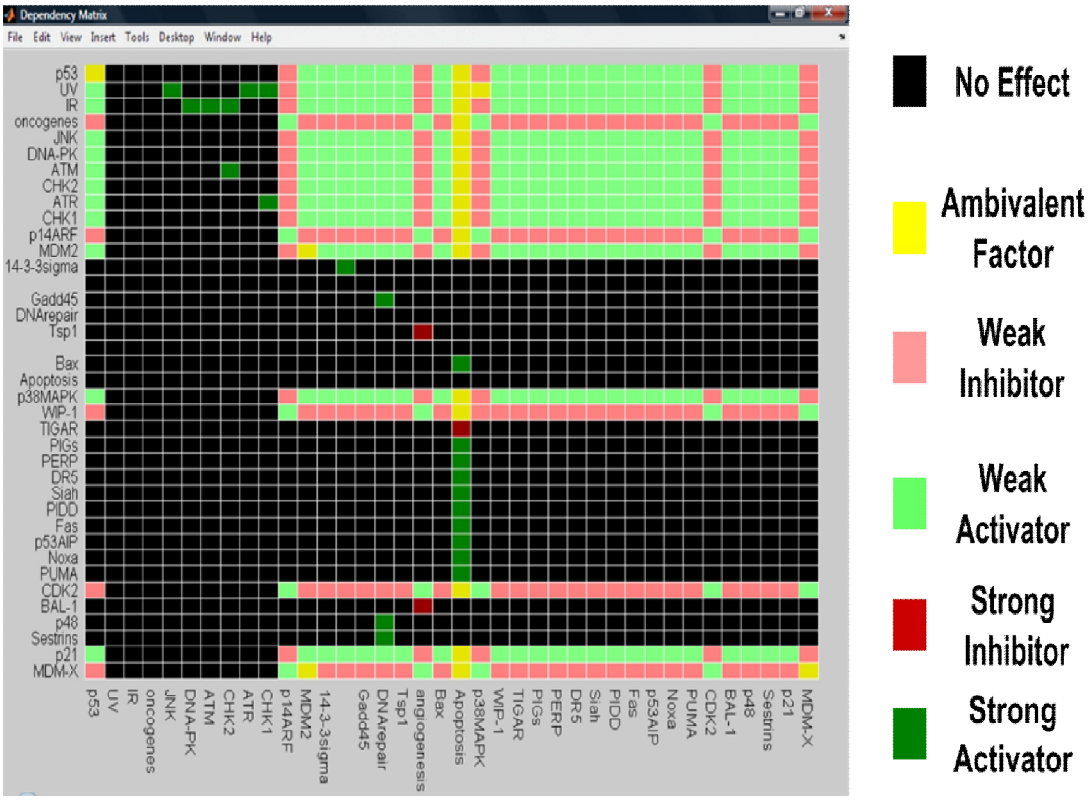
Table 3.6 lists two interactions removed from the PKT38 model so as to remove the negative feedback loop of p53-MDM2.

By comparing the dependency matrix for the modified model (Fig 3.5B) with the unmodified model Fig 3.3A), it was found that only 108 ambivalent factors changed to weak inhibitor or weak activator (Fig3.3C), and none of the effect cells changed to no effect cells (Fig 3.5C). As a result, we found that the negative feedback loop between p53 and MDM2 kept a large amount of dependency relationships between nodes at a stable state, which was represented by ambivalent factors in the dependency matrix.

A


UV
IR
Oncogenes
JNK
ATR
ATM
DNA-PK
p14ARF
p38 MAPK
CHK1
CHK2
CDK2
WIP-1
MDM2
p53
MDM-X
p21
PAG608
TIGAR
PERP
Scotin
DR5
PIDD
Fas
P53AIP
Noxa
Bax
PUMA
BAI-1
Tsp1
Gadd45
P48
14-3-3sigma
IGF-BP3
Apoptosis
angiogenesis
DNArepair
CellCycleArrest
Input
Other genes interact with p53 or MDM2
MDM2 and p53
Output

B


Dependency Matrix
File Edit View Insert Tools Desktop Window Help
p53
UV
IR
oncogenes
JNK
DNA-PK
ATM
CHK2
ATR
CHK1
p14ARF
MDM2
14-3-3sigma
Gadd45
DNArepair
Tsp1
Bax
Apoptosis
p38MAPK
WIP-1
TIGAR
PIGs
PERP
DR5
Siah
PIDD
Fas
p53AIP
Noxa
PUMA
CDK2
BAL-1
p48
Sestrins
p21
MDM-X
angiogenesis
No Effect
Ambivalent Factor
Weak Inhibitor
Weak Activator
Strong Inhibitor
Strong Activator

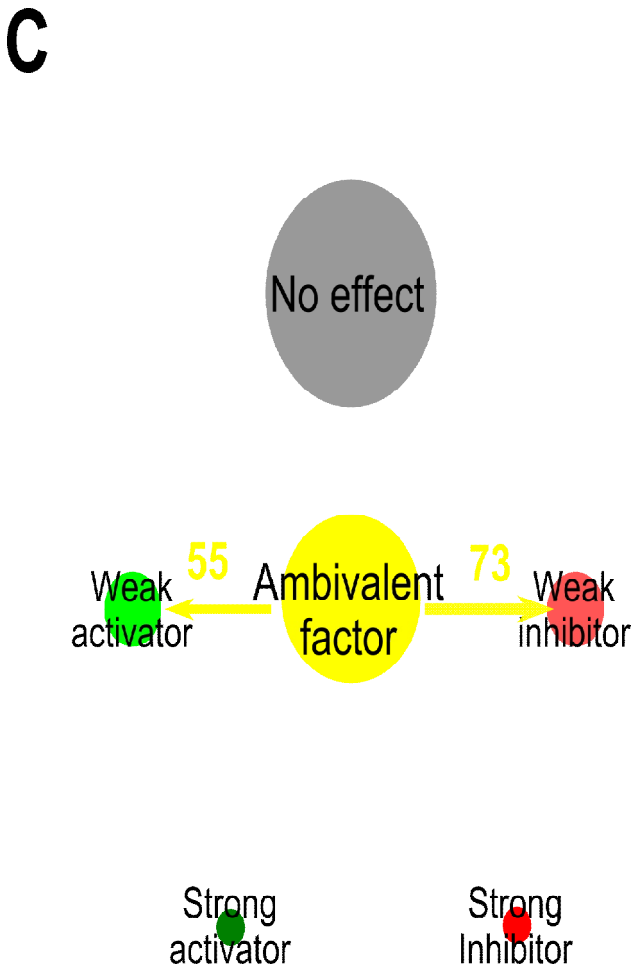


Figure 3.5: Knock-out tests for the p53-MDM2 negative feedback cycles

Figure 3.5A shows the PKT38 model without the negative feedback loop of p53-MDM2. Figure 3.5B shows a dependency matrix for the modified PKT38 model to investigate the role of the p53-MDM2 negative feedback loop. Figure 3.5C shows the distribution of effect changes between the default dependency matrix and the new dependency matrix when the negative feedback loop p53-MDM2 was removed from the PKT38 model.

Then we removed seven edges to determine the role of extended p14ARF-MDM4-p53-p21-CDK2-MDM2 negative feedback loop in p53 function. Table 3.7 below listed all the interactions removed in order to knock out the negative feedback loop of p14ARF- MDM4-p53-p21-CDK2-MDM2.

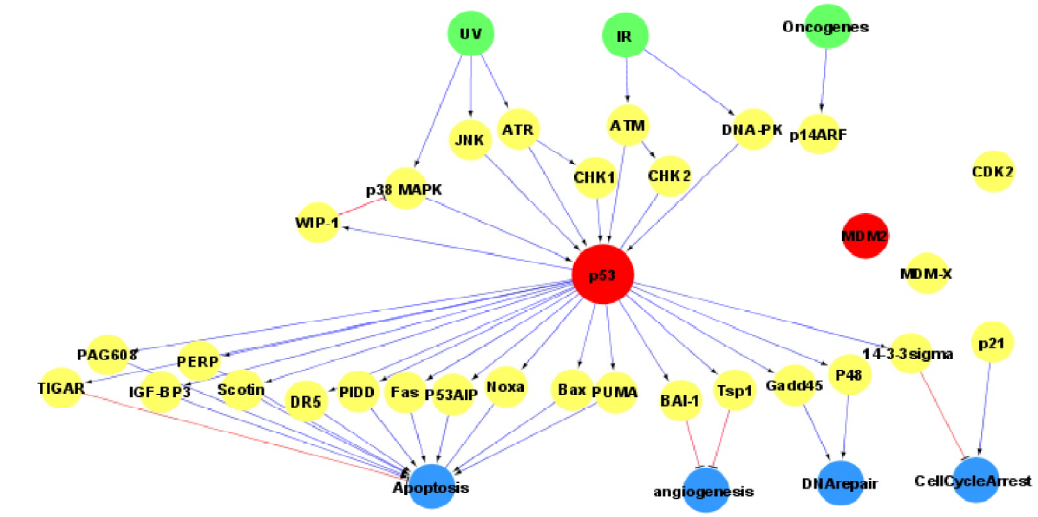
Source node of deleted interaction	Target node of deleted interaction	Deleted interaction type
P53	MDM2	+1
MDM2	P53	-1
P53	P14ARF	-1
MDM4	MDM2	+1
MDM2	MDM4	-1
CDK2	MDM2	-1
P21	CDK2	-1
P14ARF	MDM2	-1
MDM4	P53	-1

Table 3.7: Edges removed from the PKT38 model for the model in Fig 3.6A

Table 3.7 lists eight interactions removed from the PKT38 model so as to remove the negative feedback loop of p14ARF- MDM4-p53-p21-CDK2-MDM2.

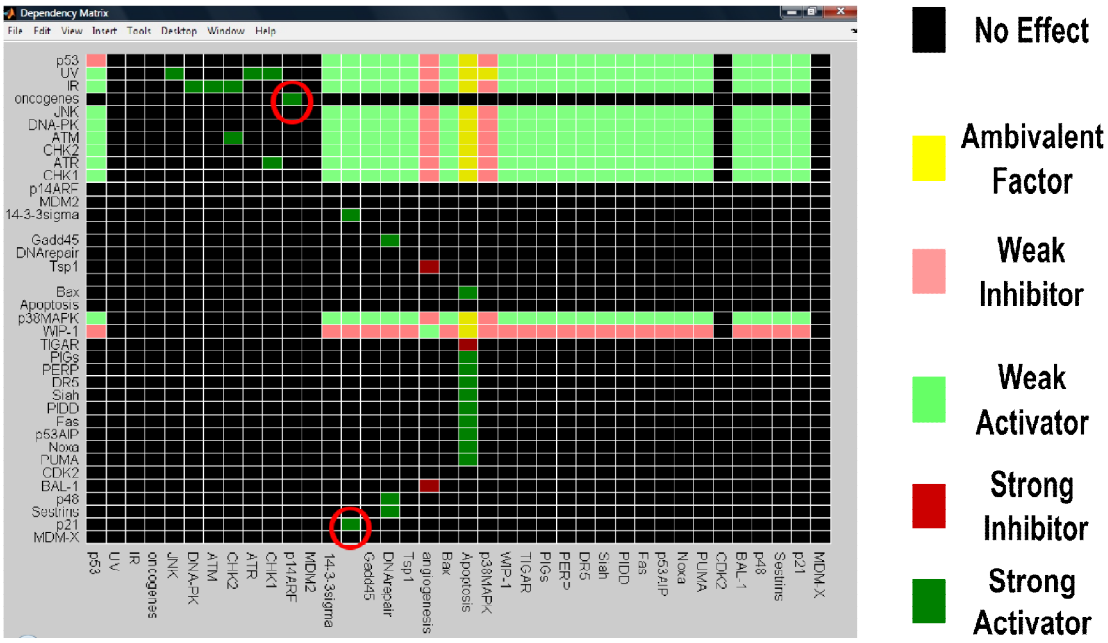
As there were more interactions involved in this negative feedback loop than the previous two negative feedback loops, the population of effect changes increased when compared with the knock-out tests of these previous two negative feedback loops and two major changes were found (Table 3.8). The effect from oncogenes onto p14ARF was strengthened from weak activator to strong activator. The effect from p21 onto cell cycle arrest was enhanced from ambivalent factor to strong activators.

A



Input | Other genes interact with p53 or MDM2 | MDM2 and p53 | Output

B



C

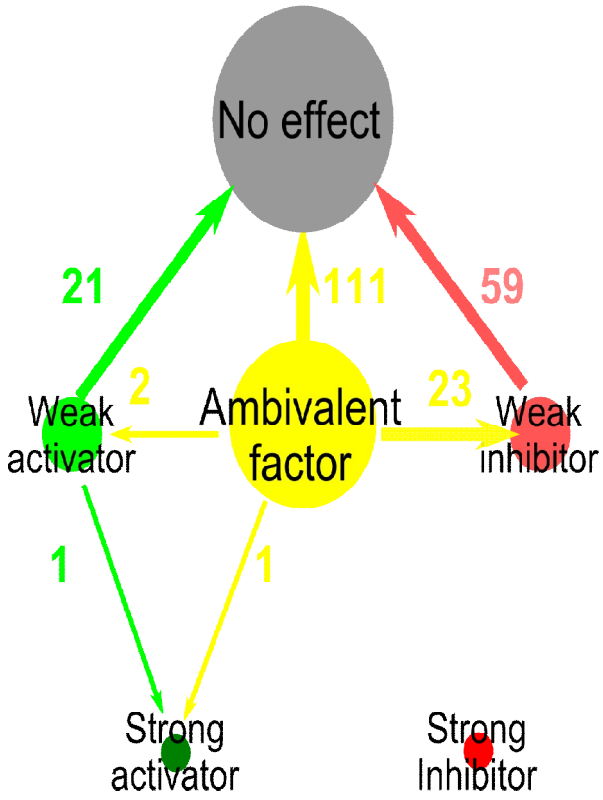


Figure3.6: **Knock-out tests for the p14ARF- MDM4-p53-p21-CDK2-MDM2 negative feedback cycles**

Figure 3.6A shows the PKT38 model without the negative feedback loop of p14ARF-MDM4-p53-p21-CDK2-MDM2. Figure 3.6B shows a dependency matrix for the modified PKT38 model to investigate the role of the p14ARF- MDM4-p53-p21-CDK2-MDM2 negative feedback loop. Two major changes were marked by red circles. Figure 3.6C shows effect change distribution was shown the distribution of effect changes between the default dependency matrix and the new dependency matrix when the negative feedback loop of p14ARF- MDM4-p53-p21-CDK2-MDM2 was removed from the PKT38 model.

Source node	Target node	Effect cell in the p53 wild type of the PKT38 model	Effect cell in the *in silico* knock-out test when all negative feedback loops were removed
Oncogenes	P14ARF	Weak Activator	Strong Activator
P21	Cell Cycle Arrest	Ambivalent Factor	Strong Activator

Table 3.8: Effect elements changed from Figure 3.7C when the p14ARF- MDM4-p53-p21-CDK2-MDM2 negative feedback loop was removed from the PKT38 model

This table list all major changes observed from the *in silico* knock-out test with depletion of the p14ARF- MDM4-p53-p21-CDK2-MDM2 negative feedback cycles..

Moreover, in order to determine the role of all negative feedback loops in p53 function, a sub network from the PKT38 model was constructed without all negative feedback loops. Table 3.8 lists all the interactions removed in order to knock out all the negative feedback loops in the PKT38 model.

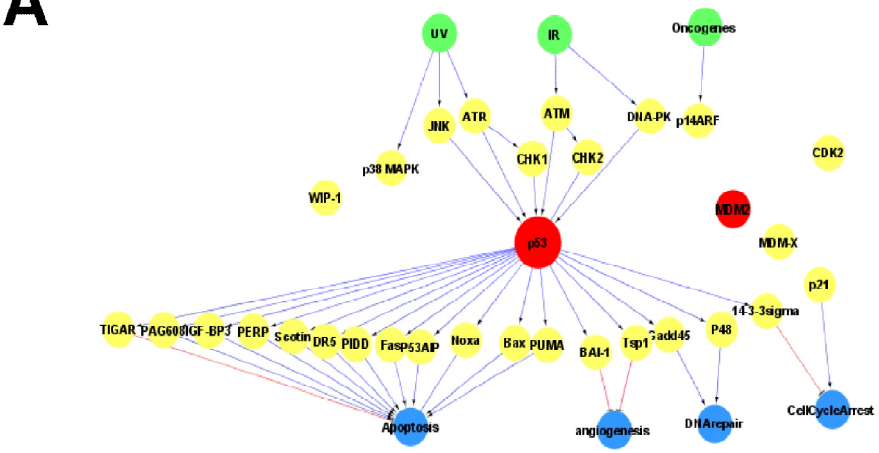
Source node of deleted interaction	Target node of deleted interaction	Deleted interaction type
P38MAPK	P53	+1
P53	Wip-1	+1
Wip-1	P53	-1
P53	MDM2	+1
MDM2	P53	-1
P53	P14ARF	-1
MDM4	MDM2	+1
MDM2	MDM4	-1
CDK2	MDM2	-1
P21	CDK2	-1
P14ARF	MDM2	-1
MDM4	P53	-1

Table 3.9: Edges removed from the PKT38 model for the model in Fig 3.7A

Table 3.9 lists eleven interactions removed from the PKT38 model so as to remove all the negative feedback loops.

Additional Table 2 shows the major effect changes from the default dependency matrix in Figure 3.3A to the dependency matrix of modified model in Figure 3.7B.It is found that population of ambivalent factors decreased faster than the decrease of the weak inhibitor and weak activators.

A



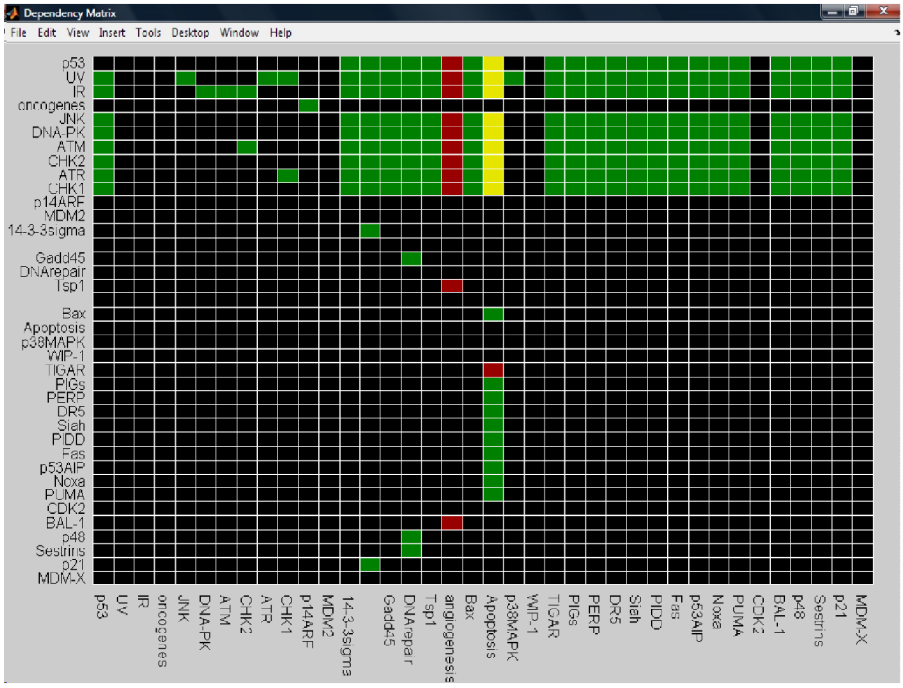
Input

MDM2 and p53

Other genes interact with p53 or MDM2

Output

B



No Effect

Ambivalent Factor

Weak Inhibitor

Weak Activator

Strong Inhibitor

Strong Activator

C

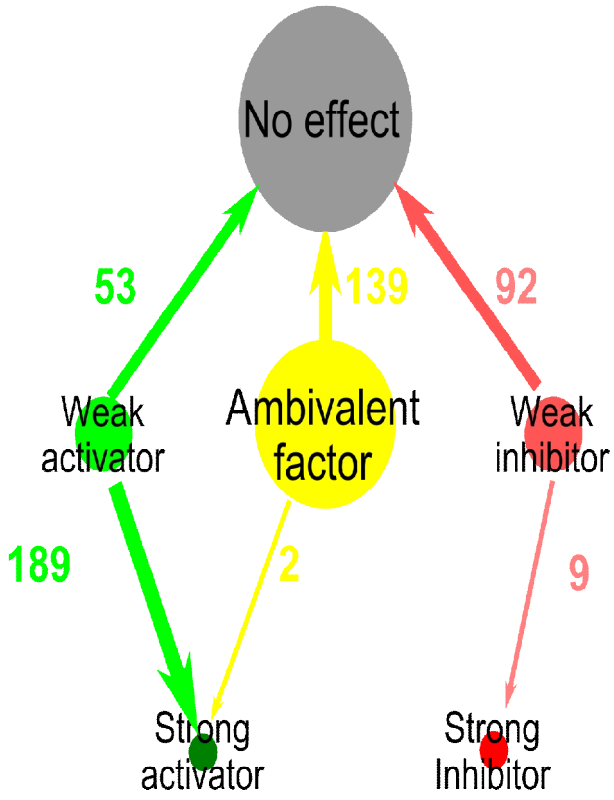


Figure 3.7: The knock-out tests without all negative feedback loops

Figure 3.7A shows the PKT38 model without all negative feedback loops. In Figure 3.7B, removing all feedback loops transforms weak activatory and inhibitory effects into strong ones. These results suggest that the inhibitions that are included in the feedback loops make the network more stable. Figure 3.7C shows the distribution of effect changes between the default dependency matrix and the new dependency matrix when all the negative feedback loops were removed from the PKT38 model.

Comparing the modified dependency matrix of the PKT38 model without feedback loops with the default one in p53 wild type (compare Figure 3.7B to Figure 3.3A and see Additional Table 2), we can conclude that the feedback loops in the p53 network play an essential role in keeping the network more stable to external perturbations. As a result, disturbing negative feedback loops in the p53 pathway may increase the sensitivity of cells to cancer treatments, such as UV and other anticancer drugs. Targeting on negative feedback loops in the p53 pathway may contribute to the improvement of cancer treatment and therapy design. Some of those applications were already identified and revealed a useful role for clinical use (Lu, 2010). For example, the peptides inhibit MDM2 described in the introduction, such as D-peptide inhibitor, and MIP (MDM2 inhibitory peptide) which target on the negative feedback loop between p53 and MDM2 to activate p53 as a tumour suppressor.

3.4 The PKT62 model

In the previous section, we constructed the PKT38 model and carried out the *in silico* knock-out analysis. Exploration of the pathways described in the PKT38 model indicated that more interactions needed to be included in order to represent realistically the complexity of the p53 pathway. Therefore more extensive literature searches were performed and a new model PKT62 was constructed. Here we describe the construction of the PKT62 model, its structure and the *in silico* analysis results. All interactions included in this model were investigated in at least one paper from the PubMed database. Figure 3.8 describes the PKT62 model and this network is based on the literature survey of more than 500 papers (Additional Table 3). Some genes or proteins, for example, DDR1 (discoidin domain receptor 1), which is represented as a node in the network above, was found to interact with p53 in only two published papers from PubMed. However for other interactions such as p21(CDKN1A)-p53 interaction, there was 8195 papers in PubMed, which both mentioned p53 and p21 and 306 papers stating that p53 activates p21. There were 62 nodes and 109 interaction edges included in the PKT62 (Figure 3.8 and Additional Table 3). This PKT62 model has three input nodes, which were DNA damage, other cellular stress, and oncogenes, and four output nodes: apoptosis, angiogenesis, DNA repair and cell cycle arrest. As MDM2 had negative two step feedback loops with p53, it was represented by red colour to indicate its role in the whole network.

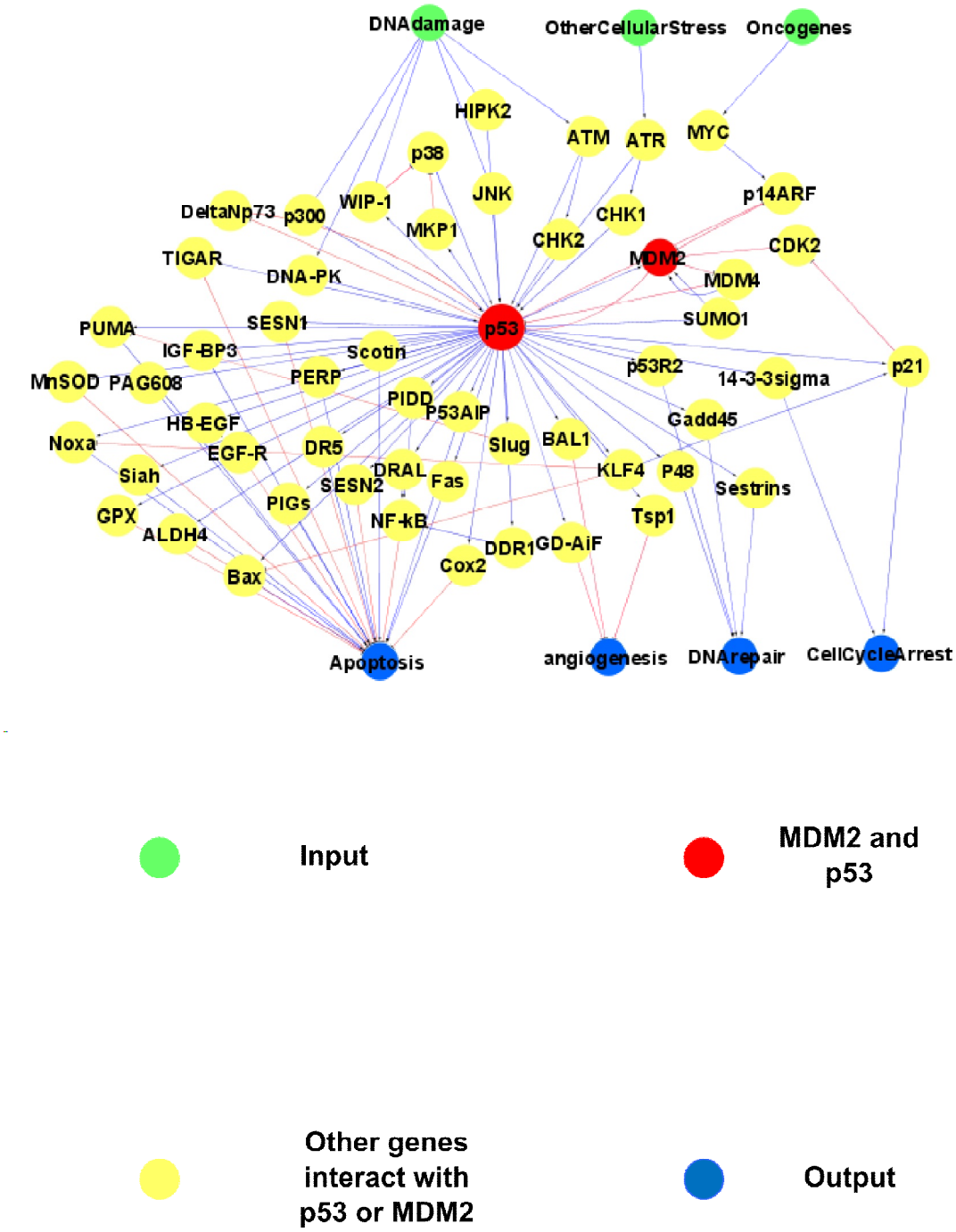


Figure 3.8: Network map of the PKT62 model

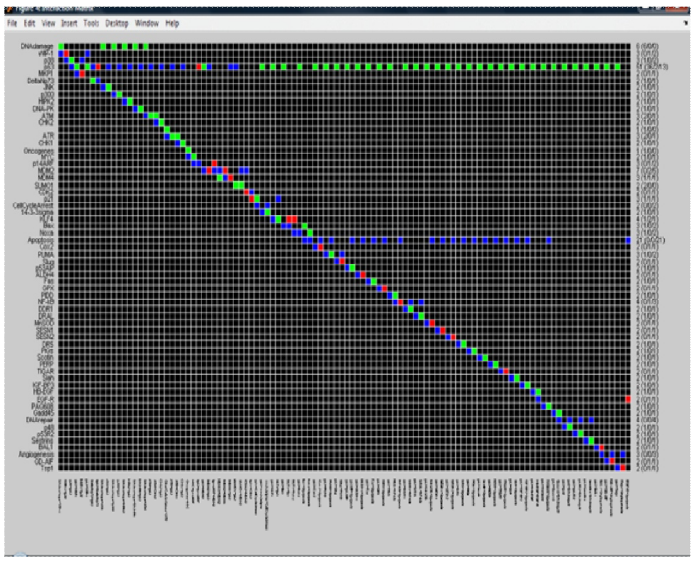
Figure 3.8 shows a network map of the PKT62 model. Three input nodes were marked by green colour and all inter medium nodes were marked by yellow colour except MDM2, which was marked by red colour. The four output node was marked by blue colour.

Next, the interaction and dependency matrix for PKT62 model was calculated to obtain a global view for the trend of the p53 network when the network became more complex than the PKT38 model towards construction of complete model describing p53 interactome (Figure 3.9A). In the Figure 3.9B, it is found that the majority of effect elements are no effect and the population of strong inhibitor and strong activator is quite small, compared with the population of weak inhibitor and weak activator. Moreover, the *in silico* knock-out tests were also performed to investigate the role of p53 and MDM2 (Figure 3.10, Figure 3.11). According to these two dependency matrices, two results were obtained. The first one is the important role of p53 effect onto the whole network. Comparing the dependency matrix in Figure 3.10 with Figure 3.9B, it was found that the majority of effect cells in the dependency matrix of the p53 null model became no effect. The effect cells of strong inhibitor and strong activator were still remaining in the dependency matrix. We could refer that those dependency relationships were independent of the p53 presence, for example, the effect from DNA damage onto ATM was strong activator and this relationship remained the same in the dependency matrix when p53 was removed. The second finding was that the depletion of MDM2 caused less changes than the removal of p53 on the whole network. Comparing Figure 3.10 and Figure 3.11, a large amount of effect cells remained the same in the knock-out test of MDM2 as in the dependency matrix of the p53 wild type. The node of p53 took part in 51 interactions in the PKT62 model, whereas MDM2 was only connected with 7 interactions. It indicated that the node with high connectivity degree may affect more onto the dependencies relationship between nodes in the network. This phenomenon was explored further in the next chapter of results.

The results shown in Figure 3.9, 3.10 and 3.11 confirmed features of the p53

pathway uncovered by the PKT38 model and highlighted the importance of p53 presence for the robustness of the whole network.

A



Inhibition Input

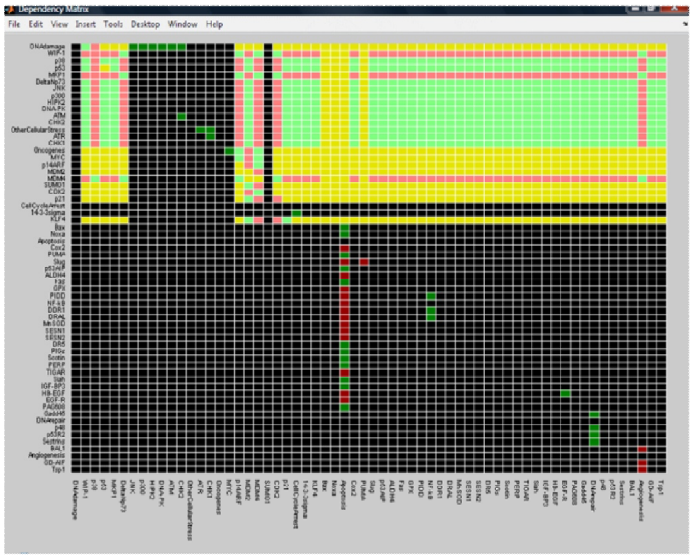
Activation Input

Output

No participation

B

No Effect



Ambivalent Factor

Weak Inhibitor

Weak Activator

Strong Inhibitor

Strong Activator

Figure 3.9: Analysis result of the PKT62 model by CellNetAnalyzer

Figure 3.9A describes the interaction matrix of p53 network at low confidence level, which was constructed by CellNetAnalyzer. Figure 3.9B shows the dependency matrix of the PKT62 model in Figure 3.8.

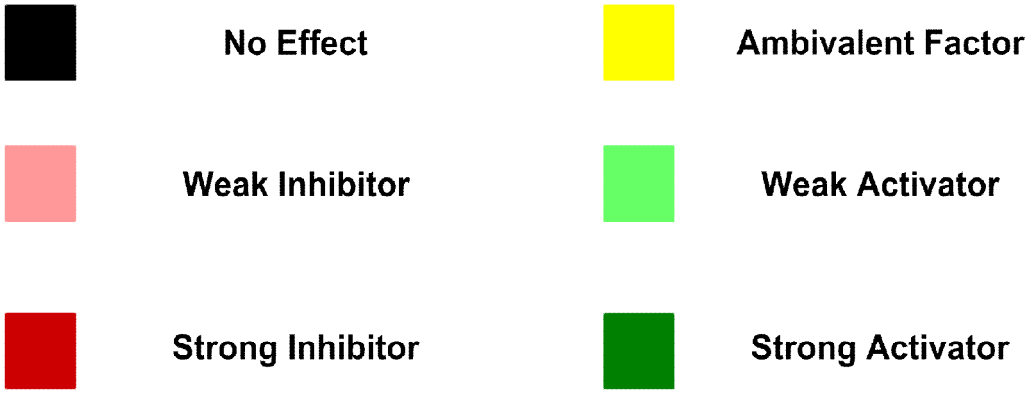


Figure 3.10: The default dependency matrix of the PKT62 model in the absence of p53.

This dependency matrix is respond to the PKT62 model when p53 and all interactions with p53 were removed from this model.

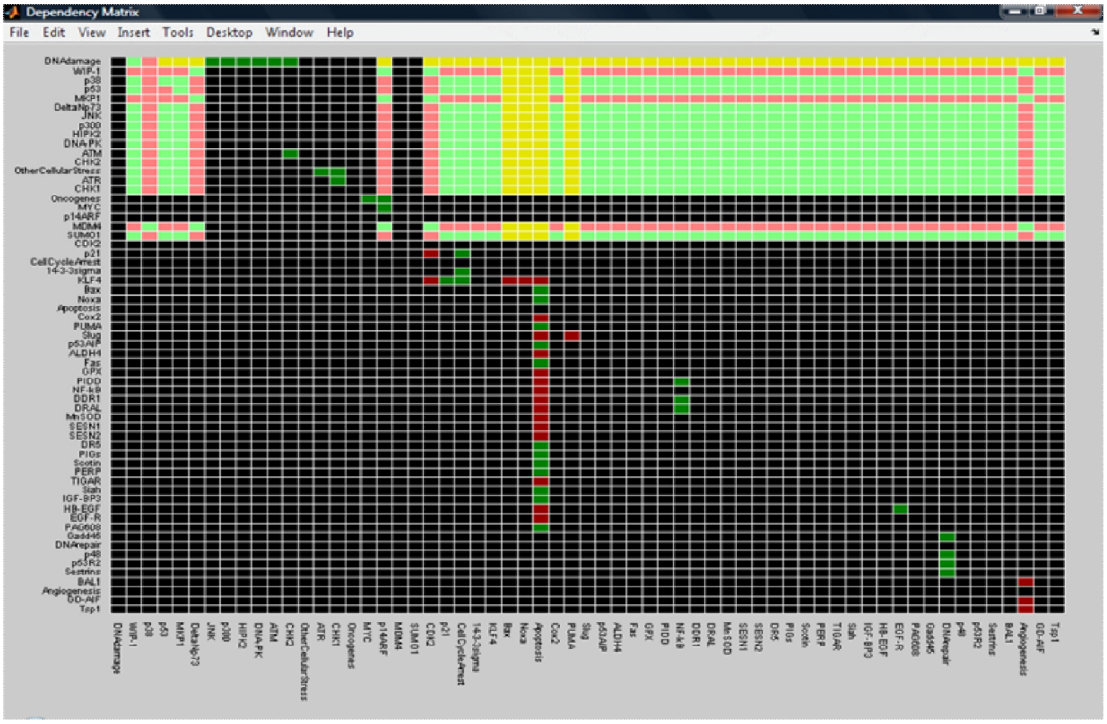


Figure 3.11: The default dependency matrix of the PKT62 model without MDM2.

This dependency matrix based on the PKT62 model was calculated when MDM2 and all interactions with MDM2 were removed from this model.

3.5 Discussion

In order to understand the cellular functions and interactions of p53 tumour suppressor, simple models for p53 pathways, PKT38 and PKT62 were built to investigate all the interactions in p53 pathways from manual literature searches. The dependency relationships between those models were analyzed using the interaction matrix and dependency matrix. In a global view of the PKT38 and PKT62 models, the number of strong inhibitor cells and strong activator cells constituted a minority part of the total effect elements in the dependency matrix. As a result, the changes which affected creation of strong inhibitors and strong activators were a hallmark to monitor the changes of network caused by perturbations in the network. Focusing on changes to strong inhibitor cell and strong activator cells facilitated the investigation of changes in dependency relationship in response to perturbations caused by modifications in the network structure.

The advantage of the PKT38 model and PKT62 models is that they provide a small-scale view of p53 network structure: input signals, upstream of p53, p53 and MDM2, downstream of p53 and output of the network based on the manual literature survey. *In silico* knock-out tests were performed on this model and the analysis result revealed the role of certain nodes and edges in the network model. Those manually curated models provide evidence that it is feasible to make p53 interactome with predictive properties from selected databases using text mining or other tools. Moreover, it revealed the crucial role of negative feedback loops for the robustness of the p53 whole network to perturbations according to the *in silico* knocks-out tests. By comparing the dependency matrix of modified PKT38 model with the unmodified one, we can

conclude that the negative feedback loops keep the network less sensitive to changes of the input signals. Similarly, Zhang (2011) in his paper pointed that more negative feedback loops of p53 increased the robustness of the p53 oscillation (Zhang et al, 2011). Their finding supported our conclusion that negative feedback loops with p53 make the whole network more robust to the perturbation. As MDM2 was the second most connected node in the PKT38 model, the knock down of it lead to more changes in the dependency relationships between other genes in the PKT38 model than the depletion of ATM or BAX. The *in silico* knock out test for the negative feedback loop between p53 and MDM2 indicated that this negative feedback loop functioned to keep the effects between other proteins in a balanced state. This finding demonstrated that the negative feedback loop between p53 and MDM2 make the whole network robust to experimental perturbations under the environmental of DNA damage (Wagner et al, 2005).

The process of PKT38 model and PKT62 model building demonstrated the feasibility of using a system biology approach to investigate the p53 pathway. The advantage of those two models was that the dependency matrix of *in silico* knock-out tests provided a direct view of the global effect changes. However, for larger networks it is impossible to see such changes visually. Once the scale of the model increased from 38 nodes to 62 nodes, it became difficult to identify all changed effects from the dependency matrix figures. Moreover, although more than 500 papers were manually curated to construct the basic model PKT38 and PKT62, the structure of these models is quite simple when compared with the real p53 pathway. Those models did not consider all possible interactions relevant to p53 and the genes or proteins that interact with p53.

Chapter 4 Early versions of the PKT205 model based on interaction information retrieved automatically from STRING

4.1 Introduction

In the previous chapter, the PKT38 model described the structure of the p53 pathways and revealed the crucial role of p53 and negative feedback loops in the whole system. However, this model was small and limited by the literature based information retrieval. Compared with 67000 papers about p53 published in PubMed and the frequency of new reports about p53 published, this simple model has limited use in contributing to further understanding of the p53 pathway mechanisms and to provide continuous updates with the increasing speed of new papers about p53 published. The population of published papers about p53 is increasing every year (Table 1.1). It was found that thousands of papers about p53 have been published every year in the last 15 years. There was an increasing strand in the last 15 years (Table 1.1). Since the information retrieval is time consuming and the conclusion of interaction type is limited by current experimental conditions, there is an urgent requirement for automated retrieval of p53 interaction information by text mining tools or other approaches. As a result, various text mining tools and databases were investigated and a Java based programme was designed to retrieve interaction information automatically from selected database.

We assumed that there existed databases which were able to provide datasets with high confidence interaction information involving p53. The hypothesis was that large numbers of p53 interactions could be retrieved by natural language processing tools and enables us to construct a “complete”

model, which would be a better representation of the real network than the previous models. The investigation process was divided into four main steps:

At the first step, we investigated all available protein-protein databases which provided interactions with p53. Their performances were compared by the following criteria:

Firstly, the database should provides p53 protein-protein interactions records. Secondly, the database should provide convincing literature evidence for these interactions; thirdly, those interactions should indicate the nature of the relationship between proteins (activation or inhibition). Then certain databases were selected among the candidate ones as our resource for interaction information.

At the second step, we constructed logical Boolean models for p53 which included protein-protein interactions relevant to p53, according to the selected database.

The third step was to perform analysis using CellNetAnalyzer to explore the features of the model.

The fourth step was to improve the model using feedback from the analysis results so that predictions could be made for the biological functions of p53 pathways.

4.2 Construction of the PKT205 model by automatic information retrieval from STRING

In Chapter 1, STRING was chosen as a database most suitable for extraction of interaction information relevant to p53. Here the p53 interactions were automatically retrieved and classified in order to generate a draft of p53 interactome, using logical Boolean models. After STRING was selected as the main resource of interaction information, the process flow of interaction extraction, model construction and analysis was designed as shown below (Figure 4.1).

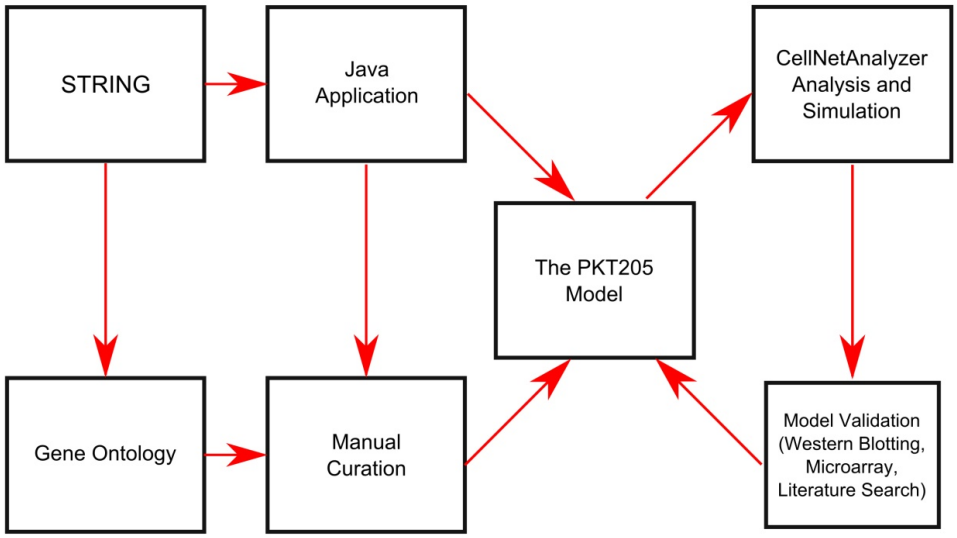


Figure 4.1: Process flow chart for the PKT205 model construction and improvement.

Figure 4.1 shows the process of model construction, simulation and analysis. Java interface programs were created to extract p53 interactions from the STRING database.

We then manually curated the data and used Gene Ontology annotations to connect the network to DNA damage input and apoptosis output. CellNetAnalyzer was used for analysis and simulations, and the results were validated using literature surveys and experimental approaches including

western blotting and microarray analysis.

Since some important protein-protein interaction were identified as post translational modification by STRING, we also turned to other available on line text mining tools, like PubMed and KEGG, to confirm the nature of these posttranslational modifications. For example, the interaction between ATM and p53 is a phosphorylation, and this phosphorylation leads to the accumulation of p53. As a result, we assumed that the interaction from ATM to p53 is activation and included it as such in our model.

4.3 Six type of errors caused by text mining

During the process of manual curation, six types of errors caused by text mining were found and they were sorted according to their frequency of occurrence in text mining:

- The first error was wrong gene name recognition. Most errors were of this type;
- The second one was wrong target recognition;
- The third error was a general term appearing in a complex context;
- The fourth error was negation and negative words not recognized by text mining;
- The fifth error was wrong relationship;
- And the sixth error was speculation or question.

There are several possible reasons that cause errors during the period of natural language processing. The incorrect protein-protein interaction in

STRING is the major factor that disturbed the construction of the novel p53 network models. The incorrect interactions may affect the global effect between nodes in the dependency matrix and lead model predictions to an unknown direction.

The first error, wrong gene name recognition, is the most common error we found in the extracted interaction records. The reasons causing this error are various: the same alias names can be used for different genes, the incorrect identification of chemical compound or drug name, etc. For example, STRING predicted that ELA2 activates p53 and showed the evidence highlighted by yellow colour (Laurora et al, 2005). However, ELA2 is ELANE (elastase, neutrophil expressed) in PubMed and has an alias name, HNE. However, HNE (4-Hydroxynonenal) in the evidence abstract was IL8 (interleukin 8).

The second error is wrong target recognition and a general term appears in complex context. This error occurred in the capture of target gene name during natural language processing. The search engine ignored some nouns which followed the gene name. For example, the noun, “target” was ignored. STRING predicted that RITA (ZNF331) activates p53. However, when we analyzed the evidence provided by STRING for further validation, the actual sentence was that RITA activates p53 targets (Zhao et al, 2010).

The third error is that a general term appears in a complex context, which prevented the text mining tools from parsing word accurately. Some genes were connected together by “-“, however, the text mining tools were not capable to indentify this connection. For example, STRING made a prediction that EP400 (p400) inhibits p53. In the evidence, it reported that the p53—p21 transcription was inhibited by p400 (Chan et al, 2005). STRING did not identify

the combined protein name”p53—p21” accurately and treated it as p53.

The fourth error is that the negation and negative words were not recognized by text mining. We found four examples of these negation and negative words: the first one is the negative word, “nor” was present in the sentence but the text mining tools did not recognize it. For example, STRING predicted that p53 activates HDAC5, however, the evidence reported that ”…nor was HDAC5 mRNA promoted by p53” (Huang et al, 2002). The second one is the use of “small interference RNA (siRNA)”. It is predicted that CDC20 activates p53 by STRING. However, the evidence provided by STRING indicated that small interference RNA (siRNA)-mediated silencing of p53 induced CDC20 (Kidokoro et al, 2008), which is opposite to the prediction that p53 activates CDC20.The third one is that the text mining tool ignored the removal of the gene. STRING predicted that TOPBP1 activates p53, but the evidence showed that the depletion of TOPBP1 up-regulated p53 target genes (Liu et al, 2009). The fourth one is that the word “the dominant negative form” was ignored by STRING. STRING predicted that MAPK9 inhibits PTGS2 but the evidence indicated that the dominant negative form of JNK1 (MAPK9) repressed Cox-2 (PTGS2) (Guan et al, 1998).

The fifth one is the wrong relationship identified for the interaction. STRING ignored the logical relationship between two genes which interact with a same gene together. For example, STRING predicted two protein-protein interactions that FOS (c-fos) inhibits ICAM1 and JUN (c-jun) inhibits ICAM1 individually. However, in the evidence, c-fos and c-jun attenuated the quercetin which had negative effect on ICAM1 (Ying et al, 2009). STRING treated this interaction as two individual interactions.

The sixth one is the speculation or question. The text mining tools did not parse the word “may”, which occurred in the extracted sentence. For example, it was predicted by STRING that POU4F1 (Brn-3a) inhibited BRCA1. However, the evidence only referred that Brn-3a may mediate the regulation of BRCA-1 (Budhram-Mahadeo et al, 2001).Therefore STRING reported speculation as a fact.

4.4 The PKT1377 and the PKT2275 model

In this chapter the details of the PKT1377 model and PKT2275 model construction are illustrated. Using a confidence score as the criteria, two models were constructed. If the confidence score of interaction record is more than 0.700 (the probability that the interaction was true was more than 70%), it was labelled as high confidence level; if the score is more than 0.400 and less than 0.700, it is at medium confidence level; and if the score is less than 0.150, it is at low confidence level. It was planned that all possible interactions which had direct or indirect relationship with p53 from the STRING database were included and the models were sorted in three layers the first layer included all interactions of genes or proteins that interact with p53; the second one is all interactions between those genes or proteins which interact with p53 directly; the third layer included all interactions which interacted with genes or proteins in the second layer. As a result, two models were constructed: PKT1377 and PKT 2275. The PKT1377 has 1377 nodes and 3612 interaction edges whose confidence score was more than 0.800, and the PKT2275 is the full extending model which included all interactions retrieved from STRING and has 2275 nodes and 13158 interaction edges. The population of different effect cells in the dependency matrix of the PKT1377 model and the PKT2275 model were

listed in Table 4.1.

Model Name	Total number of dependency effect cells	Number of No Effect cells	Number of Ambivalent Factor cells	Number of Weak Inhibitor cells	Number of Weak Activator cells	Number of Strong Inhibitor cells	Number of Strong Activator cells
PKT1377	1896129	1219766	669276	3171	3824	33	59
PKT2275	5175625	3273525	1892802	3820	5285	87	106

Table 4.1: Effect cell population result of dependency matrix for PKT1377 and PKT2275 model

This table lists number of different dependency cells found in the dependency matrix of the PKT1377 model and the PKT2275 model.

4.5 Time measurement of dependency matrix calculation

In this section, the problem caused by time cost of dependency matrix analysis and the solution of this problem is described. CellNetAnalyzer provided three different algorithms to calculate the shortest pathway for dependency matrix: exhaustive algorithm, approximative algorithm, and two-step algorithm described below. Since the complexity of these models was high, high performance computation power (the details of using high performance computation was illustrated in the method chapter) was used.

The questions that were drawn during the calculation are: why approximative algorithm was appropriate whereas the exhaustive and two-step algorithms

were not capable of calculating the dependency matrix? What prevented CellNetAnalyzer from calculating the dependency matrix using the other two algorithms? CellNetAnalyzer provided three different algorithms to calculate the shortest path distance between two nodes, which determined the effect type in the dependency matrix. Since these three algorithms had different strategies to calculate the shortest pathway distance between nodes, the PKT1377 model was analysed by all three algorithms to determine whether there was a difference between the dependency matrixes. However, during the process of dependency matrix calculation, it was observed that only the approximative algorithm had the capability to calculate the dependency matrix. The programme by the other two algorithms in CellNetAnalyzer stopped responding and the calculation failed. We suspected that the main reason is the PKT1377 model was too large to calculate all possible shortest paths between nodes. As a result, a test was designed for estimate the size of network model so that currently available computer power was capable to execute the calculation task. A consuming time test was processed with a different sub network model of PKT1377 to investigate the computer power cost of three different algorithms for dependency matrix calculation. Five sub network models were selected from the PKT1377 model with different number of nodes and edges (Table 4.2) and their dependency matrix were calculated by those three different algorithms.

Sub network name	Number of nodes	Number of edges	Time consumed by exhaustive depth-first traversal algorithm (minutes)	Time consumed by approximative algorithm (minutes)	Time consumed by two-step algorithm (minutes)
t500	500	739	3 minutes	1 minutes	1 minutes
t525	525	786	3 minutes	1 minutes	2 minutes
t550	550	828	4 minutes	1 minutes	2 minutes
t575	575	890	13 minutes	1 minutes	4 minutes
t600	600	955	99 minutes	1 minutes	12 minutes

Table 4.2: Statistic time cost of dependency matrix calculation.

The table lists time cost of dependency matrix calculation with different algorithm for sub network of the PKT1377 model.

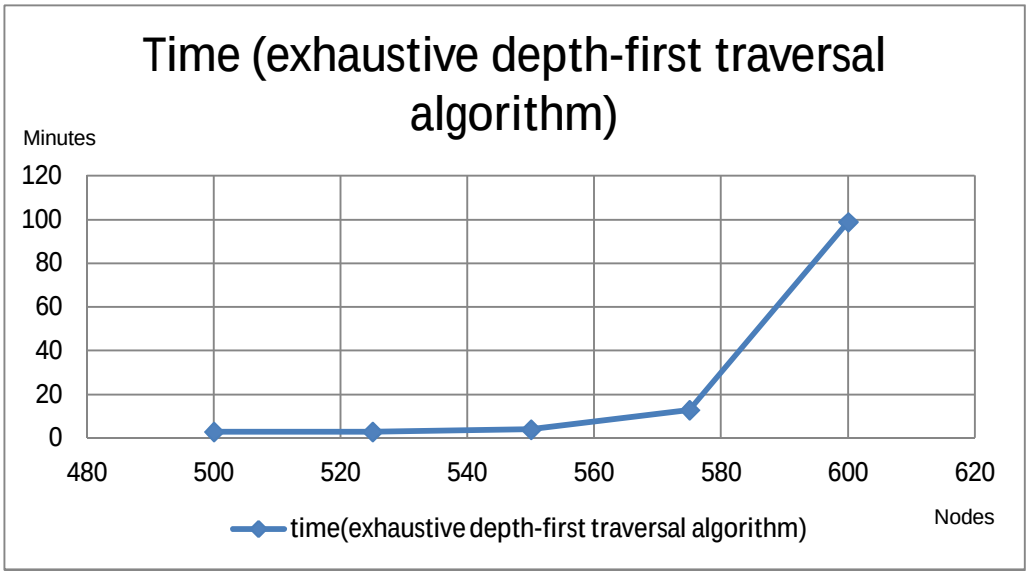
At first, the first 500 nodes were selected (we assumed that all nodes have the same priority and selected the first 500 nodes listed in the metabolite parameter file) out of the total 1377 nodes in the PKT1377 model by CellNetAnalyzer, kept their own interactions between those 500 nodes and calculated the dependency matrix by those three different algorithms. The consumed time records were listed in the Table 4.2. Next, the sub network was enlarged to the first 550 nodes in the PKT1377 model, added new interactions between them and recalculated the dependency matrix to obtain the time record for three different algorithms. In this way, the time necessary to calculate dependency matrix for the sub network model with 550 nodes, 575

nodes and 600 nodes was measured. Since the computer power could not afford the calculation of the model with more than 600 nodes and stopped responding, the measurement terminated at the scale of 600 nodes. Considering saving time for dependency matrix calculation, it was decided that the approximative algorithm should be used as the default algorithm to calculate the dependency matrix and the other two were utilized to investigate the result of dependency matrix in the *in silico* knock-out tests of PKT205/G1 model and PKT205 model in G2 version (described later), no difference was detected in the dependency matrix by those three algorithms.

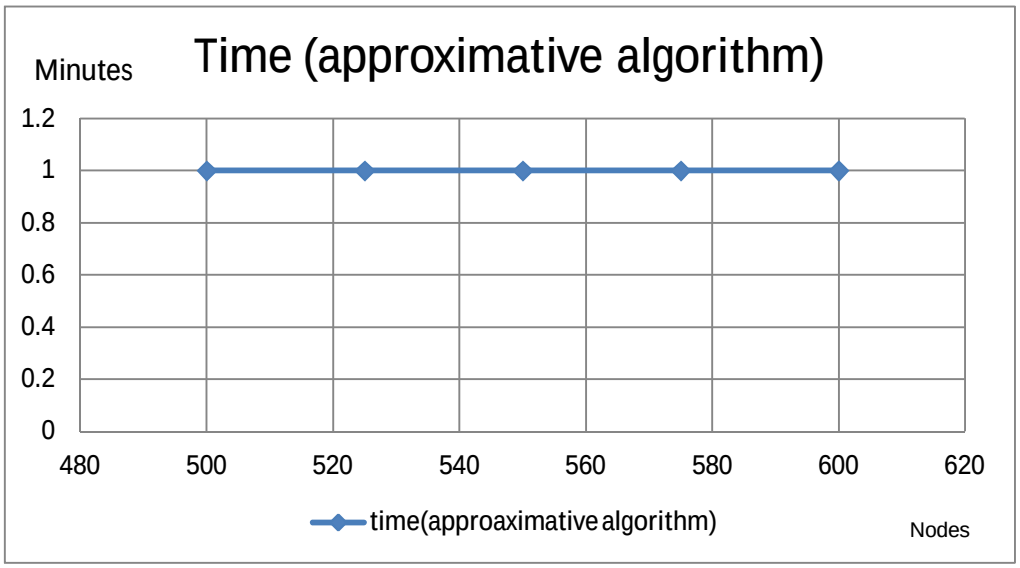
It was inferred that negative feedback loops may cause time-consuming problem for different algorithms. Our network models were represented as interaction graph in the CellNetAnalyzer and those interaction graphs were directed signed graphs, for example, “ATM activates p53” and the edge representing this activation was from ATM to p53. CellNetAnalyzer calculated the shortest positive or negative path from ATM to p53 and according to the shortest positive and negative path distance determined the global effect from ATM onto p53. As illustrated in the methods chapter, CellNetAnalyzer defined six types of effect elements in the dependency matrix. Those six types were determined by three main factors: whether there existed shortest pathway from ATM to p53, whether this shortest pathway was negative pathway or positive pathway, and whether these pathways touched negative feedback loops. CellNetAnalyzer provided three different algorithms to calculate the shortest distance of positive pathway or negative pathway between nodes. These three algorithms had different time costs due to their different mechanisms to calculate the shortest pathway distance. According to the time test shown in Figure 4.2, the computation time of the dependency matrix by the exhaustive algorithm increased exponentially with the size of the model.

The main reason is that an exhaustive search with the depth-first algorithm is required to access all possible pathways between each two nodes by enumeration. Klamt (2009) pointed that the rise in network size led to an exponential increase of the number of feedback loops (Klamt & von Kamp, 2009). As a result, the calculation of dependency matrix was time consuming due to the amount of feedback loops and cycles in the network.

A



B



C

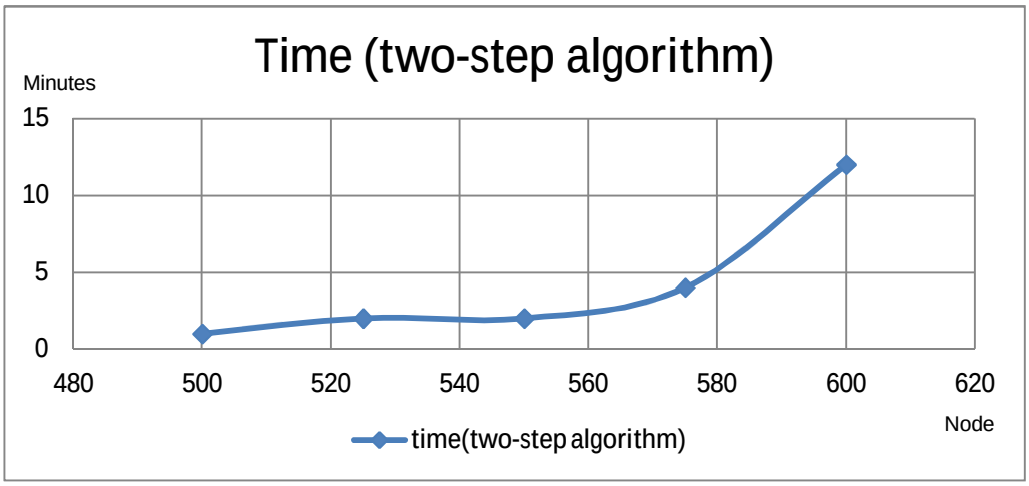


Figure 4.2: Duration of dependency matrix calculation consumed by three different algorithms for the p53 network in different nodes scale

Figure 4.2A shows the time consumed trend of the dependency matrix calculation by the exhaustive depth-first traversal algorithm. Figure 4.2B shows the time consumed trend of the dependency matrix calculation by the approximative algorithm. Figure 4.2C shows the time consumed trend of the dependency matrix calculation by the two-step algorithm.

However, the approximate algorithm utilizes a different strategy, double-label algorithm with cycle check (DLACC). DLACC algorithm is based on the Dijkstra's shortest path algorithm and the results were calculated in polynomial time. The effect of negative feedback loops or cycles decreased by the cycle check, but sometimes, the real shortest pathway with cycles may be missed. As a result, the approximate algorithms produce the results which were approximate or even equal to the real values. But this approximative algorithm was more suitable and less time-consuming for large network model which had hundreds of nodes and edges. This phenomenon was demonstrated by the time test for PKT1377 model (Figure4.2B)

The two-step algorithm is the combination of exhaustive search and DLACC. This algorithm has better performance than the exhaustive search with depth-first traversal algorithm in the responding time. However, the exhaustive search also affected the performance of shortest pathway calculation and the time cost of the calculation (Figure 4.2C).

4.6 Discussion

In summary, we discuss feasible solutions to avoid the errors caused by text mining and feasible improvements in future work. Currently, the most feasible solution to avoid the insertion of incorrect interaction information was by manual cross checking to validate these interactions by other reliable databases and publications. In the future, the improvement of the text mining results will require the improvements in text mining algorithms. Various efforts are being made to improve the accuracy and efficiency of text mining techniques for protein-protein interaction extraction (Zhou & He, 2008). For instance, Ananiadou et al (2010) insisted that database curation was needed so as to improve the efficiency and accuracy of the protein-protein interaction (PPI) extraction. They preferred deep parsing for the syntactic analysis in text mining, which considers potential relationships between entities recognized for protein-protein interaction prediction and not directly from the sentences (Ananiadou et al, 2010). Tsai (2012) utilized an integrated global association score to improve the accuracy of protein pair recognition. Not only the correlation score between proteins in each protein pair, but also the global correlations of the targeted protein with other proteins extracted from the whole article were considered in the global association scores. This method revealed its powerful role for the curation of protein-protein interactions (Tsai,

2012). Krallinger et al (2012) designed a text mining approach by associating ontology data such as the Gene Ontology (GO) and the molecular interaction ontology (PSI-MI), with protein-protein interactions and utilized an evaluation system named as BioCreative initiative (Critical Assessment of Information Extraction Systems in Biology) to test this approach. The precision of the outcome was quite high but the performance of this approach was limited by the specificity of ontology terms to match the interaction activity extracted from literature (Krallinger et al, 2012).

Algorithms for text mining will be improved resulting in better performance in the syntactic analysis and semantic analysis of large texts in publications in the next ten years. Zhou et al (2008) illustrated three major methodologies utilized for protein-protein interaction extraction work: the computational linguistics-based approach, the rule-based approach and the machine-learning approaches and compared their performance by each approach (Zhou & He, 2008). The possible solution illustrated by Zhou et al (2008) was to overcome the limitation of text mining by improving the identification of ontology terms and terminological lexicons identification, making the entity names of gene or proteins not changed frequently, establishing a confidence score scheme, advancing the validation rules for better evaluation and so on (Zhou & He, 2008). Currently the only possible approach applicable to this investigation to optimize the accuracy of protein-protein interaction results is to accept all interaction records having the high confidence score provided by STRING, validating each of them manually and regularly checking for changes with updates of STRING.

Chapter 5 Final version of the PKT205 model and analysis results

5.1 Introduction

As mentioned in the previous chapter, we have built a model of the p53 interactome but encountered two problems. First, simulations could not be performed because of the large size of the model that required excessive computation time. Second, we detected errors caused by text mining that needed to be excluded from the model. Therefore, in order to overcome these difficulties we decreased the size of the model to create a new model, the PKT205/G1 model and manually curated all records found in the PKT1377 model.

5.2 Creation of the PKT205/G1 (manual curation of protein-protein interaction records)

In order to address the time consuming problem encountered during dependency matrix calculations and simulations described above, the following research strategy was devised. Construction of the G1 version of the PKT205 model started by confirming manually all interactions using publications from PubMed so as to avoid the inclusion of incorrect interactions that could affect the model. According to run-time measurements, it was found that the time needed for calculating the dependency matrix increased sharply with the increase in network size when the shortest paths were calculated by the exhaustive depth-first traversal algorithm or two-step algorithm. With the

purpose of avoiding the inclusion of incorrect protein-protein interaction records extracted by STRING and the limitation of computer power for dependency matrix calculation, new constraints were defined to decrease the scale of the model. For this purpose incorrect interactions were not included and only two interaction layers were used including direct interactions with p53 and interactions between those genes that interact with p53 such as MDM2 and MDM4, which were p53 gene targets and were connected by a two step negative feedback loop, in which MDM4 activates MDM2 and MDM2 inhibits MDM4. The new model included p53 node and 201 other nodes which represents genes or proteins to interact with p53 at the high confidence score (more than 0.700) and all these interactions with p53 were confirmed manually by literature search. Then interactions between these 201 nodes at high confidence score were extracted from STRING, and included into the PKT205/G1 model. This PKT205/G1 model has 202 genes and 535 interactions. All nodes in Additional Table 4 represent genes or proteins which interact with p53 and obtained from STRING database. All interactions in Additional Table 4 were manually curated by literature search in PubMed or other on-line text mining tools. Those interactions were mainly from two main resources: the interaction information retrieved automatically from the STRING database and the interactions from the PKT38 model. They were combined together to create of the PKT205/G1 model whose interactions are described in Additional Table 4 below. The new model included 201 nodes which interact with p53 at the high confidence score (more than 0.700) and all these interactions with p53 were confirmed manually by literature search (Figure 5.1). Then interactions between these 201 nodes at high confidence score were extracted from STRING, and included into the PKT205/G1. This PKT205/G1 has 202 nodes and 535 interactions. Those interactions were mainly from two main resources: the interaction information retrieved

automatically from the STRING database and the interactions from the PKT38 model. They were combined together to create of the PKT205/G1.

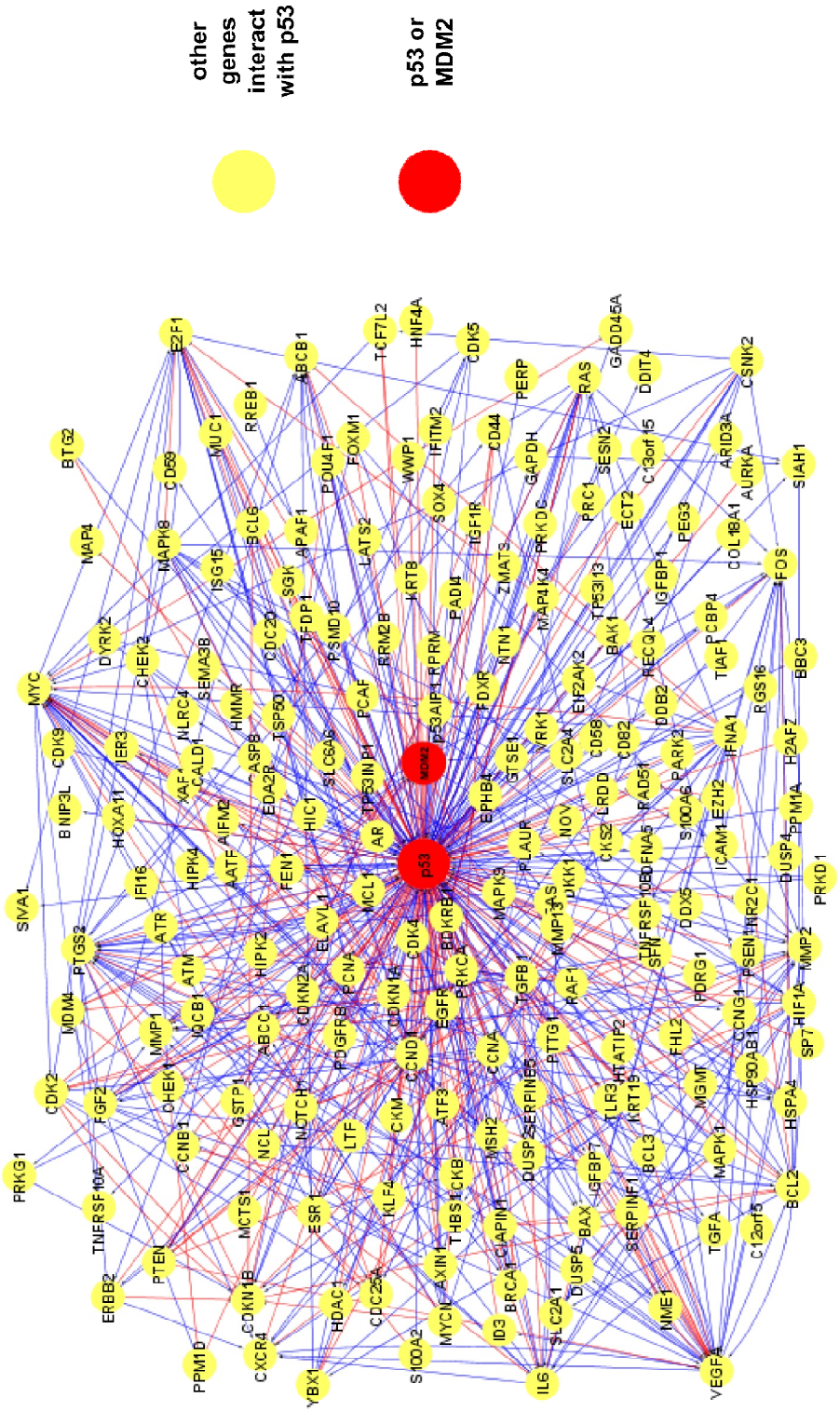


Figure 5.1: PKT205/G1 network map

202 nodes and interactions between them are shown on the map. P53 and MDM2 are represented in red colour. All activations are represented in blue colour and all inhibitions were represented in red colour.

5.3 *In silico* simulation of PKT 205 model in G1 version

After constructing the G1 version model in the previous section, we performed *in silico* knock-out tests in order to analyze the PKT205/G1 model. We performed *in silico* knock-out tests for selected genes to explore their role in the network by the comparison of dependency matrix before and after deletion. This comparison was achieved by a customs designed Java-based programme (Figure 2.5). According to the network analysis results presented in the previous chapter, it was found that individual genes or proteins were affected differently in the PKT205/G1 model depending on their connectivity degree and whether they were involved into negative feedback loops. Table 5.1 shows the connectivity degree of these 202 nodes. Table 5.2 and Figure 5.2 show the distribution of nodes with different connectivity degree in the PKT205/G1 model (Figure 5.2). It was found that only p53 participated in more than 100 interactions, which was the most connected node in the PKT205/G1model, 25 nodes took part in more than 10 interactions and the majority of nodes were connected with less than 10 interactions. According to the results of knock-out tests based on the PKT38 model, the effect of node deletion was related to its connectivity degree. We performed 30 knock-out tests for single node depletions: p53, 25 nodes with type 2 connectivity degree in Table 5.1, and four nodes with type 3 connectivity degree: ATM, BRCA1,PTEN, and CDKN1A.

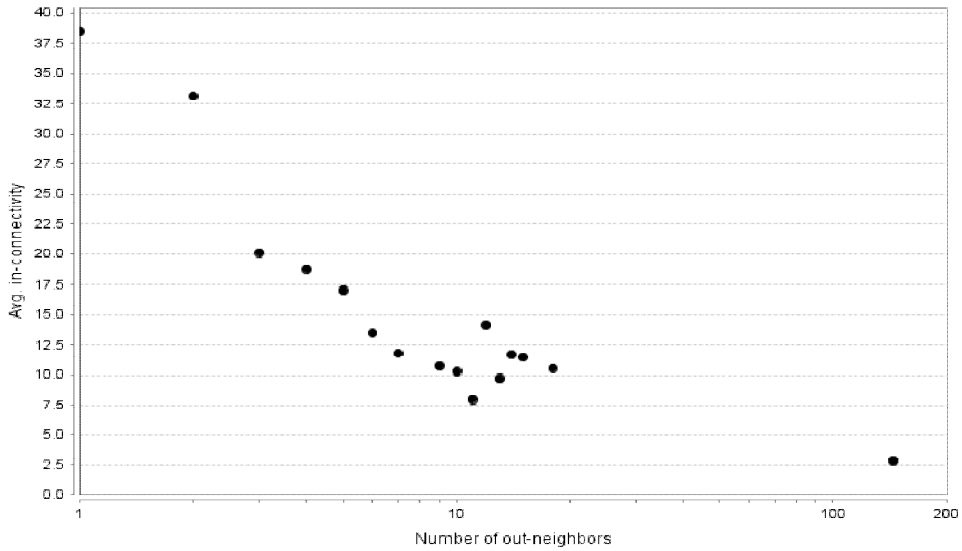


Figure 5.2: Connectivity degree distribution in the PKT205/G1 model

Both axes of the figure are in logarithmic scale and the scatter plot indicates the connectivity degree distribution.

From the statistical analysis in Table 5.1 and Table 5.2, it was found that nodes with higher connectivity degree such as p53 lead to more substantial changes in the effect cells of the dependency matrix when compared with the dependency matrix of T0, which represented the p53 wild type. Comparing the distribution of dependency cell changes in those 31 knock-out tests above with different gene deletion, we found six types of distributions (Table 5.2)

Type number	Connectivity degree range	Population of nodes of interacted type	Percentage of nodes out of total nodes
1	Connectivity >100	1	0.49%
2	10≤Connectivity ≤100	25	12.38%
3	0<Connectivity <10	176	87.13%

Table 5.1: Distribution of nodes with different connectivity degree for the G1 model version

Column 1 shows the number of types. Column 2 lists the connectivity degree scale. Column 3 shows the population of nodes with certain connectivity degree. Column 4 shows the percentage out of total 202 nodes.

Model Name	Gene deletion	Total number of effect cells	Number of No Effect cells	Number of Ambivalent Factor cells	Number of Weak Inhibitor cells	Number of Weak Activator cells	Number of Strong Inhibitor cells	Number of Strong Activator cells
T0	Null	40804	23511	14958	1109	1213	4	9
T1	P53	40401	34621	5623	53	75	9	20
T2	TGFB1	40401	23693	14245	1176	1274	4	9
T3	MDM2	40401	23375	14313	1296	1404	4	9
T4	MMP2	40401	23219	14850	1109	1213	4	6
T5	CCND1	40401	23683	14021	1262	1412	4	9
T6	CXCR4	40401	23534	14511	1117	1226	4	9
T7	IL6	40401	23375	14664	1120	1229	4	9
T8	FGF2	40401	23375	14717	1094	1202	4	9
T9	ABCB1	40401	23216	14850	1109	1213	4	9
T10	PRKCA	40401	23480	14582	1111	1213	4	11
T11	FOS	40401	23534	14445	1141	1268	4	9
T12	HIF1A	40401	23797	14274	1109	1207	4	10
T13	BCL2	40401	23375	14459	1247	1307	4	9
T14	MMP1	40401	23218	14850	1109	1213	3	8
T15	PTGS2	40401	23534	14354	1174	1326	4	9
T16	PTEN	40401	23375	14465	1124	1224	4	9
T17	VEGFA	40401	24211	13387	1323	1461	6	13

T18	MYC	40401	23534	14450	1145	1259	4	9
T19	E2F1	40401	23587	14322	1180	1297	4	11
T20	IFNA1	40401	23269	14798	1109	1213	4	8
T21	RAS	40401	23375	14639	1135	1239	4	9
T22	MAPK8	40401	23268	14798	1109	1213	4	9
T23	CSNK2	40401	23270	14798	1109	1213	4	7
T24	ATM	40401	23269	14798	1109	1213	4	8
T25	BRCA1	40401	23375	14690	1109	1214	4	9
T26	CDKN1A	40401	23375	14691	1108	1214	4	9
T27	IGF1R	40401	23375	14717	1094	1202	4	9
T28	EGFR	40401	23693	14375	1108	1212	4	9
T29	CCNA	40401	23375	14661	1123	1229	4	9
T30	CDK2	40401	23534	14530	1108	1216	4	9
T31	CDKN1B	40401	23375	14688	1111	1214	4	9

Table 5.2: Results of in silico knock-out tests for PKT205/G1.

Column 1 lists the model name. Column 2 shows the name of genes deleted from the model. Column 3 shows the total number of effect cells in the dependency matrix. Column 4 lists the number of no effect cells in the dependency matrix. Column 5 lists the number of ambivalent factor cells in the dependency matrix. Column 6 lists the number of weak inhibitor cells in the dependency matrix. Column 7 lists the number of weak activator cells in the dependency matrix. Column 8 lists the number of strong inhibitor cells in the dependency matrix. Column 9 lists the number of strong activator cells in the dependency matrix. The value ”Null” indicates that no nodes was removed from the PKT205/G1 model.

The first type of distribution is that ambivalent factors in the dependency matrix of p53 wild type were changed to the other five types of effects in the dependency matrix of the deletion model. Moreover, there were changes from

weak activator to no effect, from weak inhibitor to no effect, from weak inhibitor to strong inhibitor and from weak activator to strong activator, for example, when the p53 wild type model was compared with p53 mutant model (Figure 5.3).

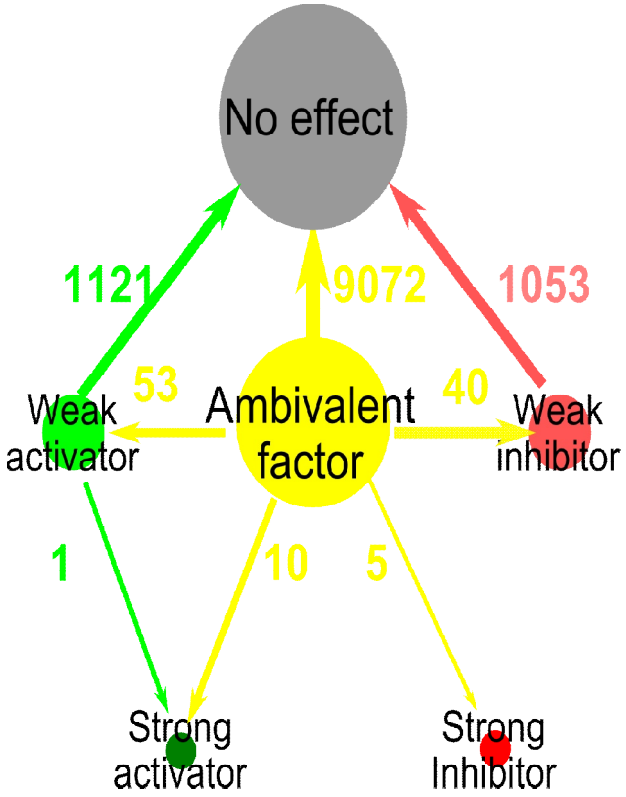


Figure 5.3: Distribution of changes in the dependency matrix of the p53 *in silico* knock-out compared to the wild-type PKT205/G1model

The gray circle represents no effect elements, the yellow cycle represents ambivalent factors, the turquoise green circle represents weak activators, the pink circle represent weak inhibitors, the red circle represents strong inhibitors, and the dark green circle represents strong activators; the direction of the arrow represents the direction of changes in the knock-out.

The major changes observed in the p53 knock-out test revealed genes whose activity changes strongly in the mutant p53 cells. Those predictions based on major changes motivated us to identify genes that are potential target for

cancer therapy. For example, model predicted that the CHEK1 activity was enhanced by ATM in the p53 negative cells. CHEK1 plays an important role in the control of the cell cycle arrest (Macip et al, 2006). As a result, certain drugs that alter the activity of CHEK1 in p53 mutant cells may be used to prevent tumour cell growth.

The second type of distribution change is that ambivalent factors were changed to the other five types of effects. There were changes from weak activator to no effect, from weak inhibitor to no effect, and from weak inhibitor to strong inhibitor. For example, these changes were found when the wild type model was compared to the model that had VEGFA removed (Figure 5.4).

The depletion of VEGFA also caused perturbations in the PKT205/G1 model. For example, it was predicted that the expression of BAX was induced by activated FOXM1 when VEGFA was absent. This prediction may provide a potential way to foster accumulation of the pro-apoptotic gene BAX to promote tumour cell death.

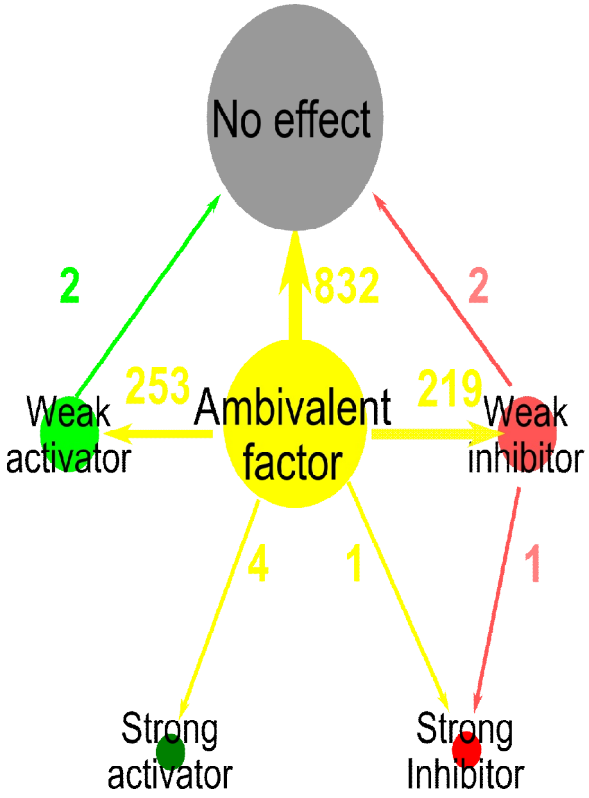


Figure 5.4: Distribution of changes in the dependency matrix of the VEGFA *in silico* knock-out compared to the wild-type for the PKT205/G1 model.

A node, VEGFA was removed from the PKT205 model. Colours symbolize the same effects as in the Figure 5.3.

The third type of change is that ambivalent factors were changed to no effect, weak inhibitor and weak activator. There were also changes from weak inhibitor to no effect. The *in silico* knock-out test of CXCR4 (Figure 5.5) is an example of this type.

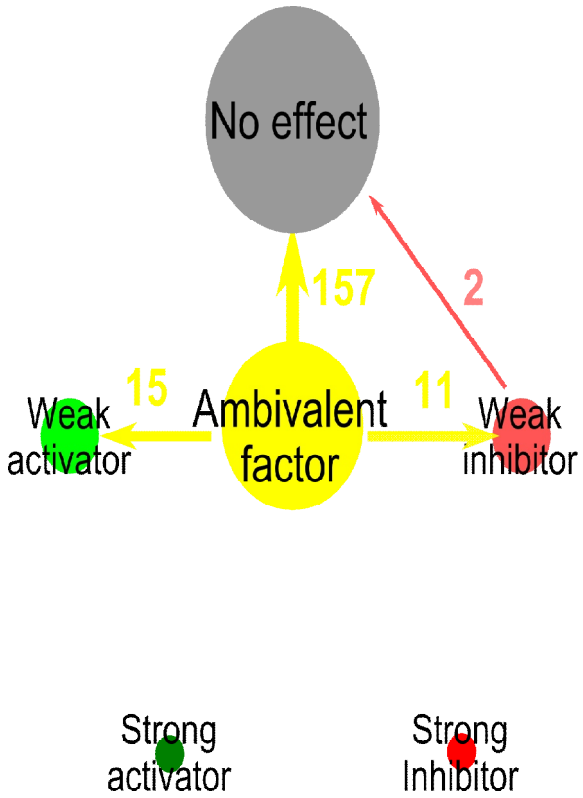


Figure 5.5: Distribution of changes in the dependency matrix of the CXCR4 *in silico* knock-out compared to the wild-type for the PKT205/G1

CXCR4 was removed from the PKT202 model. Colours symbolize the same effects as in the Figure 5.3.

The depletion of CXCR4 was predicted to result in diverse changes in p53 pathways. For example, activated p53 promoted the expression of TNFRSF10B (DR5) and DR5 was reported to induce apoptosis in p53 wild type cells (Wu et al, 1999). So this depletion may provide a direction to stimulate tumour cell death.

The fourth type of change is that ambivalent factors were changed to the other four types of effects excluding strong inhibitor. The *in silico* knock-out test of E2F1 (Figure 5.6) is an example of this type.

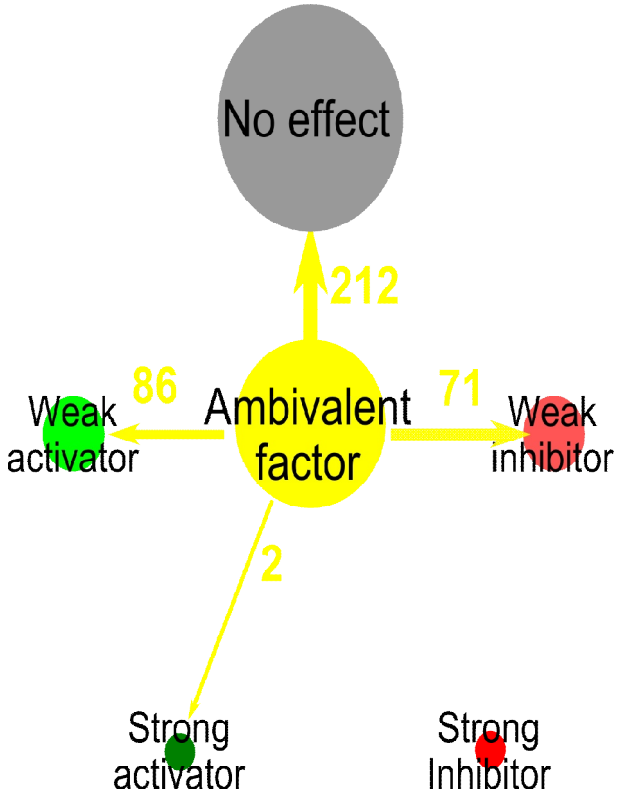


Figure 5.6: Distribution of changes in the dependency matrix of the E2F1 *in silico* knock-out compared to the wild-type for the PKT205/G1 model

E2F1 was removed from the PKT202 model. Colours symbolize the same effects as in the Figure 5.3.

It was predicted from the E2F1 knock-out test that activated ATM will strongly enhance CHEK2 in the absence of E2F1. So depletion of E2F1 may help p53 to be stabilized and function as a tumour suppressor protein.

The fifth type of distribution is that ambivalent factors were only changed to weak activator and weak inhibitor. The *in silico* knock-out test of MDM2 below (Figure 5.7) is an example of this type.

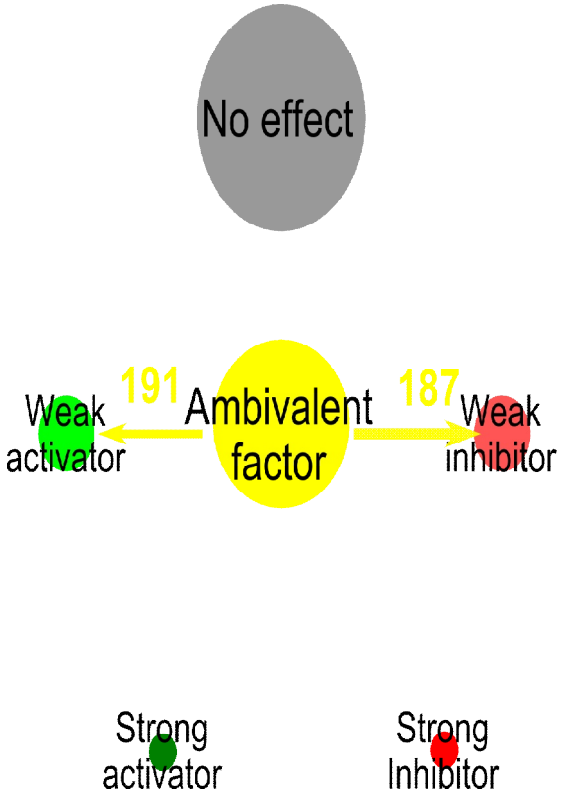


Figure 5.7: Distribution of changes in the dependency matrix of the MDM2 *in silico* knock-out compared to the wild-type for the PKT205/G1.

MDM2 was removed from the PKT202 model. Colours symbolize the same effects as in the Figure 5.3.

As there was a negative feedback loop between MDM2 and p53, the absence of MDM2 results in the instability of p53 network. The absence of MDM2 caused similar levels of both positive and negative perturbations in the network. So removing MDM2 may be not an ideal target for cancer treatment.

The final type of distribution is that there was no change found when the dependency matrix in the knock-out test was compared to the dependency matrix in the p53 wild type. For example, the effect elements in the dependency matrix of ATM depletion remained the same as in p53 wild type. The main reason is that ATM has a low connectivity degree (less than 10) in

the PKT205/G1 model. According to the distribution of changes, we can conclude that the depletion of nodes with higher connectivity degree affects more substantially the dependency relationship between remaining nodes. We also observed that most deletions result in a transformation of ambivalent factors into other types of interactions. This can be explained by the reduced number of feedback loops in deletion networks, which disrupts the wild-type balance and leads to more direct effects. These predictions demonstrated that the p53 network model like PKT205/G1 could provide predictions for dynamic mechanisms of p53 pathways. We could utilize the PKT205/G1 model to predict internal p53 pathway behaviour under perturbations by carrying out *in silico* knock-out tests. The major changes in the dependency matrix could further our understanding of gene activities in response to mutations, such as the absence of p53 or other components of the network, and facilitates the exploration of potential targets for cancer treatment.

5.4 Logical steady state analysis of the PKT205/G1

After the *in silico* knock-out tests, logical steady state analysis was performed to determine regulatory mechanisms governing the stability of the PKT205/G1 interactome. In order to investigate the details of p53 pathway mechanisms, the 202 nodes were classified into three layers: upstream of p53, p53 itself, and downstream of p53. It was found that 65 nodes functioned as upstream of p53 in the PKT205/G1 model, and 144 nodes functioned as downstream of p53. Because of feedback loops present in the system, 8 nodes functioned as both upstream and downstream of p53. Since there were many nodes that functioned upstream of p53, their role was tested by Logical Steady State (LSS) Analysis.

An input signal was added into the PKT205/G1 model to explore the role of these 65 upstream nodes on the whole network by logical steady state analysis. This input signal was connected to these nodes individually. For each node, two scenarios were predefined using a different value of the input signal. For example, ATM is an upstream node of p53. The input signal DNA damage was connected to ATM by activation (see the next chapter). The state of this input signal was preset to “ON” or “OFF”, respectively, and then the state of ATM was defined. Then logical steady state analysis was processed and states of the other 201 nodes were calculated by CellNetAnalyzer. The results for the logical steady state analysis of these 65 upstream nodes are listed in Table 5.3 below.

Name of node A	**Population of nodes with determined state in Scenario 1: node A is preset to ON**	**Population of nodes with determined state in Scenario 2: node A is preset to OFF**
HOXA11	1	155
NTN1	1	155
PSMD10	1	155
TGFB1	154	151
AATF	156	2
LTF	154	154
BCL6	155	155
SOX4	155	1
SERPINF1	154	154
MDM2	154	154

DDX5	155	1
ELAVL1	155	155
IGF1R	154	1
ERBB2	154	154
MYCN	154	154
HSPA4	154	154
ZMAT3	155	1
NCL	155	155
KLF4	156	156
PPM1A	155	1
PLAUR	1	154
HTATIP2	155	155
PTTG1	157	155
E2F1	154	152
TIAF1	155	1
CIAPIN1	1	155
BRCA1	154	1
PARK2	1	155
BTG2	155	155
SGK	1	154
MUC1	155	155
MCTS1	155	155
FAS	154	1
YBX1	154	1
HDAC1	1	155
ID3	1	154
PADI4	1	155
POU4F1	155	1

MYC	154	154
RREB1	155	1
IFNA1	156	158
CHEK2	154	2
AXIN1	154	155
H2AFZ	1	155
MAPK1	154	154
VRK1	155	1
RAF1	154	1
CSNK2	158	158
HIPK2	156	156
CDK2	154	154
CHEK1	155	155
HIPK4	155	1
CDK5	155	1
PPM1D	155	156
AURKA	1	155
PRKD1	155	1
DYRK2	155	1
MAPK9	155	1
CDK9	155	1
MAPK8	156	156
EIF2AK2	155	1
ATM	157	2
ATR	156	1
PRKDC(DNA-PK)	155	1
MDM4	154	154

Table 5.3: List of node population with determined states for the simulation of G1 version of the model

Column 1 lists the name of nodes. Column 2 shows the population of nodes with determined state when the node in column 1 is ON. Column 4 shows the population of nodes with determined state when the node in column 1 is OFF.

According to the result of logical steady state analysis (Table 5.3), these 65 nodes were classified into 3 different types:

The first type of result had more than 100 determined states when the input signal or state of the node was either OFF or ON and only 1 or 2 determined state once the input signal switched to the opposite state. For example, this was the case for ATM and ATR. This type of result applied to 36 out of total 65 upstream nodes.

The second type had the same population of determined states in both scenarios. There were 23 nodes of this type, for example, MDM2.

The third type had more than 100 determined states in both scenarios and the number of nodes with determined states was different from previous two types. There were only 6 nodes in this type: TGFB1, BCL6, PTTG1, E2F1, AXIN1 and IFNA1.

The factors leading to this phenomenon are complex: the connectivity of nodes in the PKT205/G1 model, the position of the node in the whole network, whether this node has ambivalent interactions with other nodes, and the relationship between the node and negative feedback loops. As a result, we decided to investigate by literature search the link between these 65 nodes

and an external input signal, DNA damage, which were added into the PKT205/G1 model. DNA damage was used as its effects of p53 interactome are well studied and it is the most relevant environmental signal for the function of p53, development and treatment of cancer.

5.5 Construction of the PKT205/G2 model

In the previous section, we found that the G1 version of the model did not offer the opportunity to explore the relationship between the model and external signals. As a result, we constructed the PKT205/G2 model. Considering the effect from stress signals and the cellular response, we added an input node, DNA damage, and an output node, apoptosis to the network model. All 65 upstream nodes were analysed manually by Gene Ontology (GO) terms to confirm the links from DNA damage; the same was done for downstream nodes and apoptosis. Apoptosis was chosen as it is one of the main effects of p53 activation in response to excessive DNA damage and is the most relevant for clinical use of chemotherapeutic compounds. All GO terms relevant to those effects were analysed manually using PubMed documents and other online text mining tools (Additional Table 5 and 6). Since the negative feedback loop, p53 - MAPK14 (p38) - PPM1D played an important role in the PKT38 model, we added a new node MAPK14 and all confirmed interactions with it to finalize the network, PKT205/G2. There were 205 nodes and 673 interactions in the PKT205/G2 model (Figure 5.8).

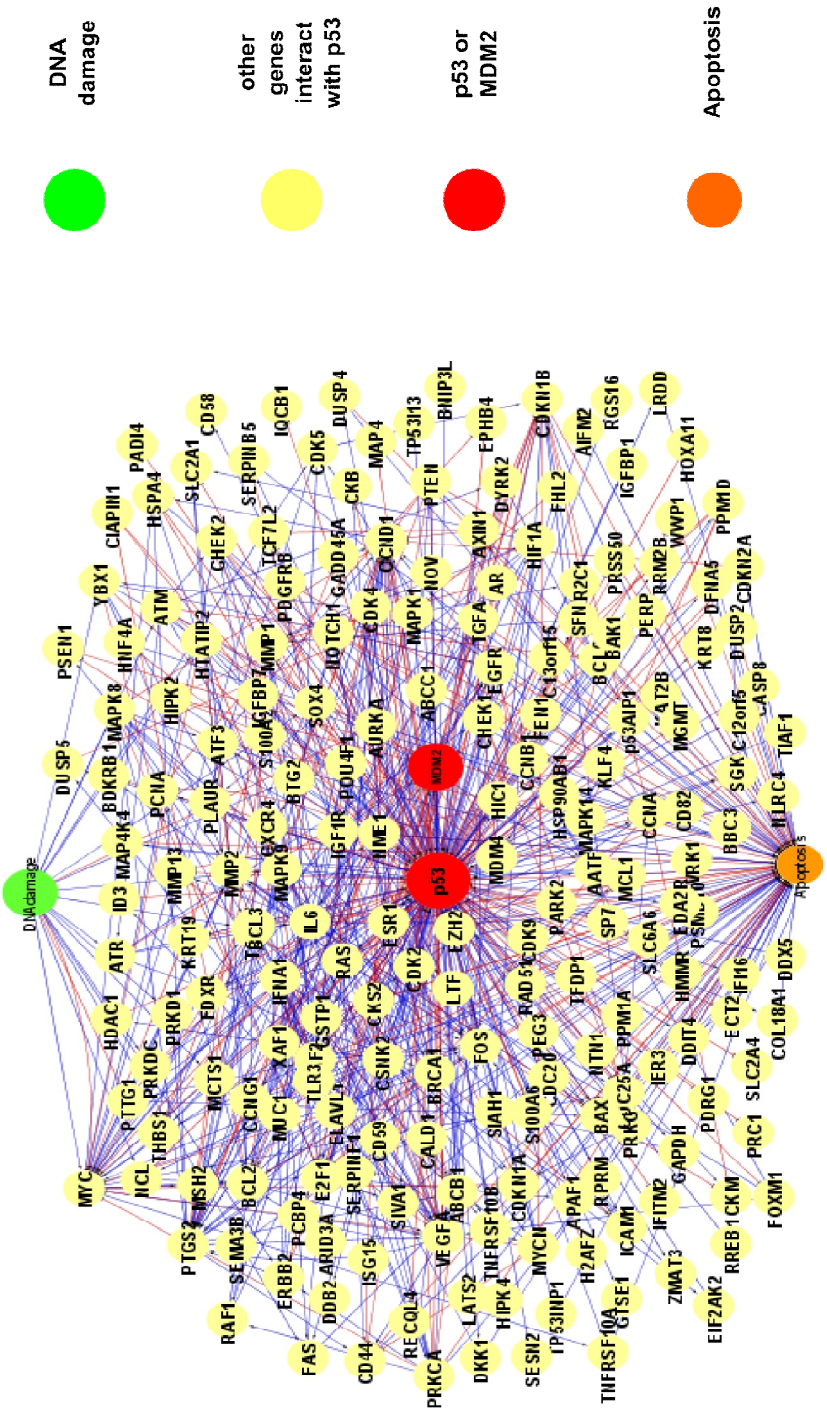


Figure 5.8: Network map of the PKT205/G2 model.

Input node, DNA damage was represented by turquoise green colour, and output node, apoptosis was represented by orange colour. The p53 and MDM2 were marked by red colour and the other nodes which interact with p53 were represented by yellow colour.

5.6 *In silico* knock-outs of the PKT205/G2 model

In this section we describe the details of the process of *in silico* knock-out tests in PKT205/G2 (Table 5.4). Since there were many different effect changes found in the *in silico* knock-out tests of the PKT202 model, we mainly focused on the *in silico* knock-out tests of p53, MDM2, MDM4, ATM, ATR, and BBC3 (PUMA). Moreover, dual gene depletion was considered for BBC3 and CDKN1A (p21). However, it was found that the depletion of CDKN1A did not have a strong effect on the BBC3 negative model. The main reason is that CDKN1A has a low connectivity degree in the PKT205/G2 and was downstream of p53.

There were four types of effect change distribution found in knock-out tests described in Table 5.4. The first one is the most complex, in which effects changed from ambivalent factor to all other types of effects, weak inhibitor changed to no effect or strong inhibitor, and weak activator changed to no effect or strong activator. The p53 *in silico* knock-out test was an example of this type (Figure 5.9).

As p53 is the most connected node in the PKT205/G2 model, its depletion results in the instability of the whole network. This was described in section 5.3 and the simulation results above verified it. For example, it was predicted that the activated ATM enhanced the expression of CHEK1 strongly in response to DNA damage in the p53 mutant model. This prediction was obtained from the major change that ATM become a strong activator of CHEK1 in the p53 null model and it was validated by experimental data in next chapter.

Gene deletion	Total number of dependency effect cells	Number of No Effect cells	Number of Ambivalent Factor cells	Number of Weak Inhibitor cells	Number of Weak Activator cells	Number of Strong Inhibitor cells	Number of Strong Activator cells
Null	42025	23468	16294	1051	1141	20	51
P53	41616	35036	6364	46	70	31	69
MDM2	41616	23443	15797	1119	1184	20	53
MDM4	41616	23336	16016	1052	1141	20	51
ATM	41616	23231	16126	1051	1141	18	49
ATR	41616	23229	16126	1051	1141	20	49
BBC3	41616	23170	16209	1040	1127	20	50
BBC3 and CDKN1A	41209	23039	15933	1039	1128	20	50

Table 5.4: Statistical result of in silico knock-out tests for PKT205/G2

Column 1 shows the name of genes deleted from the model. Column 2 shows the total number of effect cells in the dependency matrix. Column 3 lists the number of no effect cells in the dependency matrix. Column 4 lists the number of ambivalent factor cells in the dependency matrix. Column 5 lists the number of weak inhibitor cells in the dependency matrix. Column 6 lists the number of weak activator cells in the dependency matrix. Column 7 lists the number of strong inhibitor cells in the dependency matrix. Column 8 lists the number of strong activator cells in the dependency matrix. ”Null” means no gene was removed from the PKT205/G2 model.

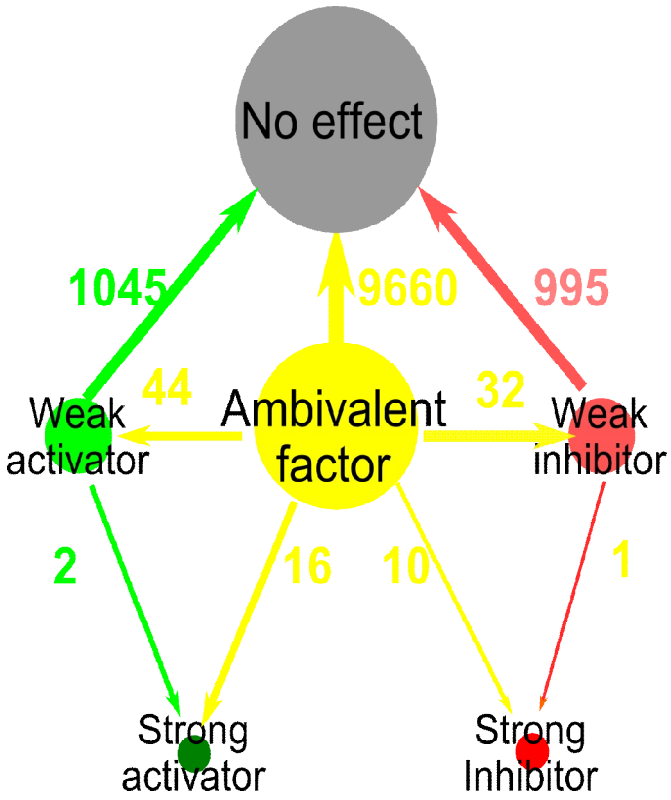


Figure 5.9: Distribution of changes in the dependency matrix of the p53 *in silico* knock-out compared to the wild-type for the PKT205/G2

P53 was removed from the PKT205/G2 model. Colours symbolize the same effects as in the Figure 5.3.

The second type of change distribution was that ambivalent factors only changed to no effect, weak inhibitor, weak activator and strong activator; the MDM2 knock-out test was an example of this type (Figure 5.10). With an input node and output node added in the PKT205/G2 model, the absence of MDM2 caused more changes in the PKT205/G2 model than in the PKT205/G1 model. But the number of major changes was less than the p53 knock-out test.

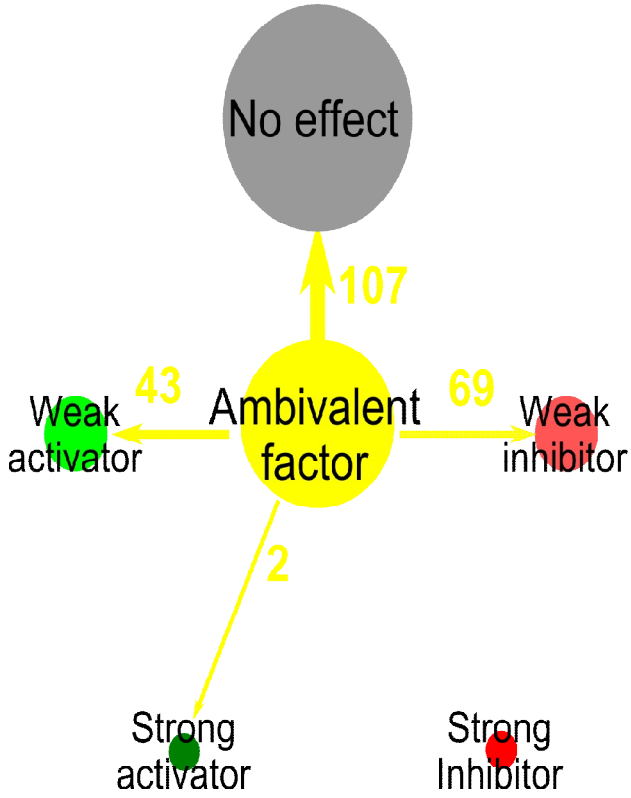


Figure 5.10: Distribution of changes in the dependency matrix of the MDM2 *in silico* knock-out compared to the wild-type for the PKT205/G2

MDM2 was removed from the PKT205/G2 model. Colours symbolize the same effects as in the Figure 5.3.

The third type of change distribution was found in the ATM knockout, whose depletion caused a single change, displayed in the Figure 5.11. The MDM4 knock-out test and ATM knout-out test had this change distribution. The final type of change distribution was that no change was found when wild type and mutant networks were compared, such as the ATR and BBC3 knock-out tests. Since ATM is poorly connected in the PKT205/G2 model, the depletion of ATM was found to only result in the abolishment of BCL6 down regulation by DNA damage.

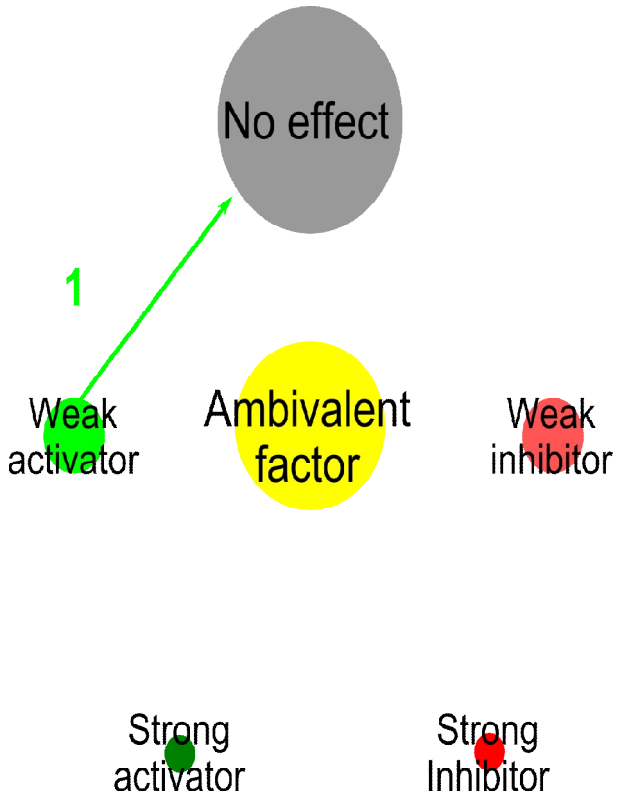


Figure 5.11: Distribution of changes in the dependency matrix of the ATM *in silico* knock-out compared to the wild-type for the PKT205/G2

ATM was removed from the PKT205/G2 model. Colours symbolize the same effects as in the Figure 5.3.

5.7 Logical steady state analysis of the PKT205/G2

In order to determine how node states changed upon perturbations caused by environmental stress signals and learn more about factors contributing to the stability of the system, a logical steady state analysis for this model was performed (Table 5.5). Logical steady state analysis simulations were performed to further the understanding of the p53 function in the PKT205/G2 model. It was found that the system in the p53-null model was less stable than in the p53 wild type model according to the decreasing number of nodes with

determined states in the p53–null model (Table 5.5). Since ATM functioned as an upstream node of p53 and was induced by DNA damage in PKT205/G2, simulations were also processed in the ATM null model. The comparison between p53 negative and ATM negative scenarios indicated that the population of nodes with determined states was linked to the connectivity of the node removed from the PKT205/G2 model. The p53 node has much higher connectivity degree than the ATM node, and as a result, the depletion of p53 led to the decreased stability of the whole system according to the number of nodes that changed from determined states in Table 5.5. The second finding was that in the simulation of the p53 knock-out tests, the state of some genes remained the same as in the p53 wild type model. Here we can draw a conclusion that the states of those genes were independent of the p53 state. The third observation was relevant to changes in the state of MDM4 that depends on p53 status. MDM4 had determined states in the simulation of the p53 wild type model. However, when p53 was removed, the MDM4 state changed to undetermined when DNA damage was ON. Therefore our prediction is that in cells exposed to DNA damage the stable state of MDM4 depends on the state of p53.

Scenario name	Input signal	Model type	Population of determined states	Percentage of determined nodes
Scenario 1	DNA damage ON	P53 wild-type	180	87.8%
Scenario 2	DNA damage OFF	P53 wild-type	181	88.3%
Scenario 3	DNA damage ON	P53 knock-out	92	45.1%
Scenario 4	DNA damage OFF	P53 knock-out	92	45.1%
Scenario 3	DNA damage ON	ATM knock-out	178	87.3%
Scenario 4	DNA damage OFF	ATM knock-out	179	87.7%

Table 5.5: List of scenarios in the logical steady state analysis for the G2 version model

Four scenarios of logical steady state analysis with different input signals are defined with their input signal, model type and percentage of nodes having a determined state.

5.8 Construction of the finalized PKT205 model (PKT205/G3)

We made further improvements to the PKT205/G2 model following an update of the STRING database. As a consequence of this update, we re-evaluated all protein-protein interactions manually and corrected some of them. The final list of interactions is shown in Additional Table 7.

We identified 202 genes or proteins that interacted with p53 in the final PKT205/G3 model and this version of the PKT205 model was used for analysis and validation in the next chapters. With the connection of DNA damage input node and apoptosis output node; this PKT205 model included 205 nodes and 677 interactions (Figure 5.12). As described in previous chapters, we merged some genes into single nodes when they could not be distinguished according to literature evidence: the genes HRAS, KRAS, NRAS and RASD1 were regarded as a single node; RAS. CCNA1 and CCNA2 were combined into a single node, CCNA; CSNK2A1 and CSNK2A2 were combined into a single node, CSKN2. The finalized PKT205 model inherited the confirmed links to DNA damage and apoptosis from the G2 version. There were 20 upstream nodes of p53 linked to DNA damage node directly. 17 out of them were stimulated by DNA damage and the other 3 were inhibited by this input node. We found 30 two-step negative feedback loops in the PKT205 model and p53 participated in 14 of them. Since the negative feedback loop between p53 and MDM2 was located in the centre of the whole network, they were marked by red in Figure 5.12. There were 77 genes in the PKT205/G3 model linked to apoptosis with 95 interactions. 56 out of these 95 interactions were pro-apoptotic, and 39 were anti-apoptotic. As a result, 18 genes functioned as ambivalent factor of apoptosis, for instance, FGF2 both

activated and inhibited the apoptosis process.

With the purpose of exploring the internal dependency relationship between nodes, we calculated the dependency matrix of the PKT205/G3 model by CellNetAnalyzer and performed *in silico* knock-out tests with particular gene deletions. For the wild-type PKT205/G3 model, we found 42025 effect cells in the dependency matrix. 23191 were of no effect type, 16,425 were of ambivalent factor type, 1,100 were weak inhibitors, 1,240 were weak activators, 20 were strong inhibitors and 49 effect cells were strong activators. With the aim of exploring the role of single gene on the whole network, we first investigated the connectivity of those 205 nodes. The connectivity calculation was performed by the NetworkAnalyzer plugin of Cytoscape (Assenov et al, 2008; Shannon et al, 2003). The results shown in Figure 5.13 revealed the distribution of connectivity for 205 nodes in the PKT205/G3 model. The only node with more than 100 interactions was p53. 29 nodes had between 10 and 100 interactions, including DNA damage and apoptosis. The remaining 173 genes or proteins had fewer than 10 interactions.

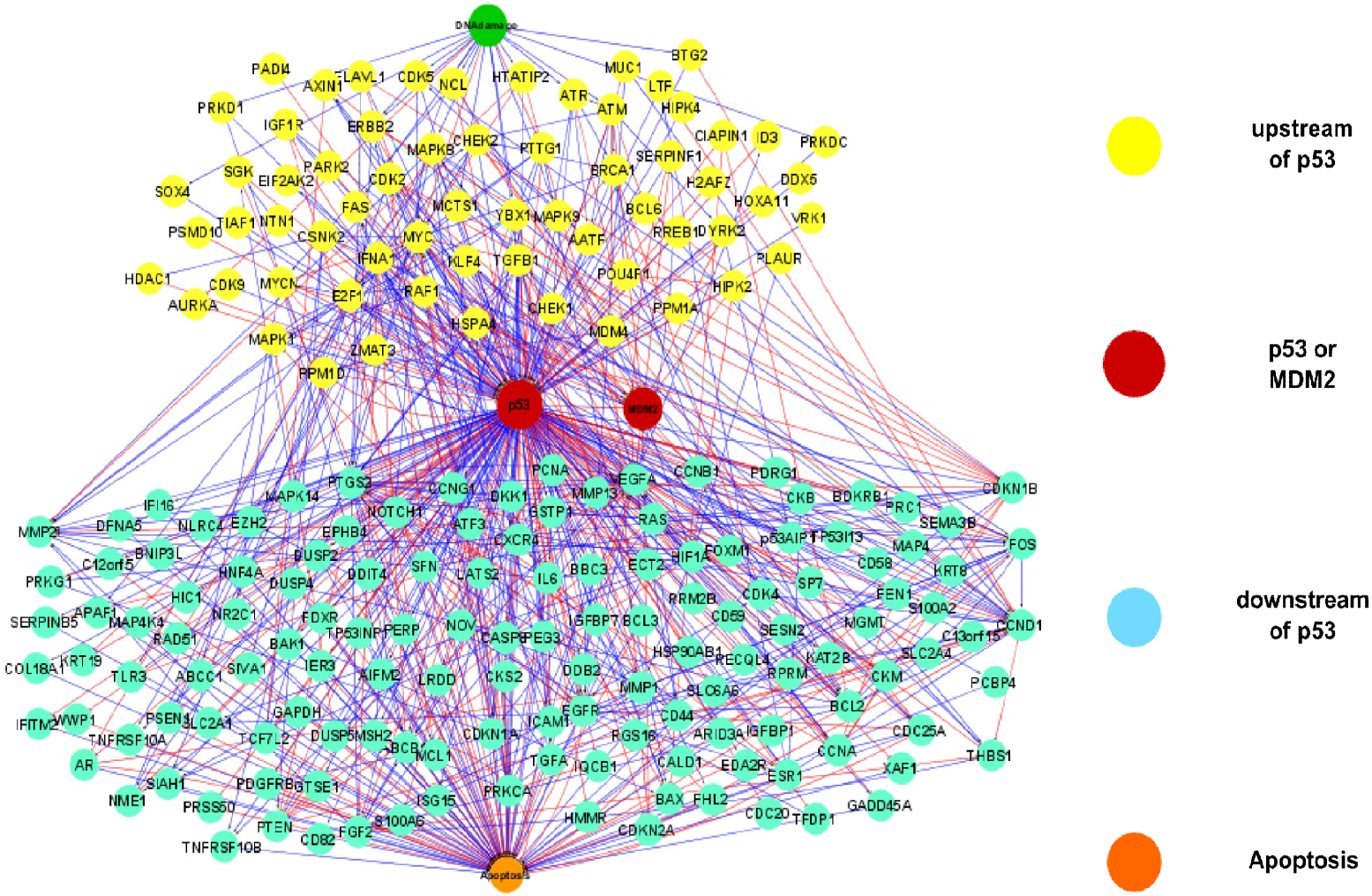

DNA damage
upstream of p53
p53 or MDM2
downstream of p53
Apoptosis
p53
MDM2
DNAdamage
Apoptosis

Figure 5.12: Final PKT205/G3 model network map

Nodes with different functions were represented by different colours. The green node was the input node, DNA damage. The light yellow nodes were the upstream nodes of p53. The p53 and MDM2 were marked by red. The light green nodes were the other downstream nodes of p53 and the output node apoptosis in the bottom was shown by orange.

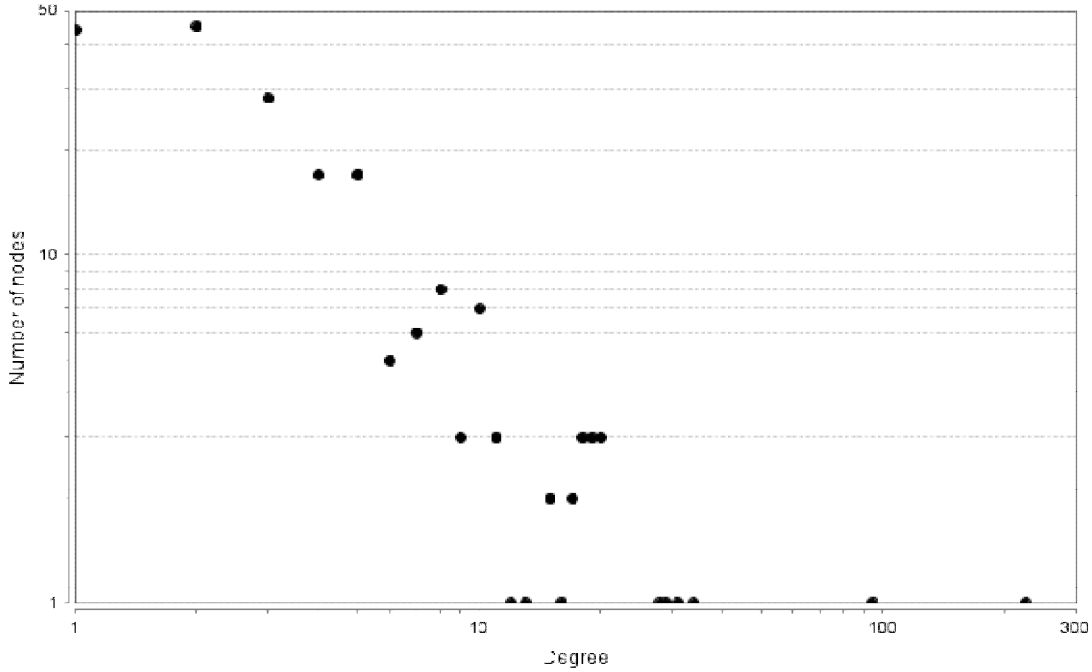


Figure 5.13: Connectivity degree distribution in the PKT205/G3 model

Both axes of the figure are in logarithmic scale and the scatter plot indicates the connectivity degree distribution.

5.9 *In silico* knock-outs of the finalized PKT 205 model

As mentioned in the previous section, the extent of perturbation caused by gene deletion was highly affected by the connectivity degree of the gene removed. We mainly focused on p53 and those 29 genes whose connectivity

was between 10 and 100. 30 *in silico* knock-out tests were performed (Table 5.6). In each knock-out test, one gene was removed from the PKT205 model and dependency matrix calculated. It was found that 11 out of 30 *in silico* knock-out tests had major changes in the new dependency matrix when a certain node was removed (Table 5.6 and Table 5.7). As a result, we identified 58 potential predictions of major changes in dependency cells. The validation of those 58 predictions is illustrated in the next chapter.

Selected Gene	Total effect elements	No Effect	Ambivalent Factor	Weak Inhibitor	Weak Activator	Strong Inhibitor	Strong Activator
Null	42025	23191	16425	1100	1240	20	49
P53	41616	34709	6690	44	79	28	66
MYC	41616	23228	15894	1135	1289	21	49
VEGFA	41616	23935	15159	1137	1307	22	56
PTGS2	41616	23229	15806	1161	1351	20	49
CCND1	41616	23395	15554	1216	1381	21	49
TGFB1	41616	23503	15545	1181	1316	20	51
IL6	41616	23061	16118	1111	1257	20	49
MDM2	41616	23169	15925	1166	1285	20	51
E2F1	41616	23489	15522	1188	1336	20	61
IFNA1	41616	22952	16256	1100	1240	20	48
EGFR	41616	23395	15813	1099	1239	21	49
FOS	41616	23229	15979	1100	1239	20	49
BCL2	41616	23061	16059	1155	1272	20	49
CDKN1B	41616	23061	16145	1101	1240	20	49
RAS	41616	23061	16119	1111	1256	20	49
FGF2	41616	23061	16173	1084	1229	20	49

MMP2	41616	22898	16314	1100	1240	20	44
PRKCA	41616	23061	16142	1100	1244	20	49
MAPK8	41616	23061	16095	1127	1264	20	49
HIF1A	41616	23502	15708	1100	1234	20	52
ESR1	41616	23171	16010	1116	1250	20	49
CXCR4	41616	23228	15956	1108	1254	20	50
CDK2	41616	23228	15977	1099	1242	21	49
ABCB1	41616	22893	16314	1100	1240	20	49
MMP1	41616	22895	16314	1100	1240	19	48
BRCA1	41616	23061	16145	1100	1241	20	49
PTEN	41616	23061	16119	1116	1251	20	49
CCNA	41616	23061	16115	1114	1257	20	49
CSNK2	41616	22953	16256	1100	1240	20	47
ATM	41616	22955	16256	1100	1240	18	47

Table 5.6: Distribution of effect changes in the dependency matrix upon *in silico* knock-out tests

Each row in the table corresponds to a single gene deleted from the PKT205/G3 model for knock-out test. The value “Null” in the selected gene column represents the wild-type.

Serial No	Name of gene knock-out	Protein A	Protein B	Effect element in p53 wild-type	Effect element in knock-out
1	P53 knock-out	SGK	Apoptosis	Ambivalent Factor	Strong inhibitor
2	P53 knock-out	KLF4	CCNB1	Ambivalent Factor	Strong inhibitor
3	P53 knock-out	IFNA1	CDK4	Ambivalent Factor	Strong inhibitor
4	P53 knock-out	IFNA1	FGF2	Ambivalent Factor	Strong inhibitor
5	P53 knock-out	PPM1D	CHEK1	Weak inhibitor	Strong inhibitor
6	P53 knock-out	SFN	CCNB1	Ambivalent Factor	Strong inhibitor
7	P53 knock-out	DNA damage	CDK4	Ambivalent Factor	Strong inhibitor
8	P53 knock-out	DNA damage	FGF2	Ambivalent Factor	Strong inhibitor
9	P53 knock-out	FGF2	CDK4	Ambivalent Factor	Strong activator
10	P53 knock-out	FOXM1	CCNB1	Ambivalent Factor	Strong activator
11	P53 knock-out	FAS	Apoptosis	Ambivalent Factor	Strong activator
12	P53 knock-out	LATS2	Apoptosis	Ambivalent Factor	Strong activator

13	P53 knock-out	PTTG1	CDK4	Ambivalent Factor	Strong activator
14	P53 knock-out	PTTG1	FGF2	Ambivalent Factor	Strong activator
15	P53 knock-out	IFNA1	TLR3	Ambivalent Factor	Strong activator
16	P53 knock-out	IFNA1	FAS	Ambivalent Factor	Strong activator
17	P53 knock-out	DYRK2	P53AIP1	Weak activator	Strong activator
18	P53 knock-out	DYRK2	Apoptosis	Ambivalent Factor	Strong activator
19	P53 knock-out	ATM	CHEK1	Ambivalent Factor	Strong activator
20	P53 knock-out	ATR	CHEK1	Ambivalent Factor	Strong activator
21	P53 knock-out	MAPK14	MMP2	Ambivalent Factor	Strong activator
22	P53 knock-out	MAPK14	BAX	Ambivalent Factor	Strong activator
23	P53 knock-out	MAPK14	SGK	Ambivalent Factor	Strong activator
24	P53 knock-out	DNA damage	CHEK1	Ambivalent Factor	Strong activator
25	P53 knock-out	DNA damage	FAS	Ambivalent Factor	Strong activator

26	MYC knock-out	TCF7L2	Apoptosis	Ambivalent Factor	Strong inhibitor
27	VEGFA knock-out	TLR3	CXCR4	Weak inhibitor	Strong inhibitor
28	VEGFA knock-out	TLR3	TNFRSF10B	Ambivalent Factor	Strong inhibitor
29	VEGFA knock-out	CXCR4	TNFRSF10B	Ambivalent Factor	Strong activator
30	VEGFA knock-out	CXCR4	Apoptosis	Weak activator	Strong activator
31	VEGFA knock-out	FOXM1	MMP2	Ambivalent Factor	Strong activator
32	VEGFA knock-out	FOXM1	BAX	Ambivalent Factor	Strong activator
33	VEGFA knock-out	FOXM1	CCNB1	Ambivalent Factor	Strong activator
34	VEGFA knock-out	FOXM1	Apoptosis	Ambivalent Factor	Strong activator
35	VEGFA knock-out	SERPINB5	Apoptosis	Ambivalent Factor	Strong activator
36	CCND1 knock-out	PDGFRB	Apoptosis	Ambivalent Factor	Strong activator
37	TGFB1 knock-out	DKK1	Apoptosis	Ambivalent Factor	Strong activator
38	TGFB1 knock-out	DNA damage	MAPK8	Ambivalent Factor	Strong activator

39	MDM2 knock-out	ATM	DYRK2	Ambivalent Factor	Strong activator
40	MDM2 knock-out	DNA damage	DYRK2	Ambivalent Factor	Strong activator
41	E2F1 knock-out	AATF	CDK5	Weak activator	Strong activator
42	E2F1 knock-out	CHEK2	AATF	Weak activator	Strong activator
43	E2F1 knock-out	CHEK2	CDK5	Weak activator	Strong activator
44	E2F1 knock-out	CSNK2	MYCN	Ambivalent Factor	Strong activator
45	E2F1 knock-out	ATM	AATF	Ambivalent Factor	Strong activator
46	E2F1 knock-out	ATM	CHEK2	Ambivalent Factor	Strong activator
47	E2F1 knock-out	ATM	CDK5	Ambivalent Factor	Strong activator
48	E2F1 knock-out	ATR	AATF	Ambivalent Factor	Strong activator
49	E2F1 knock-out	ATR	CDK5	Ambivalent Factor	Strong activator
50	E2F1 knock-out	DNA damage	AATF	Ambivalent Factor	Strong activator
51	E2F1 knock-out	DNA damage	CHEK2	Ambivalent Factor	Strong activator

52	E2F1 knock-out	DNA damage	CDK5	Ambivalent Factor	Strong activator
53	EGFR knock-out	BCL3	Apoptosis	Ambivalent Factor	Strong inhibitor
54	HIF1A knock-out	GAPDH	SIAH1	Weak activator	Strong activator
55	HIF1A knock-out	GAPDH	Apoptosis	Ambivalent Factor	Strong activator
56	HIF1A knock-out	SIAH1	Apoptosis	Ambivalent Factor	Strong activator
57	CXCR4 knock-out	TLR3	Apoptosis	Ambivalent Factor	Strong activator
58	CDK2 knock-out	CDKN1A	Apoptosis	Ambivalent Factor	Strong activator

Table 5.7: 58 predictions obtained by *in silico* knock-out tests.

58 major changes from 11 *in silico* knock-out tests are listed from page 200 to page 204.

In the PKT205/G3 model, negative feedback loops and ambivalent interactions between two nodes played an important role to make the whole system less sensitive to the external perturbations. For instance, 30 negative feedback loops and 39 pairs of ambivalent interactions (39 activations and 39 inhibitions) were found in the PKT205/G3 model. The participation of negative feedback loops may affect the stability of the model more than the ambivalent interactions due to their relationship with p53.

According to the number of major changes calculated by CellNetAnalyzer, it was found that the depletion of two-step negative feedback loops results in more perturbations than the absence of ambivalent interactions in the PKT205/G3 network.

Deleted edges	Total number of effect elements	No Effect	Ambivalent Factor	Weak Inhibitor	Weak Activator	Strong Inhibitor	Strong Activator
Null	42025	23191	16425	1100	1240	20	49
56 interactions for 30 two step negative feedback loops	42025	25650	13992	1084	1205	25	69
78 interactions for 39 pairs of ambivalent interactions	42025	23691	15916	1100	1244	22	52

Table 5.8: In silico knock out tests for two-step negative feedback loops and ambivalent interactions

This table lists the number of dependency cells calculated in two in silico knock-out tests for particular interaction depletion.

5.10 Logical steady state analysis of the finalized PKT 205 model

As we aimed to investigate the change of gene expression level under different environments in response to the change of DNA damage stress, logical steady state were performed to determine the state of nodes in the PKT205 model. Here we defined four different scenarios with different input signals and the state of p53 (Table 5.9). The state of node in those four scenarios was shown in Figure 5.14. It was observed that once p53 was absent, a large amount of nodes switched from “ON” or “OFF” to undetermined state (Figure 5.14 C, Figure 5.14 D and Table 5.9). This finding revealed that the whole p53 network became less stable with the absence of p53. As we calculated in Table 5.10, more than 50% of nodes remained unchanged between two scenarios; the number of genes down regulated was larger than the number of genes up regulated when p53 was removed from the PKT205 model. The simulation in Table 5.10 predicted that once the cells were stimulated by DNA damage, the majority of genes did not switch states.

Scenario name	Input signal	Model type	Percentage of determined nodes
Scenario 1	DNA damage ON	P53 wild-type	87.8%
Scenario 2	DNA damage OFF	P53 wild-type	88.3%
Scenario 3	DNA damage ON	P53 knock-out	45.6%
Scenario 4	DNA damage OFF	P53 knock-out	46.1%

Table 5.9: List of scenarios for the logical steady state analysis

Four scenarios of logical steady state analysis with different input signals are defined with their input signal, model type and percentage of nodes having a determined state.

Source Scenario	Target Scenario	Total number of genes for prediction	Number of genes up regulated	Number of genes unchanged	Number of genes down regulated
P53 wild type with DNA damage	P53 mutant with DNA damage	202	29 (14%)	113 (56%)	60 (30%)
P53 wild type without DNA damage	P53 mutant with DNA damage	202	30 (15%)	112 (55%)	60 (30%)
P53 wild type without DNA damage	P53 wild type with DNA damage	202	5 (2%)	185 (92%)	12 (6%)
P53 mutant without DNA damage	P53 mutant with DNA damage	202	7 (3%)	181 (90%)	14 (7%)

Table 5.10: Statistic results of Logical Steady State Analysis

The distribution of number for 202 p53 interacting genes with different change between 4 different scenarios was listed.

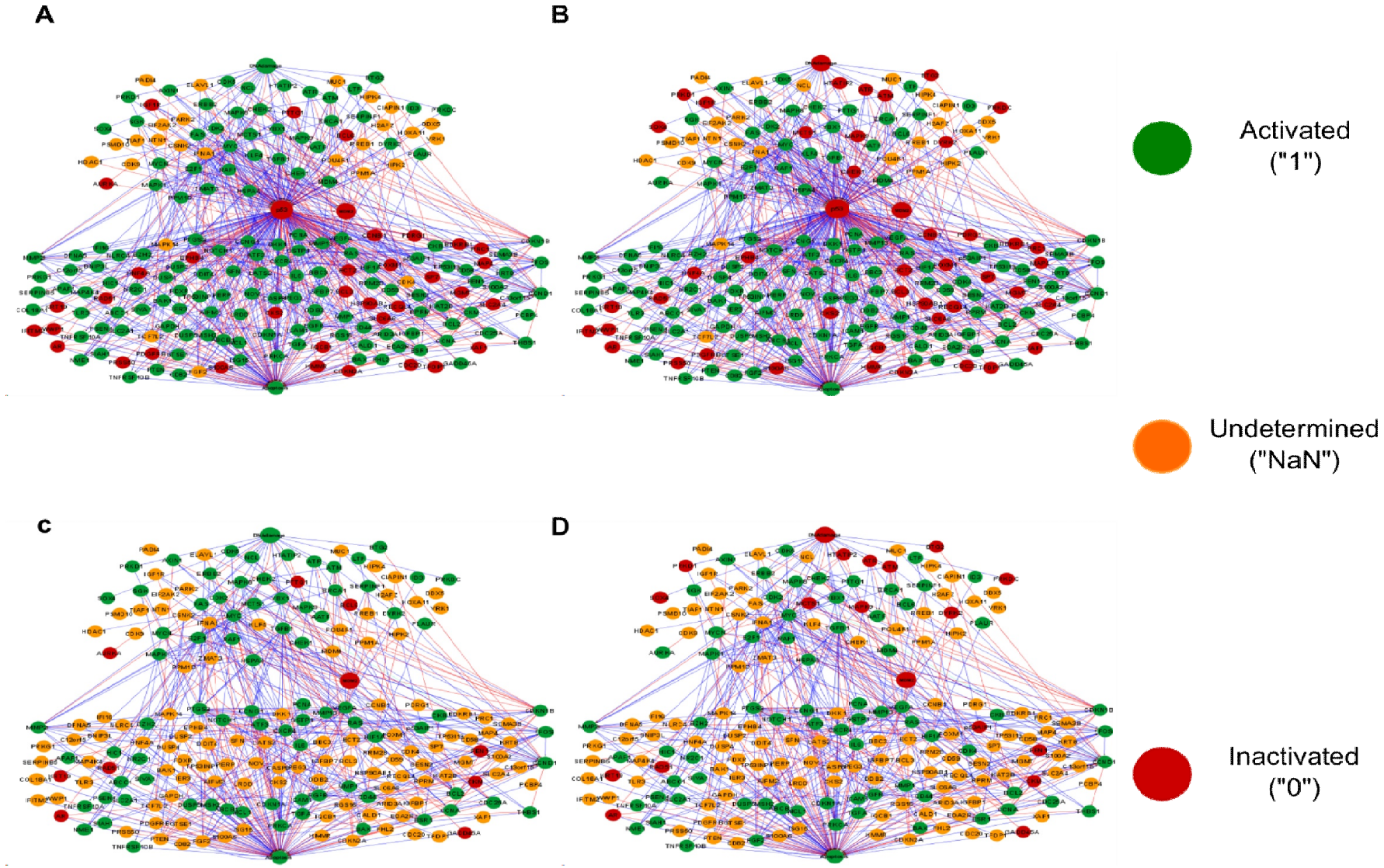

A
B
c
D
Activated
("1")
Undetermined
("NaN")
Inactivated
("0")

Figure 5.14: Logical steady state analysis result of four scenarios

The logical steady states of nodes in four scenarios are shown respectively: (a) P53 wild type when DNA damage was ”ON”; (b) P53 wild type when DNA damage was ”OFF”; (c) P53 mutant when DNA damage was ”ON”; (d) P53 mutant when DNA damage was ”OFF”. The nodes with state “ON” were represented by green, the nodes with state ”NaN” (un determined) were represented by orange, and the nodes with state “OFF” were represented by red.

In order to obtain further insight into the dynamic changes in apoptosis, we distinguished the number of 39 anti-apoptotic genes (Table 5.11) and 56 pro-apoptotic genes (Table 5.12) with different change in the comparison of different scenarios.

The changes in the state of anti-apoptotic genes are shown in Table 5.11 and those of pro-apoptotic genes are listed in Table 5.12 below. This distribution illustrates the reason why the apoptosis output node was also activated in p53 mutant cells. The majority of those 56 pro-apoptotic genes in Table 5.12 and 39 anti-apoptotic genes in Table 5.11 had no change once they were treated by DNA damage. The absence of p53 caused obvious changes of both pro-apoptotic and anti-apoptotic genes once the cells were treated with DNA damage. The number of pro-apoptotic and anti-apoptotic genes which were up regulated or down regulated increased with the depletion of p53. Among 39 anti-apoptotic genes, the expression of BCL3, PDGFRB, WWP1, IGF1R, PRSS50, EPHB4 and CKS2 was up regulated from p53 wild type to p53 mutant cells after treatment of DNA damage. The up regulation of 7 anti-apoptotic genes was also found from p53 wild type to p53 mutant cells without DNA damage treatment. Notably, IGF1R (insulin-like growth factor 1 receptor) and PDGFRB (platelet-derived growth factor receptor, beta polypeptide) were up regulated in p53 minus scenarios, which together with

FGF2 changes highlights that growth factor mediated signalling pathways are important factors contributing to survival of these tumours. Strategies targeting on inhibiting those growth factors may promote the death of tumour cells with p53 mutation and inhibitors of those anti-apoptotic genes may be suited for anti-cancer agents. Among those 56 pro-apoptotic genes, the expression of 27 genes (Table 5.12) was down regulated from p53 wild type cells to p53 mutant cells when those cells were treated by DNA damage and 29 p53 genes (Table 5.12) were repressed from p53 wild type cells to p53 mutant cells without DNA damage. The promotion of those pro-apoptotic genes may be a potential strategy for cancer treatments. Moreover, it was found that FAS and p53AIP1 were up-regulated in p53 mutant cells when treated by DNA damage. FGF2 (fibroblast growth factor 2(basic)) had both pro-apoptotic and anti-apoptotic function in the PKT205/G3 model and it was down-regulated in p53 wild type cells or p53 mutant cells in the presence of DNA damage. Approaches that will decrease expression of anti-apoptotic genes and increase expression of pro-apoptotic genes would improve cancer therapy and therefore these genes represent potential therapeutic targets.

Source Scenario	Target Scenario	Total number of genes	Population of genes up-regulated (Percentage)	Population of genes not changed (Percentage)	Population of genes down-regulated (Percentage)
P53 wild type with DNA damage	P53 mutant with DNA damage	39	7 (18%) (BCL3; PDGFRB; WWP1; IGF1R; PRSS50; EPHB4; CKS2)	24(62%) (GSTP1; ESR1; IL6; FGF2; EGFR; CD44; PRKCA; TGFA; CDC25A; FOS; EZH2; PSEN1; BCL2; CCNG1; TCF7L2; CDKN1B; ATF3; PTGS2; SGK; MCL1; NOTCH1; VEGFA; CDKN1A; AR)	8(20%) (DUSP4;DUSP2; DDIT4; MAP4K4; FHL2; IER3; C12orf5; SFN)
P53 wild type without DNA damage	P53 mutant without DNA damage	39	7 (18%) (BCL3; PDGFRB; WWP1; IGF1R; PRSS50; EPHB4; CKS2)	24(62%) (GSTP1; ESR1; IL6; FGF2; EGFR; CD44;; PRKCA; TGFA; CDC25A; FOS; EZH2; PSEN1; BCL2;CCNG1; TCF7L2; CDKN1B;; ATF3; PTGS2; SGK; MCL1; NOTCH1; VEGFA; CDKN1A; AR)	8(20%) (DUSP4; DUSP2; DDIT4; MAP4K4; FHL2; IER3; C12orf5; SFN)

P53 wild type without DNA damage	P53 wild type with DNA damage	39	0(0%)	38(97%) (BCL3; GSTP1; ESR1; DUSP4; IL6; PDGFRB; WWP1; I GF1R; EGFR; CD44; PRKCA; DUSP2; TGFA; CDC25A; FOS; DDIT4; MAP4K4; EZH2; PSEN1; PRSS50; BCL2; FHL2; CCNG1; TCF7L2; EPHB4; CDKN1B; ATF3; PTGS2; SGK; MCL1; IER3; NOTCH1; VEGFA; CDKN1A; AR; CKS2; C12orf5; SFN)	1(3%) (FGF2)
P53 mutant without DNA damage	P53 mutant with DNA damage	39	0(0%)	38(97%) (BCL3; GSTP1; ESR1; DUSP4; L6; PDGFRB; WWP1; IGF1R; EGFR; CD44; PRKCA; DUSP2; TGFA; CDC25A; FOS; DDIT4; MAP4K4; EZH2; PSEN1; PRSS50; BCL2; FHL2; CCNG1; TCF7L2; EPHB4; CDKN1B; ATF3; PTGS2; SGK; MCL1; IER3; NOTCH1; VEGFA; CDKN1A; AR; CKS2; C12orf5;SFN)	1(3%) (FGF2)

Table 5.11: Number of anti-apoptotic genes with altered expression depending on p53 and DNA damage

The distribution of anti-apoptotic genes in the PKT205 model that change their expression between four different scenarios were calculated by comparing the steady state of the source and the target scenario.

Source Scenario	Target Scenario	Total number of genes	Population of genes up-regulated (Percentage)	Population of genes not changed (Percentage)	Population of genes down-regulated (Percentage)
P53 wild type with DNA damage	P53 mutant with DNA damage	56	5 (9%) (ECT2;XAF1; IFITM2;HNF4A; CDKN2A)	24 (43%) (ESR1; MSH2; CXCR4; I L6; FGF2; EGFR; CD44; PRKCA; BAX; CDC25A; FOS; SIVA1; CCNG1; CDKN1B; APAF1; ATF3; PTGS2; NOTCH1; FAS; AR; TNFRSF10A; TNFRSF10B; SIAH1; p53AIP1)	27(48%) (DUSP4; DUSP2; FDXR; IFI16; TLR3; DDIT4; MAP4K4; SEMA3B; PEG3; PCBP4; DFNA5; TP53INP1; CASP8; COL18A1; BBC3; BAK1; NLRC4; GADD45A; PTEN; AIFM2; DKK1; BNIP3L; IGFBP7;LATS2; SERPINB5; PERP; LRDD)
P53 wild type without DNA damage	P53 mutant without DNA damage	56	5 (9%) (ECT2;XAF1; IFITM2;HNF4A; CDKN2A)	22 (39%) (ESR1; MSH2; CXCR4; IL6; FGF2; EGFR; CD44; PRKCA; BAX; CDC25A; FOS; SIVA1; CCNG1; CDKN1B; APAF1; ATF3; PTGS2; NOTCH1; AR;T NFRSF10A; TNFRSF10B; SIAH1)	29 (52%) (DUSP4; DUSP2; FDXR; IFI16; TLR3; DDIT4; MAP4K4; SEMA3B; PEG3; PCBP4; DFNA5; TP53INP1; CASP8; COL18A1; BBC3; BAK1; NLRC4; GADD45A; FAS; PTEN; AIFM2; DKK1; BNIP3L; IGFBP7; LATS2; SERPINB5; PERP; LRDD; p53AIP1)

P53 wild type without DNA damage	P53 wild type with DNA damage	56	0 (0%)	55 (98%) (ESR1; ECT2; MSH2; DUSP4; CXCR4; XAF1; L6; EGFR; CD44; PRKCA; DUSP2; FDXR; BAX; IFI16; TLR3; CDC25A; FOS; DDIT4; MAP4K4; SEMA3B; PEG3; IFITM2; SIVA1; PCBP4; DFNA5; HNF4A; TP53INP1; CCNG1; CASP8; CDKN1B; COL18A1; BBC3; APAF1;BAK1; NLRC4; ATF3; PTGS2; GADD45A; NOTCH1; FAS; PTEN; AIFM2; DKK1; AR; BNIP3L; TNFRSF10A; TNFRSF10B; IGFBP7; LATS2; SERPINB5; PERP; SIAH1; LRDD; p53AIP1; CDKN2A)	1(2%) (FGF2)

P53 mutant without DNA damage	P53 mutant with DNA damage	56	2 (4%) (FAS p53AIP1)	53 (94%) (ESR1; ECT2; MSH2; DUSP4; CXCR4; XAF1; IL6; EGFR; CD44; PRKCA; DUSP2; FDXR; BAX; IFI16; TLR3; CDC25A; FOS; DDIT4; MAP4K4; SEMA3B; PEG3; IFITM2; SIVA1; PCBP4; DFNA5; HNF4A; TP53INP1; CCNG1; CASP8; CDKN1B; COL18A1; BBC3; APAF1; BAK1; NLRC4; ATF3; PTGS2; GADD45A; NOTCH1; PTEN; AIFM2; DKK1; AR; BNIP3L; TNFRSF10A; TNFRSF10B; IGFBP7; LATS2; SERPINB5; PERP; SIAH1; LRDD; CDKN2A)	1(2%) (FGF2)

Table 5.12: Number of pro-apoptotic genes with altered expression depending on p53 and DNA damage

The distribution of pro-apoptotic genes in the PKT205 model that change their expression between four different scenarios were calculated by comparing the steady state of the source and the target scenario.

5.11 Discussion

The construction process of the PKT205/G1 model revealed six different types of errors caused by text mining. Those errors indicated that there was a requirement for manual curation to avoid the inclusion of incorrect interactions when the size of our model increased. However, it was also clear that automated extraction substantially decreased the time required to build the model and provided the only feasible solution of generating the complete p53 interactome. It also highlighted the need for building databases with better NLP (Natural Language Processing) and text mining techniques to avoid those artifacts and allow automated extraction for any process with minimal curation.

The simulation results of the PKT205/G1 model indicated that the change of dependency relationships between nodes in the model was substantially affected by the connectivity degree of the genes. The genes with higher connectivity had more significant effect on the internal relationships between remaining genes in the model. Those genes which were at the lowest connectivity level had no or small effect on the model. The two step negative feedback loops and ambivalent interactions between two genes played an important role to make the whole system robust to external perturbations; for instance, there were both positive and negative pathways between DNA damage and apoptosis. The involvement of two step feedback loops may

increase the stability of the model compared to ambivalent interactions due to their relationship with p53, according to the knock-out tests of two step feedback loops in the PKT205/G3 model (Table 5.8). Due to the complexity of the network, only two step feedback loops and ambivalent interactions were considered here. As we reported in the *in silico* knock-out tests, p53 was the most important factor and the presence of p53 made the whole network more stable (Table 5.9). The p53 node was involved in 47% of the two-step negative feedback loops but the ambivalent interactions occupied a minority of interactions connected to p53 (less than 10%).

By logical steady state analysis, it was found that those upstream genes had different effects in determining the state of other nodes and this finding may help us evaluate the global effects of these p53 upstream genes onto the whole network.

When the DNA damage input signal was switched from ON to OFF, the state of the output node, apoptosis was always ON. Once upstream nodes which were connected to negative feedback loops or had ambivalent interactions were induced by input signals, the majority of nodes in the model had determined states irrespectively of the state of the input signal. This finding above indicated the importance of input node selection and the limits of the PKT205/G1 model, which was an isolated system and could not communicate with the environment. As a result, the PKT205/G1 model was extended into PKT205/G2. The advantage of PKT205 in G2 version is that the input signal, DNA damage and the output signal, apoptosis were connected to the model using Gene Ontology (GO) terms, text mining and manual curation so that the effects from the model onto external signals could be investigated. Through

the logical steady state analysis of the PKT205/G2, it was found that the whole system became less stable when p53 was removed from the network.

The achievement of the PKT205/G1 model was feasibility of the approach to construct a p53 complex network model from vast amounts of interaction information. The same approach can be used to create an interactome for any other gene. The disadvantage of PKT205/G2 is that this was a limited model and it only considered the genes or proteins which interact with p53 directly. The potential effect from other genes or proteins which interact with p53 indirectly was ignored. By the investigation of the logical steady state analysis in the PKT205/G1model, it was found that the connectivity of nodes played an important role in the stability level of the whole network model. Those interactions not included in the PKT205/G2 model may have potential effects on determining the state of other genes or proteins in the p53 pathways. This factor may weaken the predictive power of PKT205/G2 and may cause errors in predictions when the simulations are compared with the experimental data. In order to improve the PKT205 model, we developed the G3 version of this model and finalized PKT205/G3 as the PKT205 model. Table 5.11 and 5.12 revealed genes that may be targeted as potential cancer strategies. As I mentioned in the previous sections, in cells with mutant p53 not treated with chemotherapy inducing DNA damage, 29 out of 58 pro-apoptotic genes were down regulated, 22 pro-apoptotic genes did not change and only 5 pro-apoptotic genes were up regulated (Table 5.12). Meanwhile, 38 out of 39 anti-apoptotic genes remained the same and only FGF2 was down-regulated (Table 5.11). This finding illustrated that in tumour cells with p53 mutant, the probability of apoptosis was decreased and cells survived. As a result, those genes promoting cell death were regarded as potential chemotherapy target. For example, we found that 2 pro-apoptotic genes, FAS and p53AIP1 became

up regulated, and lead to an increase in apoptosis probability to promote tumour cell death (Table 5.12), when cells with the mutant p53 were treated by DNA damage. They may be selected as potential target of cancer treatment. Several important predictions were obtained from our model, which will help us to get deeper insights into the mechanisms of p53 pathways.

Chapter 6 Validation of model predictions

6.1 Introduction

As described in previous chapters, we finalized the PKT205 model and performed *in silico* analysis. A large amount of predictions were produced from the analysis results. With the purpose of measuring the predictive strength of our model and validating the simulation predictions we compared those predictions with results obtained through literature survey, western blotting experiments, microarray expression profiles and ChIP-sequencing data.

6.2 *In silico* knock-out predictions and literature validation

Since we performed 30 *in silico* knock-out tests for the finalized PKT205 model presented in the previous chapter, 58 predictions were obtained from those 11 *in silico* knock-out tests which had major changes in strong inhibitor or strong activator activities (Table 6.1). All those 58 predictions were investigated manually by PubMed and Google on line searching. Changes that we didn't confirm through this approach were named potentially novel predictions (PNP) reflecting limitations of manual searches of the vast p53 related literature. We were able to confirm 4 out of these 58 predictions through literature searches, focusing on major changes caused by the p53 deletion which were expected to have strong experimental effects and they are listed below:

1. The effect of DNA damage onto FAS (Fas (TNF receptor superfamily, member 6)) changed from an ambivalent factor in the p53 wild-type model to a strong activator when p53 was removed. Manna et al (2011) have determined that in p53 minus cells, Fas protein levels are elevated under DNA damage compared to p53 wild-type cells, which is in agreement with our prediction

(Manna et al, 2011).

2. Similarly to FAS, the effect of LATS2 (LATS, large tumor suppressor, homolog 2 (Drosophila)) onto apoptosis was changed from an ambivalent factor in the p53 wild-type model to a strong activator when p53 was removed. It was found that in both p53 wild-type (A549) and p53 minus cells (H1299), LATS2 was able to induce apoptosis and that apoptosis is slightly increased in H1299 as measured by PARP and caspase 9 cleavage (Ke et al, 2004).

3. We observed that the effect of DNA damage onto CHEK1 (checkpoint kinase 1) changed from an ambivalent factor in the p53 wild-type to a strong activator when p53 was removed. CHEK1 protein levels were found to be higher in p53 -/- cells than in p53 +/+ HCT116 colorectal cancer cells treated by daunorubicin (Gottifredi et al, 2001), which also matches our predictions (Table 6.1).

4. It was reported that KLF4 (Kruppel-like factor 4(gut)) caused more reduction of CCNB1 (cyclin B1) expression in p53 -/- HCT116 than in p53 +/+ HCT116 cells (Yoon & Yang, 2004), which matched our model prediction.

However, one prediction out of those 58 predictions was found opposite to the literature evidence. The prediction pointed out that IFNA1 (interferon, alpha 1) enhanced TLR3 (toll-like receptor 3) in p53 mutant cells compared to p53 wild type cells. But this was opposite to the fact reported by Taura et al that IFNA1 with the effect of DNA damaging drug 5-fluoro-uracil(5-FU) reduced the expression of TLR3 in p53 -/- HCT116 cell compared to p53 +/+ HCT116 cells (Taura et al, 2010).

Activated node	Deleted gene	Reported effects from literature	References	Predictions	Verified state
DNA damage	p53	Expression level of Fas enhanced	(Manna et al, 2011)	DNA damage promoted more up-regulation of FAS than it in p53 wild type	Verified by literature
LATS2	p53	cell death enhanced	(Ke et al, 2004)	LATS2 induced more apoptosis than it in p53 wild type	Verified by literature
DNA damage	p53	Expression level of CHEK1 enhanced	(Gottifredi et al, 2001)	DNA damage promoted more up-regulation of CHEK1 than it in p53 wild type	Verified by literature
KLF4	p53	CCNB1 reduced	(Yoon & Yang, 2004)	KLF4 reduced more expression of CCNB1 than in p53 wild type	Verified by literature
ATM	p53			ATM enhanced more CHEK1 than it in p53 wild type	Verified by lab work
ATR	p53			ATR enhanced more CHEK1 than it in p53 wild type	Verified by lab work
MAPK14	p53	Stimulation of BAX	(Gomez-Lazaro et al, 2008)	BAX enhanced	Consistent with Prediction

SERPINB5	VEGFA	Apoptosis enhanced with the presence of MMP3 and MMP9 inhibition	(Rossiter et al, 2008)	Apoptosis enhanced	Consistent with Prediction
ATM	MDM2	DYRK2 induced in the presence and absence of DNA damage	(Taira et al, 2010)	DYRK2 enhanced	Consistent with Prediction
ATR	MDM2	DYRK2 induced in the presence and absence of DNA damage	(Taira et al, 2010)	DYRK2 enhanced	Consistent with Prediction
CDKN1A	CDK2	Apoptosis decreased but not confirmed directly	(Delavaine & La Thangue, 1999) (Satyanarayana et al, 2008)	Apoptosis reduction	Consistent with Prediction
IFNA1	p53			CDK4 reduction	PNP
IFNA1	p53			FGF2 reduction	PNP
PPM1D	p53			CHEK1 reduction	PNP

SFN	p53			CCNB1 reduction	PNP
DNA damage	p53			CDK4 reduction	PNP
DNA damage	p53			FGF2 reduction	PNP
FGF2	p53			CDK4 enhanced	PNP
FOXM1	p53			CCNB1 enhanced	PNP
FAS	p53			Apoptosis enhanced	PNP
PTTG1	p53			CDK4 enhanced	PNP
PTTG1	p53			FGF2 enhanced	PNP
IFNA1	p53			FAS enhanced	PNP
DYRK2	p53			P53AIP1 enhanced	PNP
DYRK2	p53			Apoptosis enhanced	PNP
MAPK14	p53			MMP2 enhanced	PNP

MAPK14	p53			SGK enhanced	PNP
TCF7L2	MYC			Apoptosis reduction	PNP
TLR3	VEGFA			CXCR4 reduction	PNP
TLR3	VEGFA			TNFRSF10B reduction	PNP
CXCR4	VEGFA			TNFRSF10B enhanced	PNP
CXCR4	VEGFA			Apoptosis enhanced	PNP
FOXM1	VEGFA			MMP2 enhanced	PNP
FOXM1	VEGFA			BAX enhanced	PNP
FOXM1	VEGFA			CCNB1 enhanced	PNP
FOXM1	VEGFA			Apoptosis enhanced	PNP
PDGFRB	CCND1			Apoptosis reduction	PNP
DKK1	TGFB1			Apoptosis reduction	PNP

DNA damage	TGFB1			MAPK8 enhanced	PNP
AATF	E2F1			CDK5 enhanced	PNP
CHEK2	E2F1			AATF enhanced	PNP
CHEK2	E2F1			CDK5 enhanced	PNP
CSNK2	E2F1			MYCN enhanced	PNP
ATM	E2F1			AATF enhanced	PNP
ATM	E2F1			CHEK2 enhanced	PNP
ATM	E2F1			CDK5 enhanced	PNP
ATR	E2F1			AATF enhanced	PNP
ATR	E2F1			CDK5 enhanced	PNP
DNA damage	E2F1			AATF enhanced	PNP
DNA damage	E2F1			CHEK2 enhanced	PNP

DNA damage	E2F1			CDK5 enhanced	PNP
BCL3	EGFR			Apoptosis reduction	PNP
GAPDH	HIF1A			SIAH1 enhanced	PNP
GAPDH	HIF1A			Apoptosis enhanced	PNP
SIAH1	HIF1A			Apoptosis enhanced	PNP
TLR3	CXCR4			Apoptosis enhanced	PNP
IFNA1	p53	TLR3 reduction	(Taura et al, 2010)	TLR3 enhanced	Opposite to Prediction

Table 6.1: Validations of model predictions of the *in silico* knock-out test

This table lists 58 predictions in the selected gene deletion background. Some of these predictions were verified by existing literature survey or laboratory based experiments, and the other was potential novel predictions (PNP).

6.3 Western blotting validation

In the previous section, we investigated 58 potential predictions by literature survey, and found that four predictions were confirmed by literature evidence, whereas one prediction was opposite to the literature evidence. We performed lab experiments for further validations, which enabled us to confirm 2 predictions by western blotting. As seen in Table 6.1, once p53 was absent, the stimulation of CHEK1 by ATM or ATR was enhanced by its phosphorylation. The western blotting experiments obtained are displayed (Figure 6.1C) and have validated those two predictions as CHEK1 phosphorylation on ATM/ATR specific site increased in SAOS2 cells that are p53 deficient when compared with U2OS cells that have wild type p53.The bars representing stimulated CHEK1 expression were marked in a red rectangle (Figure 6.1C). These results also demonstrated that ATM and ATR maybe more active in p53 minus background. Figure 6.1 explains this change by presenting a subset of our model structure. With the presence of p53, there were both positive and negative paths from ATM to CHEK1. The negative path through p53 made ATM to be an ambivalent factor of CHEK1. Once p53 was deleted, only the positive path from ATM to CHEK1 remained and the expression level of CHEK1 was stimulated by the activation of ATM.

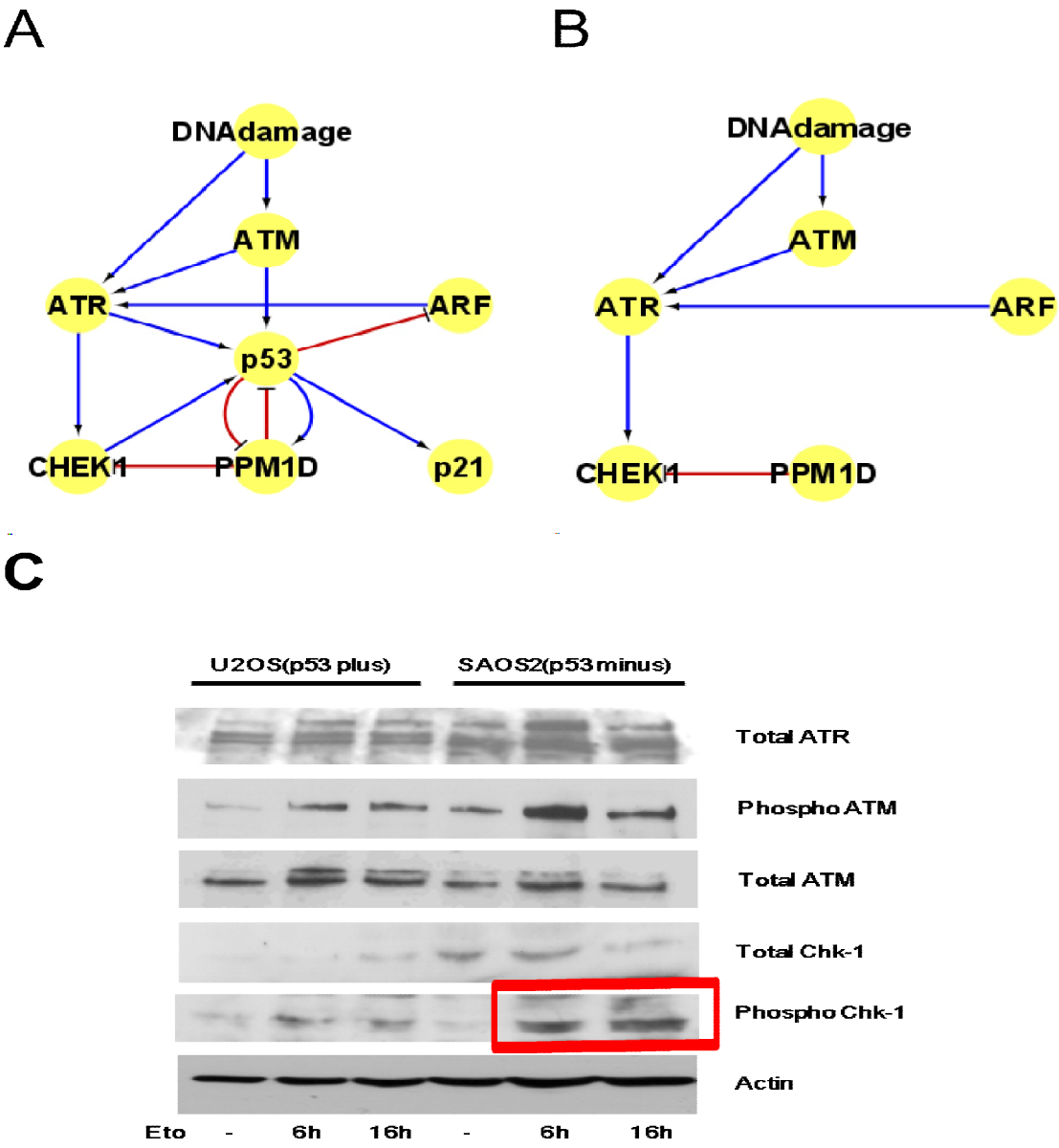


Figure 6.1: Positive and negative pathways from ATM/ATR to CHEK1

(a) Positive and negative pathways from ATM/ATR to CHEK1 in p53 wild type cells as known from literature survey; (b) Positive and negative pathways from ATM/ATR to CHEK1 in p53 minus cells; (c) Chk1 (CHEK1) activation is increased in p53 negative background. U2OS cells that have functional p53 and SAOS2 cells that lack functional p53 were treated with 10 μM etoposide for 16 hours. Cell extracts were analyzed by SDS PAGE and western blot analysis using antibodies against total Chk1, ATR and ATM. ATM phosphorylated Chk1 at serine 1981 and ATR phosphorylated Chk1 at serine 345.

.

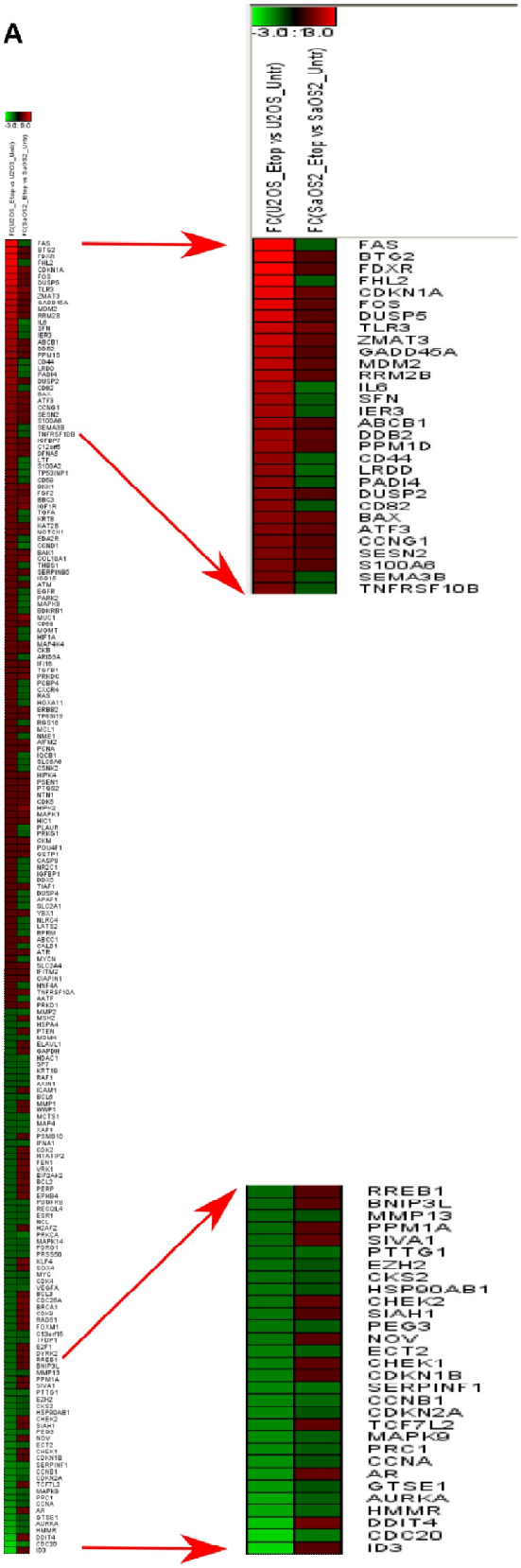
6.4 Genome wide experimental validation

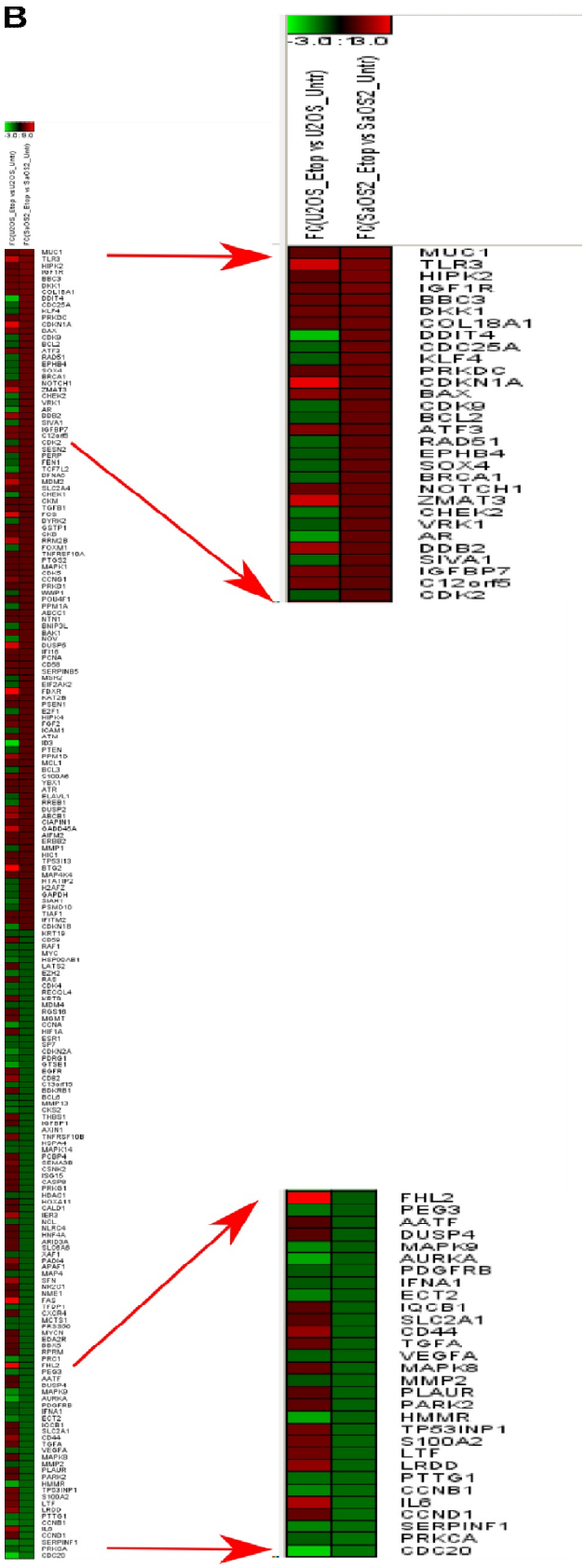
Although we verified those potential predictions by literature survey and western blotting, those confirmed predictions constituted a small percentage of the total predictions. As a result, we turned to genome wide experiments for further validation of the predictive strength of our model. Here microarray experimental data in three different types of cell lines were utilized for validation: U2OS human osteosarcoma cells, SAOS2 cells are also human osteosarcoma cells, and HCT116 human colon carcinoma cells. U2OS human osteosarcoma cells were p53 wild type; SAOS2 cells had non-functional p53. HCT116 human colon carcinoma cells included two different types for p53: p53+/+ and p53-/-. Results of microarray experiments in colon cancer cells were obtained from results published in GSE10795 from PubMed (Wilhelm et al, 2008). All microarray results described in this study will be deposited in publicly accessible database. We compared our model prediction from logical steady state analysis with microarray experimental data for those four cell lines. DNA microarray data provides gene expression data at a steady state rather than in a dynamic process and it is difficult to obtain time series data in certain situations (Siegel et al, 2006), such as the research of individual cancer patients. As a result, DNA microarray data was regarded as qualitative data and we assigned finite signs to represent the gene expression level so as to compare gene expression between different cell lines or samples. Meanwhile, logical steady states include finite state values and those states were correlated or determined by logical functions such as AND, OR and NOT. This simplification facilitates the comparison of gene expression between different cells or samples. As we mentioned in the method chapter, we used an approach to compare model predictions and microarray expression data.

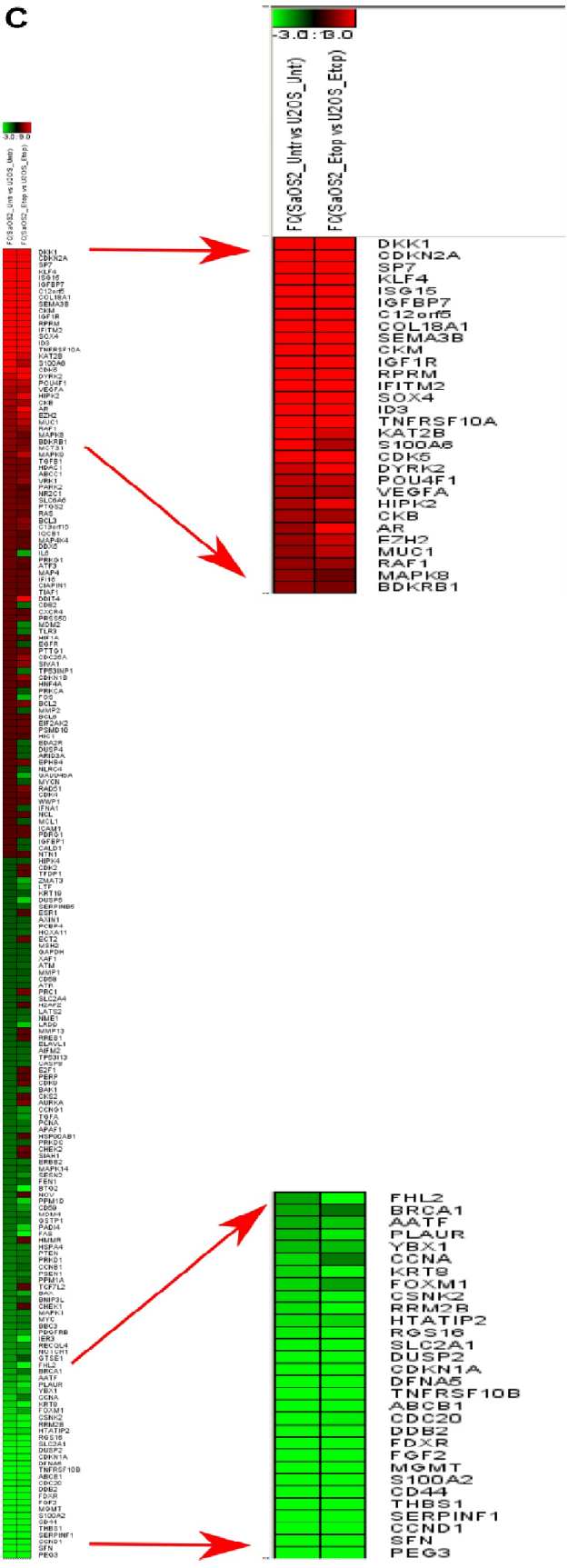
Firstly, the distribution of genome wide gene expression changes between different cell lines either treated by DNA damage or not was investigated. Then the median of expression values for genes corresponding to multiple probes was calculated, in order to reduce noise. The fold changes between different scenarios were calculated according to the approach described in the method chapter to measure the gene expression change. The heat map of median expression value were represented by Genesis (Figure 6.2) (Sturn et al, 2002). The gene expression profile was analysed by fold changes and p-values. These data were provided by the Faculty of Life Science Microarray Core Facility. All expression profiles were filtered by fold change more than 1.5 and p-value less than 0.05.

After significant genes whose expression level were up regulated or down regulated, were identified, the functional annotation analysis of those probe ids was performed using DAVID (Huang et al, 2009) (Additional Table 8) to explore the biological functions of activated genes which interact with p53.

Those GO terms detected by DAVID in Additional Table 8 indicated that the target genes of p53 exert diverse biology activities in the p53 wild type cells in response to DNA damage. Once p53 lost its function and was absent in the p53 mutant cells, the whole network became instable and biological functions occurring in normal cells were disturbed in the p53 mutant cells. Additional Table 8 revealed a shift in apoptosis, cell cycle, genes senescence and DNA repair process which p53 target genes were involved in, when p53 was deficient.


A
-3.0 : 3.0
FC(U2OS_Etop vs U2OS_Untr)
FC(SaOS2_Etop vs SaOS2_Untr)
FAS
BTG2
FDXR
FHL2
CDKN1A
FOS
DUSP5
TLR3
ZMAT3
GADD45A
MDM2
RRM2B
IL6
SFN
IER3
ABCB1
DDB2
PPM1D
CD44
LRDD
PADI4
DUSP2
CD82
BAX
ATF3
CCNG1
SESN2
S100A6
SEMA3B
TNFRSF10B
RREB1
BNIP3L
MMP13
PPM1A
SIVA1
PTTG1
EZH2
CKS2
HSP90AB1
CHEK2
SIAH1
PEG3
NOV
ECT2
CHEK1
CDKN1B
SERPINF1
CCNB1
CDKN2A
TCF7L2
MAPK9
PRC1
CCNA
AR
GTSE1
AURKA
HMMR
DDIT4
CDC20
ID3


B
-3.0 : 18.0
FC(U2OS_Etop vs U2OS_Untr)
FC(SaOS2_Etop vs SaOS2_Untr)
MUC1
TLR3
HIPK2
IGF1R
BBC3
DKK1
COL18A1
DDIT4
CDC25A
KLF4
PRKDC
CDKN1A
BAX
CDK9
BCL2
ATF3
RAD51
EPHB4
SOX4
BRCA1
NOTCH1
ZMAT3
CHEK2
VRK1
AR
DDB2
SIVA1
IGFBP7
C12orf5
CDK2
FHL2
PEG3
AATF
DUSP4
MAPK9
AURKA
PDGFRB
IFNA1
ECT2
IQCB1
SLC2A1
CD44
TGFA
VEGFA
MAPK8
MMP2
PLAUR
PARK2
HMMR
TP53INP1
S100A2
LTF
LRDD
PTTG1
CCNB1
IL6
CCND1
SERPINF1
PRKCA
CDC20


C
-3.0 : 3.0
FC(SaOS2_Untr vs U2OS_Untr)
FC(SaOS2_Etop vs U2OS_Etop)
DKK1
CDKN2A
SP7
KLF4
ISG15
IGFBP7
C12orf5
COL18A1
SEMA3B
CKM
IGF1R
RPRM
IFITM2
SOX4
ID3
TNFRSF10A
KAT2B
S100A6
CDK5
DYRK2
POU4F1
VEGFA
HIPK2
CKB
AR
EZH2
MUC1
RAF1
MAPK8
BDKRB1
FHL2
BRCA1
AATF
PLAUR
YBX1
CCNA
KRT9
FOXM1
CSNK2
RRM2B
HTATIP2
RGS16
SLC2A1
DUSP2
CDKN1A
DFNA5
TNFRSF10B
ABCB1
CDC20
DDB2
FDXR
FGF2
MGMT
S100A2
CD44
THBS1
SERPINF1
CCND1
SFN
PEG3

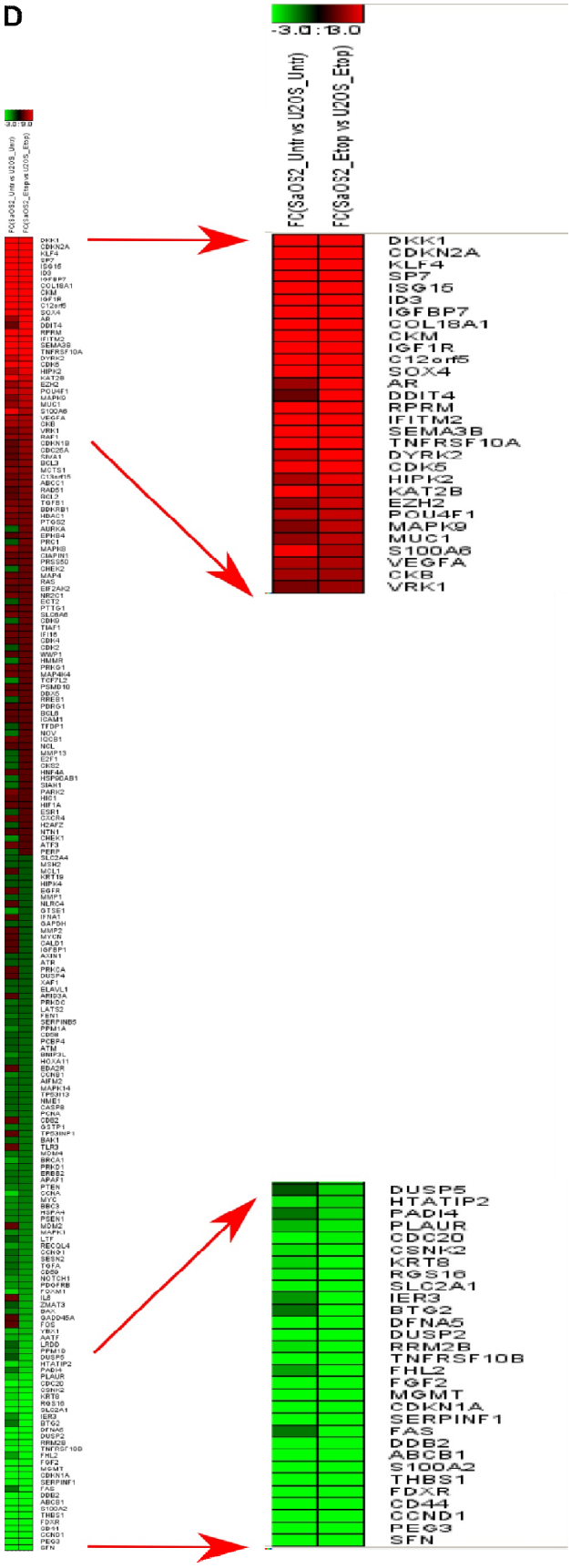

D
-3.0 : 3.0
FC(SaOS2_Untr vs U2OS_Untr)
FC(SaOS2_Etop vs U2OS_Etop)
DKK1
CDKN2A
KLF4
SP7
ISG15
ID3
IGFBP7
COL18A1
CKM
IGF1R
C12orf5
SOX4
AR
DDIT4
RPRM
IFITM2
SEMA3B
TNFRSF10A
DYRK2
CDK5
HIPK2
KAT2B
EZH2
POU4F1
MAPK9
MUC1
S100A6
VEGFA
CKB
VRK1
DUSP5
HTATIP2
PADI4
PLAUR
CDC20
CSNK2
KRT8
RGS16
SLC2A1
IER3
BTG2
DFNA5
DUSP2
RRM2B
TNFRSF10B
FHL2
FGF2
MGMT
CDKN1A
SERPINF1
FAS
DDB2
ABCB1
S100A2
THBS1
FDXR
CD44
CCND1
PEG3
SFN

Figure 6.2: Genome-wide analysis of p53 targets in U2OS and SAOS2 cells

In Figure 6.2A, fold change of U2OS cells treated with etoposide versus untreated U2OS cells (first column), and fold change of SAOS2 cells treated with etoposide versus untreated SAOS2 cells (second column). Enlargements show 30 genes with the largest positive (top) and negative (bottom) fold change. In Figure 6.2B, fold change of U2OS cells treated with etoposide versus untreated U2OS cells (first column), and fold change of SAOS2 cells treated with etoposide versus untreated SAOS2 cells (second column). Enlargements show 30 genes with the largest positive (top) and negative (bottom) fold change. In Figure 6.2C, fold change of untreated SAOS2 cells versus untreated U2OS cells (first column), and fold change of SAOS2 cells treated with etoposide versus U2OS cells treated with etoposide (second column). Enlargements show 30 genes with the largest positive (top) and negative (bottom) fold change. In Figure 6.2D, fold change of untreated SAOS2 cells versus untreated U2OS cells (first column), and fold change of SAOS2 cells treated with etoposide versus U2OS cells treated with etoposide (second column). Enlargements show 30 genes with the largest positive (top) and negative (bottom) fold change.

As it was found that there exists different biological function between the microarray experimental data and the prediction of the PKT205/G3 model (Additional Table 8), it was necessary to compare the gene expression changes in microarray experimental data with the PKT205/G3 model to evaluate the predictive strength of my model.

Our model simulation results were compared with the microarray experimental data mentioned in the beginning of this section. Predictions were classified as true, small error or large error as described in Materials and Methods (section 2.6.6).

Experiment source scenario	Experiment target scenario	Model LSSA simulation	Total number of genes	Number of true predictions (percentage of true prediction)	p-value of correct prediction	Number of small error predictions (percentage of small error prediction)	p-value of total success (true prediction and small error)	Number of large error predictions (percentage of large error prediction)
U2OS cell under DNA damage	SAOS2 cell under DNA damage	P53 wt with DNA damage ON vs p53 null with DNA damage ON	200	109 (54.5%)	3.8×10^{-10}	80 (40%)	1.1×10^{-21}	11 (5.5%)
U2OS cell without DNA damage	SAOS2 cell without DNA damage	P53 wt with DNA damage OFF vs p53 null with DNA damage OFF	200	111 (55.5%)	6.3×10^{-11}	77 (38.5%)	9.0×10^{-21}	12 (6%)

U2OS cell line without DNA damage	U2OS cell line under DNA damage	P53 wt with DNA damage ON vs p53 wt with DNA damage OFF	200	142 (71%)	1.4×10^{-27}	56 (28%)	3.0×10^{-32}	2 (1%)
SAOS2 cell line without DNA damage	SAOS2 cell line under DNA damage	P53 null with DNA damage ON vs p53 null with DNA damage OFF	200	131 (65.5%)	1.2×10^{-20}	65 (32.5%)	2.4×10^{-29}	4 (2%)
HCT116 cell p53+/+ without DNA damage	HCT116 cell p53 -/- without DNA damage	P53 null with DNA damage OFF vs p53 wt with DNA damage OFF	169	88 (52.1%)	2.2×10^{-7}	72 (42.6%)	8.5×10^{-19}	9 (5.3%)

Table 6.2: Model evaluation by logical steady state and microarray analysis by approach described in method chapter

The changes of gene expression in experimental microarray data were compared with model simulation results. The number of true predictions, small errors, large errors and their percentage were calculated and listed.

The true prediction percentage ranged from 52% to 71% and the large error predictions occupied less than 6% of the total (Table 6.2). These notable results revealed the strong predictive strength of my model and provided positive feedback to test our hypothesis that a Boolean model is an efficient tool to investigate p53 pathways induced by DNA damage will contribute to knowledge. It was found that the comparison between experimental microarray data in different cell types and model simulations produced different evaluation results. Although there may be noise affecting the microarray experiments under different conditions, one possibility is that some of those 202 genes which interact with p53 in the PKT205/G3 model may be cell specific.

For further exploration, we classified the number of pro-apoptotic and anti apoptotic genes verified in those microarray data with the different change trends (Table 6.3).

Pro-apoptotic and anti-apoptotic genes from the PKT205 model were identified in microarray analysis of U2OS (p53 positive), SAOS2 (p53 negative) human osteosarcoma cell line (200 genes from the model were analysed), and HCT 116 (p53 positive and negative) human colon cancer cell lines (169 genes from the model were analysed). The changes of these genes between different samples were classified into three categories: gene expression level

up regulated, gene expression level unchanged, gene expression level down regulated. It was found that most of these genes fall in to the category of gene expression unchanged. Then the behaviour of 77 genes that regulate apoptosis in the PKT205/G3 model among those genes selected from the experimental data was investigated. Comparing microarray data for U2OS cells with DNA damage to SAOS2 cells with DNA damage, eight pro-apoptotic genes (FGF2, CD44, DUSP2, FDXR, PEG3, DFNA5, FAS, and TNFRSF10B) were found to be down regulated and four anti-apoptotic genes (IGF1R, DDIT4, AR, and C12orf5) were found to be up regulated. Activating these eight pro-apoptotic genes and inhibiting of those four anti-apoptotic genes may be targets of anti cancer agents to promote tumour cell death in p53 mutants. Comparing microarray data for SAOS2 cells without DNA damage with SAOS2 cells with DNA damage, 11 pro-apoptotic genes (BAX, TLR3, CDC25A, DDIT4, SIVA1, COL18A1, BBC3, ATF3, NOTCH1, DKK1, and AR) were found to be up regulated and 7 anti-apoptotic genes (DUSP4, IL6, PDGFRB, CD44, PRKCA, TGFA and VEGFA) were found to be down regulated. Those genes may be utilized as potential markers to monitor p53 mutant tumour cell death for cancer patients with UV or IR treatment. Remarkably, the growth factors and receptors FGF2 and IGF1R were identified as common factors, and PDGFR and TGFA as specific factors, contributing to U2OS human osteosarcoma and HCT116 colon cancer cells growth, respectively (Table 6.3). For example, IGF1R is an anti-apoptotic gene up regulated in SAOS2 cells when compared to U2OS cells, whereas FGF2 which can be both pro and anti-apoptotic is up regulated in SAOS2 cells. In HCT116 cells with mutant p53 similar to SAOS2, there is up regulation of anti-apoptotic IGF1R, but PDGFRB and TGFA (transforming growth factor, alpha) are also up regulated and FGF2 does change in these cells (Table 6.3), indicating that both general (IGF1R) and cell type specific (PDGFRB and TGFA) pathways were uncovered by the

model. These results provide a genome wide validation of the model and the first evidence of its potential use as a therapeutic tool in cancer care.

A

Experiment source condition	Experiment target condition	Total number of genes	Number of up-regulated genes	Number of unchanged genes	Number of down-regulated genes
U2OS cells under DNA damage	SAOS2 cells under DNA damage	200	19 (10%)	161 (80%)	20 (10%)
U2OS cells without DNA damage	SAOS2 cells without DNA damage	200	19 (10%)	164 (81%)	17 (9%)
U2OS cells without DNA damage	U2OS cells under DNA damage	200	25 (13%)	153 (76%)	22 (11%)
SAOS2 cells without DNA damage	SAOS2 cells under DNA damage	200	27 (14%)	145 (72%)	28 (14%)
HCT116 cells p53+/+ without DNA damage	HCT116 cells p53-/- without DNA damage	169	19 (11%)	125 (74%)	25 (15%)

B

Experiment Source condition	Experiment target condition	Total number of pro-apoptotic genes identified	Number of pro-apoptotic genes up-regulated	Number of pro-apoptotic genes not changed	Number of pro-apoptotic genes down-regulated
U2OS cells under DNA damage	SAOS2 cells under DNA damage	57	9 (DDIT4; SEMA3B IFITM2 COL18A1 DKK1 AR; TNFRSF10A IGFBP7 CDKN2A)	38 (ESR1; ECT2; MSH2; DUSP4; CXCR4 XAF1; IL6; EGFR;; PRKCA; BAX; IFI16; TLR3; CDC25A; FOS; MAP4K4; SIVA1; PCBP4; HNF4A; TP53INP1; CCNG1 CASP8; CDKN1B; BBC3; APAF1 BAK1; NLRC4; ATF3; PTGS2; GADD45A NOTCH1; PTEN; AIFM2; BNIP3L; LATS2 SERPINB5; PERP; SIAH1; LRDD)	8 (FGF2; CD44; DUSP2; FDXR; PEG3; DFNA5; FAS; TNFRSF10B)

U2OS cells without DNA damage	SAOS2 cells without DNA damage	55	7 (SEMA3B; IFITM2; COL18A1; DKK1; TNFRSF10A; IGFBP7; CDKN2A)	41 (ESR1; ECT2; MSH2; DUSP4; CXCR4; XAF1; IL6; EGFR; PRKCA; BAX; IFI16; TLR3; CDC25A; FOS; DDIT4; MAP4K4; SIVA1; PCBP4; HNF4A; TP53INP1;CCNG1; CASP8; CDKN1B; BBC3;; APAF1; BAK1; NLRC4; ATF3; PTGS2; GADD45A; NOTCH1; FAS; PTEN; AIFM2; AR; BNIP3L; LATS2; SERPINB5; PERP; SIAH1; LRDD)	7 (FGF2; CD44; DUSP2; FDXR; PEG3; DFNA5; TNFRSF10B)
U2OS cells without DNA damage	U2OS cells under DNA damage	55	11 (IL6; CD44; DUSP2; FDXR; BAX; TLR3; FOS; ATF3; GADD45A; FAS; LRDD)	37 (ESR1; MSH2; DUSP4; CXCR4; XAF1; FGF2; EGFR; PRKCA; IFI16; CDC25A MAP4K4; SEMA3B; IFITM2; SIVA1 PCBP4; DFNA5; HNF4A; P53INP1 CCNG1; CASP8; COL18A1; BBC3 APAF1; BAK1; NLRC4; PTGS2 NOTCH1; PTEN; AIFM2; DKK1 BNIP3L; TNFRSF10A; TNFRSF10B IGFBP7; LATS2; SERPINB5 PERP)	7 (ECT2; DDIT4; PEG3; CDKN1B; AR; SIAH1; CDKN2A)

SAOS2 cells without DNA damage	SAOS2 cells under DNA damage	55	11 (BAX; TLR3; CDC25A; DDIT4; SIVA1; COL18A1; BBC3; ATF3; NOTCH1; DKK1; AR)	37 (ESR1; MSH2; CXCR4; XAF1; FGF2; EGFR; DUSP2; FDXR; IFI16; FOS; MAP4K4; SEMA3B; PEG3; IFITM2; PCBP4; DFNA5; HNF4A; CCNG1; CASP8; CDKN1B; APAF1; BAK1; NLRC4; PTGS2; GADD45A; FAS; PTEN; AIFM2; BNIP3L; TNFRSF10A; TNFRSF10B; IGFBP7; LATS2; SERPINB5; PERP; SIAH1; CDKN2A)	7 (ECT2; DUSP4; IL6; CD44; PRKCA; TP53INP1; LRDD)
HCT116 cells p53+/+ without DNA damage	HCT116 cells p53-/- without DNA damage	48	3 (IFI16; GADD45A; SERPINB5)	38 (ESR1; ECT2; MSH2; DUSP4; FGF2; EGFR; PRKCA; DUSP2; BAX; CDC25A FOS; DDIT4; MAP4K4; SEMA3B; IFITM2; PCBP4; DFNA5; HNF4A; TP53INP1; CCNG1; CASP8; CDKN1B; COL18A1; BAK1; ATF3; NOTCH1; PTEN; DKK1; AR; BNIP3L; TNFRSF10A; TNFRSF10B; IGFBP7; LATS2; PERP; SIAH1; LRDD; CDKN2A)	7 (CXCR4; IL6; CD44; FDXR; TLR3; BBC3; APAF1)

Experiment source condition	Experiment target condition	Total number of anti-apoptotic genes identified	Number of anti-apoptotic genes up-regulated	Number of anti-apoptotic genes not changed	Number of anti-apoptotic genes down-regulated
U2OS cells under DNA damage	SAOS2 cells under DNA damage	38	4 (IGF1R DDIT4 AR C12orf5)	28 (BCL3; GSTP1; ESR1; DUSP4; IL6; PDGFRB; WWP1; EGFR; PRKCA; TGFA; CDC25A; FOS; MAP4K4; EZH2; PSEN1; PRSS50; BCL2; CCNG1; TCF7L2; EPHB4; CDKN1B; ATF3; PTGS2; MCL1; IER3; NOTCH1; VEGFA CKS2)	6 (FGF2; CD44; DUSP2; FHL2; CDKN1A; SFN)

U2OS cells without DNA damage	SAOS2 cells without DNA damage	38	2 (IGF1R; C12orf5)	31 (BCL3; GSTP1; ESR1; DUSP4; IL6 PDGFRB; WWP1; EGFR; PRKCA; TGFA; CDC25A; FOS; DDIT4; MAP4K4 EZH2; PSEN1; PRSS50; BCL2; FHL2; CCNG1; TCF7L2; EPHB4; CDKN1B; ATF3; PTGS2; MCL1; IER3; NOTCH1; VEGFA; AR; CKS2)	5 (FGF2; CD44; DUSP2; CDKN1A; SFN)
U2OS cells without DNA damage	U2OS cells under DNA damage	38	9 (IL6; CD44; DUSP2; FOS; FHL2; ATF3; IER3; CDKN1A; SFN)	25 (BCL3; GSTP1; ESR1; DUSP4;PDGFRB; FGF2; WWP1; IGF1R; EGFR; PRKCA; TGFA; CDC25A; MAP4K4; EZH2; PSEN1; PRSS50; BCL2 CCNG1; EPHB4; PTGS2; MCL1 NOTCH1; VEGFA; CKS2; C12orf5)	4 (DDIT4; TCF7L2; CDKN1B; AR)

SAOS2 cells without DNA damage	SAOS2 cells under DNA damage	38	9 (IGF1R; CDC25A; DDIT4; BCL2; EPHB4; ATF3; NOTCH1; CDKN1A; AR)	22 (BCL3; GSTP1; ESR1; FGF2; WWP1; EGFR; DUSP2; FOS; MAP4K4; EZH2; PSEN1; PRSS50; FHL2; CCNG1; TCF7L2; CDKN1B; PTGS2; MCL1; IER3; CKS2; C12orf5; SFN)	7 (DUSP4; IL6; PDGFRB; CD44; PRKCA; TGFA; VEGFA)
HCT116 cells p53+/+ without DNA damage	HCT116 cells p53-/- without DNA damage	36	5 (PDGFRB; IGF1R; TGFA; SGK; IER3)	26 (GSTP1; ESR1; DUSP4; FGF2; WWP1; EGFR; PRKCA; DUSP2; CDC25A; FOS; DDIT4; MAP4K4; EZH2; PSEN1; BCL2; FHL2; CCNG1; EPHB4; CDKN1B; ATF3; MCL1; NOTCH1; AR; CKS2; C12orf5; SFN)	5 (BCL3; IL6; CD44; TCF7L2; CDKN1A)

Table 6.3: Number of pro- and anti-apoptotic genes with altered expression depending on p53 and DNA damage in human osteosarcoma and colon cancer cell lines.

Table 6.3A lists all genes from the PKT206 model were identified in microarray analysis of U2OS (p53 positive), SAOS2 (p53 negative) human osteosarcoma (200 genes from the model were analysed) and HCT116 (p53 positive and negative) human colon cancer cell lines (169 genes from the model were analysed). Table 6.3B lists pro-apoptotic genes from the PKT206 model identified in microarray analysis. Table 6.3C shows anti-apoptotic genes from the PKT206 model identified in microarray analysis. Results were obtained by comparing the steady state of the source and target scenario.

6.5 ChIP-seq result validation

In order to refine the model and increase its predictive powers we wanted to determine the differences of p53 isoforms interactions that have been phosphorylated or acetylated in response to DNA damage. Many posttranslational modifications of p53 have been described (Figure 1.7) and are a results of activation of upstream regulators and in response to cellular stress. The novel technique ChIP-seq has the capacity to differentiate between activities of different p53 isoforms. These isoforms of p53 are created by posttranslational modifications by several enzymes like kinases including ATM, ATR, DNA-PK for serine 15 (Tibbetts et al, 1999), HIPK2 (homeodomain interacting protein kinase 2) (Hofmann et al, 2002), DYRK2 (dual-specificity tyrosine-phosphorylation-regulated kinase 2) (Taira et al, 2007), AMPKα (AMP-activated protein kinase catalytic subunit α) (Okoshi et al, 2008) and MAPK14 (p38 mitogen activated protein kinase) (Perfettini et al, 2005) for serine 46. But mainly S15 is targeted by ATM/ATR kinases in response to DNA damage and S46 is involved in p53AIP1 induction of

apoptosis (Oda et al, 2000). ChIP-seq is a high-throughput sequencing technology and consists of the following steps using the SOLiD^{TM} system: at first, the DNA-binding protein is cross-linked to DNA, secondly, chromatin was isolated and DNA was sheared, then chromatin was precipitated with protein specific antibody, after that, the cross-link was reversed and the protein was digested (Tallack et al, 2010). The SOLiD fragment library was constructed and certain adapters were ligated to DNA fragment. We analysed previously published data (GSE22186) that describes how total p53 genome occupancy changes comparing to p53 phosphorylated on serine 15 and serine 46 in U2OS cells treated with etoposide (Smeenk et al, 2011). Smeenk et al (2011) treated U2OS cells by Actinomycin D and Etoposide to explore cell cycle arrest and apoptosis caused by phosphorylation of serine 15 of p53 and serine 46 of p53. According to comparison results from ChIP-seq analysis, they found that phosphorylation of serine 46 of p53 has a stronger effect on apoptosis than phosphorylation of serine 15. We compared the p53 target genes using their official gene symbol, which is a simplification of gene name description in different ChIP-sequencing samples respectively in Table 6.4. We obtained ChIP-seq results from GSE22186 in PubMed, which were reported by Smeenk et al, identified genes or proteins included in the PKT205 model (Smeenk et al, 2011).

I compared the genes reported as targets in U2OS cells exposed to certain treatments listed in Table 6.4, with 202 gene nodes in our PKT205 model. Those 177 genes found in the experiment of U2OS cell treated with Etoposide where p53 –DO1 antibody was used for immunoprecipitation are shown in Table 6.4, 13 out of 177 genes matched the name of nodes in our PKT205 model in Table 6.4 below. Comparing the 13 genes found in the experiment of U2OS cells with Etoposide treatment with 77 genes which regulate apoptosis

in the PKT205/G3 model, it was found that 9 out of those 13 matched my model prediction as regulators of apoptosis in p53 wild type cells in response to DNA damage. However, the other 4 genes below: ZMAT3, RRM2B, DDB2, and EDA2R were not found in the PKT205/G3 model. Those 4 genes may be potential apoptosis regulators predicted by the ChIP-seq experimental data.

Compared with previous results for ChIP-seq data of U2OS cells with Etoposide and p53-DO1 antibody treatment, PCNA was a new gene found to be involved in apoptosis. We can infer that PCNA is specific to the phosphorylation of serine 46 of p53 and is involved in the apoptosis activity caused by this phosphorylation.

In summary, this ChIP-seq data from literature (Smeenk et al, 2011) verified that some of the p53 target genes in the PKT205/G3 model functioned as regulators of apoptosis and new genes involved in apoptosis were revealed, which were uncovered in the PKT205/G3 model. Moreover, PCNA was found to be specific to the apoptosis caused by the phosphorylation of serine 46 of p53.

Number of significant genes in the experiment	Cell type in the experiment	Treatment utilized in the experiment	The significance criteria	Number of genes found in the PKT205 model
177	U2OS cell	Etoposide treatment and p53-DO1 antibody	Expression changed more than 1.7 fold	13 (BTG2; ATF3; TGFA; ZMAT3; TLR3; CDKN1A; TP53INP1; RRM2B; FAS; TCF7L2; DDB2; SERPINB5; EDA2R)
94	U2OS	P53-pS46 antibody	More than1.2 fold in expression between Actinomycin D and Etoposide	8 (GADD45A; TGFA; FHL2; ZMAT3; CDKN1A; DDB2; PCNA; EDA2R)

Table 6.4: P53 target gene verified by ChIP-seq data

This table shows p53 target genes verified by ChIP-seq experimental data (Smeenk et al, 2011).

6.6 Discussion

By performing *in silico* knock-out tests, 58 novel predictions were produced. Four predictions out of them were confirmed by previous literature evidence and one of them did not agree with predictions. Two predictions were verified by western blotting experiments. The other predictions were consistent with literature work or potential novel prediction. Taking into account the limitation of text mining methods from literature, these novel predictions can be used for developing novel targets in cancers carrying mutations mimicked in the *in silico* knock-out tests. Furthermore, combining multiple mutations *in silico* to match mutations found by tumour genome sequencing will provide exciting opportunities for personalizing treatment specifically for each individual patient or tumour. During the literature search, evidence for p53 knock-out test predictions were easier to find than for other gene knock-out test. This may be due to particularly strong interest in this gene and current research achievements on cancer research.

Our findings highlighted the possibility of using CHEK1 modulators as a novel cancer therapy. Since there are defects in p53 pathways of most tumour cells, the CHEK1 kinase plays an important role to mediate cell cycle arrest in those tumour cells that lost p53 function. It was found that tumour cells are deficient in the G1 checkpoint, and arrest in S and G2 check points to repair DNA damage. The S and G2 checkpoint is mediated by CDC25A (cell division cycle 25 homolog A (S. pombe)), which is a target of CHEK1; siRNA (small interfering RNA) targeting CHEK1 was able to prevent the degradation of

CDC25A and led to abrogation of the checkpoint (Chen et al, 2006). Our model suggested that upon DNA damage, ATM, ATR and CHEK1 were all up-regulated in the absence of p53, and that CHEK1 inhibits CDC25A. Those predictions from our model can better explain why CHEK1 pathway are regarded as a potential chemotherapeutics target for cancer treatments (Chen et al, 2009). Furthermore, these predictions indicate that any potential treatment should take into account whether the tumour is p53 positive or negative.

To verify our model prediction by additional method to literature survey, we turned to genome experimental data for validations. The comparison between model simulation results and microarray experimental data were designed to estimate the predictive strength of our model. As there is a difference in gene expression levels in different cell lines, a dynamic threshold which was determined by the mean value and standard deviation value of the log ratios was defined. Then the mean value and standard deviation of fold change distribution in comparisons between different scenarios was calculated. But a larger difference was found between two samples from different cell lines were compared. For instance, the mean value of the fold change of gene expression in U2OS compared to SAOS2 not treated by etoposide was 3 and the standard deviation value was 15.9. This indicated that there was a large difference between the expression levels in U2OS and SAOS2, which is the reason why distributions of expression values had to be normalised. As we found in Table 5.14, more than 50% of genes in our model simulation did not change between different scenarios. However, overall model was a success and true prediction ranged from 52% to 71% depending on the cancer cell type used for validation.

For further validation of our model structure, we utilized ChIP-sequencing data to verify the p53 target genes. According to the comparison results between different samples, the binding regions of p53 target genes differed with different antibody treatment. This finding revealed the difference in p53 isoforms and is important for the establishment of sub network model from the PKT205 model corresponding to different types of cancer cells (Figure 6.3). Figure 6.3 shows a sub network of the p53 pathway for p53 regulated cell cycle arrest in response to DNA damage (UV and IR). This model will be analysed by simulations with input signals which are predefined by random states so as to explore the dynamic activity of genes that are involved in cell cycle arrest. Considering the isoforms of p53, the PKT205 model may be improved in two different directions: the first one is to divide the node of p53 into several nodes, which correspond to the isoforms of p53. For instance, the ATM can phosphorylate p53 at serine 15 and this phosphorylation will be linked to the node representing the isoforms of p53 phosphorylated at serine 15 (Figure 6.3). This improvement may help us obtain further understanding of the mechanisms in contrast with the current situation where all interactions were linked to the same node, p53. Main point of making models specific for p53 isoforms is to increase selectivity of the flow of information from the environmental signal to the cellular process. At the moment apoptosis prediction relies on the dependency matrix calculations and number of pro-apoptotic and anti-apoptotic genes. The power of ChIP-seq superimposition to the model is that it will identify specific upstream signals (for example, facilitate development/use of chemotherapy compounds) that will selectively target kinases that phosphorylate S46 on p53 to only activate subset of genes specific for that isoforms.

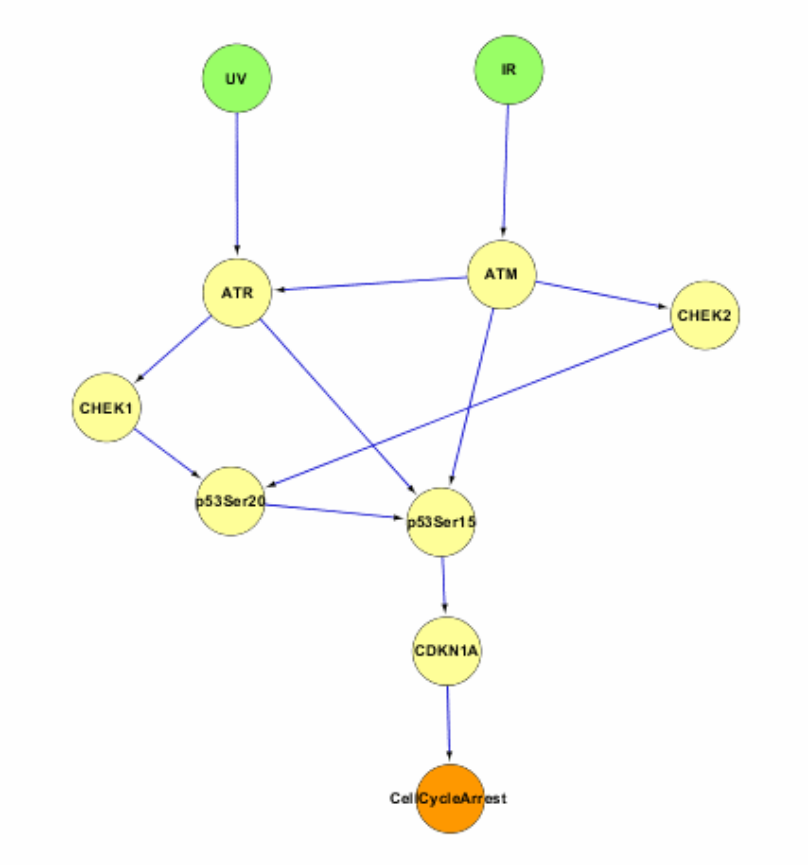


Figure 6.3: Sub network established on the basis of the PKT205/G3 model

The input nodes of this sub network were represented in green colour and the output node was marked in orange colour. The other nodes in the network were represented in yellow colour. The node"p53Ser15" represents phosphorylation of serine 15 of p53 and the node "p53Ser20" represents phosphorylation of serine 20 of p53. They are connected by activation links which are represented in blue colour.

The second improvement is to establish a sub network of the PKT205 model which is cell specific. For example, we could generate a p53 network model specific for breast cancer cells. Genes can be chosen according to clinical microarray or ChIP-seq experimental data and the interactions could be introduced into this sub network model on the basis of interactions included in

the PKT205/G3 model. This may reduce the noise caused by the complexity of the network, in which some interactions may be cell specific, and will facilitate further simulations.

Chapter 7 Discussion

7.1 Use of logical models in cancer research

The use of Boolean networks in cancer research has been reported in a few other studies. For example, Ghaffari et al (2011) designed a Boolean model of gastrointestinal cancer comprising 17 genes (Ghaffari et al, 2011), whereas Chaves et al (2009) constructed a Boolean network with 20 nodes to investigate the dynamics of the NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) pathway in controlling apoptosis (Chaves et al, 2009). Calzolari et al (2007) designed a Boolean network with 47 genes that regulate apoptosis and investigated the relationship between genes and selective control of cell populations (Calzolari et al, 2007). Zhang et al (2008) constructed a Boolean network for T cell large granular lymphocyte (T-LGL) survival (Zhang et al, 2008), which consisted of 58 nodes and 123 edges and provided an insight into the long-term survival of cytotoxic T lymphocyte (CTL) in T-LGL leukaemia. Mai et al (2009) constructed a Boolean network including 40 nodes involved in apoptotic pathways and demonstrated that apoptosis is an irreversible process (Mai & Liu, 2009). Ge et al (2009) constructed Boolean networks to investigate the dynamics of negative feedback loops of p53 pathways. They compared the dynamics of a stochastic versus deterministic model and showed that the Boolean model was able to predict the dominant process in the system (Ge & Qian, 2009).

There were two new published research articles for p53 that used network modelling approaches. The first one utilized Boolean modelling approach to investigate the p53 and NF-κB pathway in response to DNA damage (Plotz &

Naumann, 2012). This Boolean network consists of 96 nodes and 98 interactions and represents the p53 and NF-κB pathway to induce apoptosis in response to SSBs (DNA single strand breaks) and DSBs (DNA double-strand breaks) stress signals. Predictions from models with p53 presence and p53 absence were both validated by literature work and experimental results. In this report, a core regulatory network was generated by decreasing the size of the original network and simulations were performed to identify candidate target proteins for cancer therapies. In addition, simulations of certain genetic disorders were performed and predictions were produced. The 117 protein defects which contributed to carcinogenesis were explored on the basis of simulations. Comparing with this achievement, the PKT205 model has more extensive coverage of p53 pathways and includes the most complete p53 interactome to date. In addition, the PKT205/G3 model provides evidence of its potential use in individualized cancer therapy through assessment of apoptotic potential of chemotherapy in osteosarcoma and colon cancer. The second one was reported by Isik et al and they investigated cyclic pathways in which p53 was involved, by using a newly designed score flow algorithm (Isik et al, 2012). A Cytoscape plug-in was developed on the basis of this algorithm and was used to analyse 30 KEGG pathways. The initial raw data for input for network models was based on ChIP-seq experimental data and microarray experimental data, which were utilized to predefine the score of nodes. They also performed *in silico* knock-out tests for p53 and verified that apoptosis activity was prevented by the absence of p53. These simulation results indicated that the signal transduction score flow algorithm played a useful role for the improvement of drug design to promote apoptosis and disturb cell cycle in tumour cells. This Cytoscape plug-in provides an alternative way to validate my PKT205 model and is useful for further improvement and analysis, it will help me to gain more details about p53 biological activities in response to

diverse cellular stress stimuli, such as DNA damage.

The simulations using the PKT38 model indicated different effects on the outcome of cellular stress signals, which were obtained by the analysis of dependency matrix in section 3.3. Different input signals exerted different effects on the same output node. Since we established the PKT38 and PKT62 models using manual literature surveying, the model simulation results revealed that extracting interaction information from literature to construct a p53 pathway model is a promising approach to investigate dynamic mechanisms in p53 pathways induced by DNA damage, but also requires the assistance of computational tools to extract protein-protein interaction information more efficiently. This led us to creating an interface program for automated extraction of literature from STRING database that has confidence score schema using existing databases. This substantially accelerated the literature survey and generation of the final model. Analysis of this small network also indicated that the negative feedback loops play an important role in decreasing the sensitivity of the p53 network to external environmental stress signals and retain the stability of p53 network.

p53 acts as a tumour suppressor and plays a crucial role in protecting cells against cancer and genetic instability caused by DNA damage (Brown & Attardi, 2005). The loss of p53 function is common in many cancer cells, highlighting its importance for medicine. However, the vast number of reports presents a problem for selecting and extracting relevant p53 related information. In order to overcome this problem, we surveyed protein-protein interaction databases, and selected the STRING database as the main data resource to extract protein-protein interactions as it provided information about the nature of interactions (activation or inhibition) and was frequently updated.

During the manual curation of interaction records, we found six types of errors in the STRING database. Because of the variety of natural language, which makes it challenging to extract protein-protein interaction information from publications accurately and efficiently and limits the performance of text mining tools, manual curation was performed for all automatically extracted information so as to reduce the perturbations caused by text mining errors.

Since there are thousands of reported gene interactions with p53, we automatically extracted all genes interacting with p53 from the STRING database. This led to a model with more than 2000 nodes that included several layers of direct and indirect p53 interactants. According to measurements of simulation times, it was found that a large amount of feedback loops in the p53 interaction network was the major barrier which prevented CellNetAnalyzer from producing functional analysis results in a reasonable amount of time. This model was simplified by eliminating indirect interactants, and further manual curation resulted in the generation of the present PKT205 model.

Once the protein-protein interaction information relevant to p53 was successfully retrieved and curated, the PKT205/G1 model and the PKT205/G2 models were constructed. The advantage of PKT205/G1 model is that it facilitates the simulation and functional analysis by reducing the network size and provided a series of predictions resulting from *in silico* knock-out tests. However, the PKT205/G1 model lacks the communication with environmental factors, which is why it was extended into the PKT205/G2 model. This model included the interaction links with DNA damage and apoptosis. Since the STRING database had been updated in the meantime, we further updated this

model and created the PKT205/G3 model. In the analysis of these three versions of the PKT205 model,

PKT205/G3 is the final version of the PKT205 model, and was used for *in silico* knock-out tests and for the logical steady state analysis in order to predict the dynamic changes of the p53 pathways in response to particular gene depletion perturbations. Those predictions furthered our understanding of the role of p53 target genes in the apoptosis process, such as p53AIP1, which induced apoptosis (Oda et al, 2000), and FGF2 which both promotes (Kim et al, 2004) and inhibits apoptosis (Karsan et al, 1997). The major finding obtained through knock-out tests and experimental validations are discussed in the next sections.

7.2 Prediction of logical steady states in p53 knock-out test

Knock-out simulations allowed us to mimic p53 mutants potentially found in cancer and generate predictions of the effects of DNA damage on cellular fate. The percentage of change for pro-apoptotic and anti-apoptotic genes is shown in Table 5.11 and Table 5.12. These distributions of changes in Table 5.11 and Table 5.12 illustrate the probability of cell death and potential mechanisms of cancer treatment. For instance, in cells with mutant p53 not treated with chemotherapy inducing DNA damage, 29 out of 58 pro-apoptotic genes were down-regulated, 22 pro-apoptotic genes do not change and only 5 pro-apoptotic genes were up-regulated. Meanwhile, 38 out of 39 anti-apoptotic genes remained the same and only FGF2 was down-regulated. This finding illustrated that in tumour cells with p53 mutations, the probability of apoptosis process was decreased and cells survived. When cells with the mutant p53

were treated by DNA damage, only two pro-apoptotic genes, FAS and p53AIP1 became up-regulated, and led to an increase in apoptosis probability.

Analysis of the expression changes of genes that control apoptosis using steady state comparisons between different scenarios, *in silico* and in two different types of cancer cell types (Tables 5.11-5.12, Table 6.3), produced several important predictions that may have direct therapeutic implications. First, FGF2 which can both inhibit and activate apoptosis is the only factor altered in DNA damage treated cells that do not have its p53 status altered, indicating its important role in p53 mediated apoptosis and highlighting its therapeutic potential. Furthermore, this type of analysis identified seven anti-apoptotic genes that are up regulated in the p53 mutant scenario and potentially contribute to the proliferative and resistant phenotype of p53 minus tumours (Table 5.11). Therefore these genes should be targeted with inhibitors to successfully treat cancer carrying p53 mutations. On the other hand, a large number of pro-apoptotic genes are down regulated in p53 mutant cells according to the model, identifying them as potential therapeutic targets for activation (Table 5.12). Further analysis of the subset of genes relevant to the model that we found changed in microarray data revealed that the loss of p53 up regulates the IGFR1 gene in both osteosarcoma and colon cancer cell lines. Remarkably, our data identify growth factors as a major level of control of anti-apoptotic activities in p53 negative cells irrespectively of DNA damage. IGF1R is up regulated *in silico* in both SAOS2 and HCT116 p53 minus cells (Table 6.3). In addition, PDGFRB, IGFR1R and TGFA are all up regulated in HCT116 p53-/- cell lines when compared to the HCT116 p53+/+ plus cell lines (Table 6.3). Our analysis highlighted one factor (IGF1R) that is found up regulated in p53 negative cells in the model and at least two different cancer cell lines, and in addition indicated that different cell lines may have additional

growth factor combinations and dependencies, as colon cancer cells not exposed to DNA damage had up regulated PDGFR whereas SAOS2 cell did not, when compared to their p53 positive counterparts. This, together with the mentioned role of FGF2 highlights the crucial role of growth factors and their receptors as therapeutic targets in p53 negative cancer.

7.3 Experimental validations of predictions

As the system biology approach is a combination of computational and experimental methodologies, several diverse experimental approaches were used to validate the predictions generated from my model simulations. Those validations consist of literature based validations, protein measurement (western blotting) result validation, microarray experimental validation and ChIP-seq analysis validation. These approaches all have their own advantages and disadvantages.

7.3.1 Literature based validations

58 predictions were produced according to the *in silico* knock-out tests and 25 out of them were derived in the absence of p53. Combining this distribution with the survey of gene depletion effect in the PKT205/G1 model, we could conclude that the depletion of a gene or protein that has high connectivity degree affects more strongly the whole p53 network than those with low connectivity degree. The advantage of literature based validation was that it provided direct evidence from publications to validate the model predictions.

However, this literature survey is time consuming and limited by current large number of p53 related publications. For instance, in those 58 predictions, there were four confirmed predictions and one prediction was found opposite to the literature evidence; they were all relevant to the research of p53 knock-outs. However, the evidence of other gene knock-outs was still not clear. It is also difficult to validate these predictions because of the vast literature that needs to be analysed for most of these predictions, which strengthens the case for development of better text mining tools.

7.3.2 Validations by protein measurement

The extent of the knock-out effects depends on the connectivity and position of the knocked-out protein in the network. For example, ATM is upstream of p53 with a connectivity of ten, and it does not involve feedback loops. Therefore, the knock-out of ATM resulted in few changes in the dependency matrix. However, predictions of knock-out tests identified genes which have a significant effect on the whole p53 network. Some of them have been used as cancer drug targets, for example, ERBB2 (v-erb-b2, erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)), and EGFR (epidermal growth factor receptor) are targets in breast cancer treatment (Cameron & Stein, 2008). This experimental approach provided an alternative way to validate our model prediction and is not limited to previous publications. However, the limited availability of resources and time prevented us to validate all predictions by western blots. Therefore we turned to a genome-wide type of validation.

7.3.3 Validation by microarray analysis

Logical steady state analysis in the p53 knock-out indicated that negative feedback loops are crucial for the robustness of the p53 system to external perturbations. The results of logical steady state analysis indicated that state changes between different DNA damage input conditions and different p53 status could be predicted with significantly better precision than random. The correct prediction percentage was ranging between 52 % and 71 % depending on the cancer type which substantially exceeds the expected probability of 33.3 %. All the combined percentages of true predictions and small errors were over 90 %, compared to an expected probability of 66.7 % (Table 6.3). The microarray experimental data provided high through-put genome data to estimate the prediction strength of the model. But the limitations of microarray technology may affect the results as well; for example, discrepancies between model predictions and microarray data may be caused by the fact that only one probe is available for CKM (creatine kinase, muscle) in the microarray platform used (indicated in material and methods), thus suggesting the need for more a reliable estimate of the expression of this gene.

7.3.4 Validation by ChIP-seq analysis

My model was also validated by published ChIP-seq experimental data from literature (Smeenk et al, 2011). The comparison confirmed nine gene nodes in the PKT205/G3 model, which regulated apoptosis activity. Moreover, this ChIP-seq data revealed that the apoptosis regulator function of four genes was uncovered in my model and PCNA was specific to regulate apoptosis in response to the phosphorylation of serine 46 of p53. This finding demonstrated the possibility to improve the PKT205/G3 model by combining it

with high through-put experimental data, such as ChIP-seq data, in future work. For example, Nikulenkov et al performed ChIP-seq analysis on p53 induced by nutlin3a, RITA and a drug 5-fluorouracil (5-FU). The treatment of nutlin3a was known to cause growth arrest. The drug 5-FU could induce DNA damage and result in G1/S arrest. RITA (reaction of p53 and induction of tumour cell apoptosis) was able to induce apoptosis. Comparing the peak of binding sites in MCT7 cells treated by those three molecules, they found binding sets that were common under these three treatments at high confidence level and that the p53 binding activity acted as a 'p53 default program'. They found that the degradation of AURKA is activated by p53 in MCF7 (human breast carcinoma line) and colon carcinoma HCT116 cells with these three treatments. This interesting finding indicated that the level of p53 target gene, AURKA was decreased (Nikulenkov et al, 2012). But my current PKT205/G3 model only includes the interaction that AURKA inhibits p53. Combining pathways generated from the PKT205/G3 model and literature work, I inferred two negative pathways from p53 to AURKA in Figure 7.1 below. ERK2 (MAPK1) was reported to activate AURKA (Frau et al, 2010) and p53 may inhibit AURKA through MAPK1.

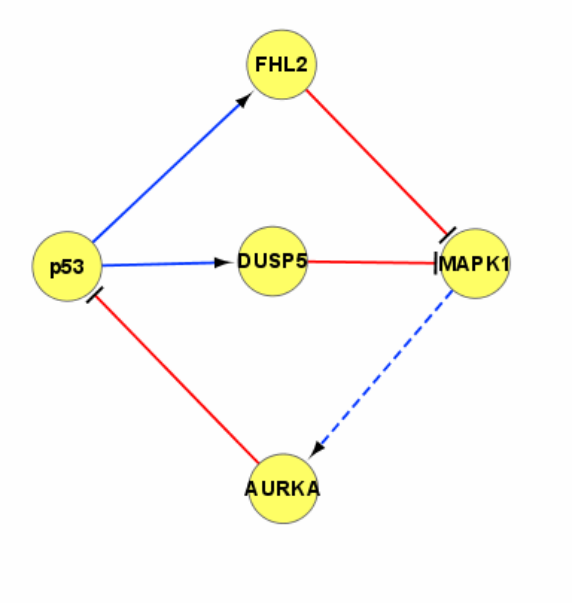


Figure 7.1: Possible p53 pathway to inhibit AURKA

This figure depicts the negative pathway from p53 to AURKA. The activation that MAPK1 activates AURKA was represented by a dashed line arrow.

7.4 Significance and future applications of the PKT205 model

In summary, I have presented a series of logical models for p53 pathways induced by DNA damage in this thesis. My analysis results highlighted the important role of p53 in these signalling pathways and provided new insight into the dynamic mechanisms of p53 pathways using model simulation combined with microarray and ChIP-seq experimental data. Predictions produced by model simulation revealed that p53 target genes such as p53 AIP1, FGF2, PDGFRB, IGFR1R and TGFA may be potential targets to promote tumour cell death in p53 mutated cells. In summary, the simulation results verified the strength of this approach to establish interaction network models for particular genes or proteins on the basis of automatic retrieval of interaction information from databases. Although challenges to describe the

cellular signalling pathways more accurately by text mining results still remain, due mainly to the limitation of text mining techniques, cancer researchers could establish protein-protein interaction network models for other genes or proteins of interest by this same approach. *In silico* knock-out tests for selected genes or interactions furthered my understanding of functions of genes which interact with p53 in response to external perturbations and led to a deeper insight into the role of feedback loops and effects exerted by those hub-nodes in the p53 pathway. As is known, computation model simulation makes it easy to perform *in silico* experiments instead of time consuming and expensive experiments performed in the lab. These *in silico* model simulations are able to predict targeted gene activity for cancer treatment. Although the high predictive strength of predictions for genome wide gene expression indicated that my PKT205 model was reliable and promising to predict genome wide gene activities, those simulation results still need to be further validated. Combining high through-put experimental data such as microarray data and ChIP-seq data with my model simulations will improve the predictive performance of the model and make it more representative of the real p53 signalling pathways. Moreover, model simulations could be utilized to reveal potential targets for cancer therapy, test whether particular drugs, which target p53 or its target genes, are effective for cancer treatment, and measure the effects of cancer treatments. Due to the complexity of cancer, this combined analysis by model simulations and high through-put experimental data could help researchers to explain the reason why these drugs mentioned above did not perform as expected by model simulations.

As we mentioned in Chapter 6, the correct prediction percentage of my model prediction ranged between 52% and 71%. Given the qualitative nature of our model, these are very promising values which show that this model is a useful

tool to predict p53 target gene expression changes under different conditions, dynamic behaviour in response to external perturbations, explore the role of two step feedback loops for the robustness of the network in response to external perturbations and gain an deeper insight into the dynamic mechanism of p53 pathways in response to DNA damage. Some negative errors are unavoidable due to the fact that a Boolean network is an approximation of the real system. It does not take into account continuous changes in gene expression levels and time delays caused by feedback loops. However, the consideration of different time scales could be included in the future and the modelling simulations could consider the time delay occurring in different pathways so as to make the model more descriptive of the real network dynamics. Nevertheless the advantage of the Boolean network approach is its completeness, since it would be unrealistic to model the exact dynamics of so many proteins using differential equations. But we can also consider adding weight values to interactions and more state values to nodes, so as to make the Boolean model more sensitive to expression levels. For instance, if ATR is activated and the inhibition pathway to repress ATR is inactivated, the state of ATR would be represented by “2” to interpret its high activated state. Otherwise, the state of activated ATR would be represented by “1” to stand for a lower level of ATR expression. This will help us to distinguish between different levels of gene expression as well as link the model to the dose of the drug or strength of DNA damage

Additionally, our model uses an interaction graph where only two genes are involved in each interaction, but some interactions may require the action of more than two genes. In the future, these factors should be considered to refine and develop enhanced versions of the PKT205 model that are based on hypergraphs and are specific for p53 post translational modification isoforms, different cell or cancer types, and other types of input and output signals. The

PKT205 model will also have to be improved by taking into account the target genes specificity in response to different environmental stress signals. In this improved model, p53 will be divided into several connected nodes to represent the different binding regions of p53. For instance, ATM will activate the node p53S15, which represents the phosphorylation of ATM at serine 15 of p53, while p300/CBP will activate the p53K382 node, which represents the acetylation of p300/CBP at lysine 382 of p53. This will further our understanding of particular pathway selectivity in response to environmental signals and reveal key mechanisms in the targeted biological process. This modelling approach will have to be supported by series of experimental approaches using ChIP-Seq and different DNA damage signals activating different pathways (for example, UV treatment will activate ATR, hydroxy urea will activate DNA-PK, c-Myc activation will affect ARF and so on.)

Several important predictions were obtained from our model, which will help us to get deeper insight into the mechanisms of p53 pathways. These findings highlighted the possibility of using CHEK1 pathway as a target for novel cancer therapy. Since there are defects in p53 pathways of most tumour cells, the CHEK1 kinase plays an important role to mediate cell cycle arrest in those tumour cells that lost p53 function. It was found that tumour cells are deficient in the G1 checkpoint, and arrest in S and G2 checkpoints to repair DNA damage. The S and G2 checkpoints are mediated by CDC25A (cell division cycle 25 homolog A), which is a target of CHEK1; siRNA (small interfering RNA) targeting CHEK1 was able to prevent the degradation of CDC25A and led to abrogation of the checkpoint (Chen et al, 2006). One of the model predictions was that upon DNA damage, ATM, ATR and CHEK1 were all up-regulated in the absence of p53, and that CHEK1 inhibits CDC25A. Those predictions from our model can better explain why CHEK1 inhibitors are

regarded as a potential chemotherapeutics target for cancer treatments (Chen et al, 2009). Furthermore, these predictions indicate that any potential treatment should take into account whether the tumour is p53 positive or negative.

Moreover, we could also improve the current model efficiency by testing the model on more cell lines in the future. All interactions relevant to p53 could be retrieved from the STRING database and other online resources to establish a "complete" network model, to include all interactions retrieved from databases, whose structure would be similar to the PKT2275 model described in Chapter 4. However, the interactions included in this PKT2275 model need to be updated because STRING was updated from version 8.3 to version 9.05. In addition, manual curation is required to confirm all interactions included. Those new added interactions in the model can be manually curated at the moment so as to provide a convincing interaction "pool" for users to choose interactions which were relevant to particular genes they are interested in, or according to individual genome-wide changes in the patient's genome/transcriptome/proteome profile. This interaction information from STRING database could be combined with information from other reliable databases, such as Gene Ontology terms, by a more advanced approach to cross check interaction massively and automatically with the assistance of certain computation tools. These confirmed interactions will help establish a candidate network, and then this candidate network will be combined with high through-put experimental data for individual tumour cells to provide potential personalised treatment strategies for individual cancer patients.

References

Adimoolam S, Ford J (2003) p53 and regulation of DNA damage recognition during nucleotide excision repair. *DNA Repair(Amst)* 2: 947-954

Ainhoa P-D, Andrey M, Natalia S (2007) Microarrays for Cancer Diagnosis and Classification. In *Microarray Technology and Cancer Gene Profiling*, Simone M (ed), 6, pp 74-85. LCC: Landes Bioscience and Spring Science + Business Media

Albeck J, Burke J, Aldridge B, Zhang M, Lauffenburger D, Sorger P (2008) Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol Cell* 30: 11-25

Anand P, Kunnumakkara A, Sundaram C, Harikumar K, Tharakan S, Lai O, Sung B, Aggarwal B (2008) Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* 25: 2097-2116

Ananiadou S, Pyysalo S, Tsujii J, Kell D (2010) Event extraction for systems biology by text mining the literature. *Trends Biotechnol* 28: 381-390

Anderson C, Appella E (2002) Signalling to the p53 tumour suppressor through pathways activated by genotoxic and non-genotoxic stresses. In *Handbook of Cell Signaling*, Bradshaw RA, Dennis E (eds), pp 237-247. New York: Academic Press

Anderson CW, Appela E. (2005) Pstranlational Modifications of p53: Uptream Signaling Pathways. In Zambetti G (ed.), *The p53 Network*, Vol. 2, pp. 95-114.

Assenov Y, Ramírez F, Schelhorn S, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* 24: 282-284

Aylon Y, Michael D, Shmueli A, Yabuta N, Nojima H, Oren M (2006) A positive feedback loop between the p53 and Lats2 tumor suppressors prevents tetraploidization. *Genes Dev* 20: 2687-2700

Banin S, Moyal L, Shieh S, Taya Y, Anderson C, Chessa L, Smorodinsky N, C P, Reiss Y, Shiloh Y, Ziv Y (1998) Enhanced Phosphorylation of p53 by ATM in Response to DNA Damage. *Science* 281: 1674-1677

Barabási A, Oltvai Z (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101-113

Bargonetti J, Manfredi J (2002) Multiple roles of the tumor suppressor p53. *Curr Opin Oncol* 14: 86-91

Bauer M, Berleant D (2012) Usability survey of biomedical question answering systems. *Hum Genomics* 6: 17

Bos J (1989) ras oncogenes in human cancer: a review. *Cancer Res* 49: 4682-4689

Bradford M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254

Brown J, Attardi L (2005) The role of apoptosis in cancer development and treatment response. *Nat Rev Cancer* 5: 231-237

Bruggeman F, Westerhoff H (2007) The nature of systems biology. TRENDS in Microbiology. *Trends Microbiol* 15: 45-50

Budhram-Mahadeo V, Moore A, Morris P, Ward T, Weber B, Sassone-Corsi P, Latchman D (2001) The closely related POU family transcription factors Brn-3a and Brn-3b are expressed in distinct cell types in the testis. *Int J Biochem Cell Biol* 33: 1027-1039

Bulavin D, Saito S, Hollander M, Sakaguchi K, Anderson C, Appella E, Fornace AJ (1999) Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *EMBO J* 18: 6845-6854

Burhans W, Weinberger M (2007) DNA replication stress, genome instability and aging. *Nucleic Acides Res* 35: 7545-7556

Calzolari D, Paternostro G, Harrington PJ, Piermarocchi C, Duxbury P (2007) Selective control of the apoptosis signaling network in heterogeneous cell populations. *PLoS One* 2: e547

Cameron D, Stein S (2008) Drug Insight: intracellular inhibitors of HER2--clinical development of lapatinib in breast cancer. *Nat Clin Pract Oncol* 5: 512-520

Carter B, Mak D, Schober W, Koller E, Pinilla C, Vassilev L, Reed J, Andreeff M (2010) Simultaneous activation of p53 and inhibition of XIAP enhance the activation of apoptosis signaling pathways in AML. *Blood* 115: 306-314

Chan H, Narita M, Lowe S, Livingston D (2005) The p400 E1A-associated protein is a novel component of the p53 --> p21 senescence pathway. *Genes Dev* 19: 196-201

Chang H, Nuyten D, Sneddon J, Hastie T, Tibshirani R, Sørlie T, Dai H, He Y, van't Veer L, Bartelink H, van de Rijn M (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 102: 3738-3743

Chaves M, Essing T, Allgöwer F (2009) Regulation of Apoptosis via the NFκB Pathway: Modeling and Analysis. In *Dynamics On and Of Complex Networks: Applications to Biology, Computer Science, and the Social Sciences*, Ganguly N, Deutsch A, Mukherjee A (eds), pp 19-33. New York: Springer-Verlag

New York, LLC

Chen C, Kennedy R, Sidi S, Look A, D'Andrea A (2009) CHK1 inhibition as a strategy for targeting Fanconi Anemia (FA) DNA repair pathway deficient tumors. *Mol Cancer* 8: 24

Chen Z, Xiao Z, Gu W, Xue J, Bui M, Kovar P, Li G, Wang G, Tao Z, Tong Y, Lin N, Sham H, Wang J, Sowin T, Rosenberg S, Zhang H (2006) Selective Chk1 inhibitors differentially sensitize p53-deficient cancer cells to cancer therapeutics. *Int J Cancer* 119: 2784-2794

Chène P (2001) The role of tetramerization in p53 function. *Oncogene* 20: 2611-2617

Chiu S, Xue L, Usuda J, Azizuddin K, Oleinick N (2003) Bax is essential for mitochondrion-mediated apoptosis but not for cell death caused by photodynamic therapy. *Br J Cancer* 89: 1590-1597

Christensen T, Oliveira A, Nielsen J (2009) Reconstruction and logical modeling of glucose repression signaling pathways in Saccharomyces cerevisiae. *BMC Syst Biol* 3: 7

Cohen A, Hersh W (2005) A survey of current work in biomedical text mining. *Brief Bioinform* 6: 57-71

de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9: 67-103

Debridge A, Valente L, Strasser A (2012) The Role of the Apoptotic Machinery in Tumor Suppression. *Cold Spring Harb Perspect Biol* 4: 1-14

Delavaine L, La Thangue N (1999) Control of E2F activity by p21Waf1/Cip1. *Oncogene* 18: 5381-5392

Dimitriadi M, Poulogiannis G, Liu L, Bäcklund L, Pearson D, Ichimura K, Collins V (2008) p53-independent mechanisms regulate the P2-MDM2 promoter in adult astrocytic tumours. . *Br J Cancer* 99: 1144-1152

Dimri G (2005) What has senescence got to do with cancer? *Cancer Cell* 7: 505-512

Donzelli M, Draetta G (2003) Regulating mammalian checkpoints through Cdc25 inactivation. *EMBO Rep* 4: 671-677

Dutta A, Ruppert J, Aster J, Winchester E (1993) Inhibition of DNA replication factor RPA by p53. *Nature* 365: 79-82

Eapen B (2007) Photosensitivity in Smith-Lemli-Opitz syndrome: a flux balance analysis of altered metabolism. *Bioinformation* 2: 78-82

el-Deiry W, Kern S, Pietenpol J, Kinzler K, Vogelstein B (1992) Definition of a consensus binding site for

p53. *Nat Genet* 1: 45-49

Emonet T, Macal C, North M, Wickersham C, Cluzel P (2005) AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics* 21: 2714-2721

Felsher D, Bishop J (1999) Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol Cell* 4: 199-207

Feng Z, Levine A (2010) The regulation of energy metabolism and the IGF-1/mTOR pathways by the p53 protein. *Trends Cell Biol* 20: 427-434

Ferbeyre G, de Stanchina E, Lin A, Querido E, McCurrach M, Hannon G, Lowe S (2002) Oncogenic ras and p53 cooperate to induce cellular senescence. *Mol Cell Biol* 22: 3497-3508

Frau M, Biasi F, Feo F, Pascale R (2010) Prognostic markers and putative therapeutic targets for hepatocellular carcinoma. *Mol Aspects Med* 31: 179-193

Gallagher S, Kefford R, Rizos H (2006) The ARF tumour suppressor. *Int J Biochem Cell Biol* 38: 1637-1641

Gardner L, Li Q, Park M, Flanagan W, Semenza G (2001) Hypoxia inhibits G1/S transition through regulation of p27 expression. *J Biol Chem* 276: 7919-7926

Ge H, Qian M (2009) Boolean network approach to negative feedback loops of the p53 pathways: synchronized dynamics and stochastic limit cycles. *J Comput Biol* 16: 119-132

Ghaffari N, Ivanov I, Qian X, Dougherty E (2011) A CoD-based stationary control policy for intervening in large gene regulatory networks. *BMC Bioinformatics* 12 S10

Giaccia A, Kastan M (1998) The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev* 12: 2973-2983

Gianchandani E, Papin J, Price N, Joyce A, Pallson B (2006) Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol* 2: e101

Goda N, Ryan H, Khadivi B, McNulty W, Rickert R, Johnson R (2003) Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia. *Mol Cell Biol* 23: 359-369

Gomez-Lazaro M, Galindo M, Concannon C, Segura M, Fernandez-Gomez F, Llecha N, Comella J, Prehn J, Jordan J (2008) 6-Hydroxydopamine activates the mitochondrial apoptosis pathway through p38 MAPK-mediated, p53-independent activation of BAX and PUMA. *J Neurochem* 104: 1599-1612

Gottifredi V, Karni-Schmidt O, Shieh S, C P (2001) p53 down-regulates CHK1 through p21 and the

retinoblastoma protein. *Mol Cell Biol* 21: 1066-1076

Gregory M, Qi Y, Hann S (2005) The ARF tumor suppressor: keeping Myc on a leash. *Cell Cycle* 4: 249-252

Guan Z, Buckman S, Miller B, Springer L, Morrison A (1998) Interleukin-1beta-induced cyclooxygenase-2 expression requires activation of both c-Jun NH2-erminal kinase and p38 MAPK signal pathways in rat renal mesangial cell. *J Biol Chem* 273: 28670-28676

Guda C, King B, Pal L, Guda P (2009) A top-down approach to infer and compare domain-domain interactions across eight model organisms. *PLoS One* 4: e5096

Gudkov A, Gurova K, Komarova E (2011) Inflammation and p53: A Tale of Two Stresses. *Genes Cancer* 2: 503-516

Gudkov AV. (2005) Therapeutic Strategies Based on Pharmacological Modulation of p53 Pathway. In Zambetti G (ed.), *The p53 tumor suppressor pathway and cancer.*, Vol. 2, pp. 225-243.

Hanahan D, Weinberg R (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646-674

Harmston N, Filsell W, Stumpf M (2010) What the papers say: Text mining for genomics and systems biology. *Hum Genomics* 5: 17-29

Hasty J, McMillen D, Isaacs F, Collins J (2001) Computational studies of gene regulatory networks: in numero molecular biology. *Nat Rev Genet* 2: 268-279

Hofmann T, Möller A, Sirma H, Zentgraf H, Taya Y, Dröge W, Will H, Schmitz M (2002) Regulation of p53 activity by its interaction with homeodomain-interacting protein kinase-2. *Nat Cell Biol* 4: 1-10

Huang dW, Sherman B, Lempicki R (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acides Res* 37: 1-13

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

Huang Y, Tan M, Gosink M, Wang K, Sun Y (2002) Histone deacetylase 5 is not a p53 target gene, but its overexpression inhibits tumor cell growth and induces apoptosis. *Cancer Res* 62: 2913-2922

Hubert A, Paris S, Priet J, Ninane N, Raes M, Michiels C (2006) Casein kinase 2 inhibition decreases hypoxia inducible factor-1 activity under hypoxia through elevated p53 protein level. *J Cell Sci* 119: 3351-3362

Hunter A, LaCasse E, Korneluk R (2007) The inhibitors of apoptosis (IAPs) as cancer targets. *Apoptosis*

12: 1543-1568

Isik Z, Ersahin T, Atalay V, Aykanat C, Cetin-Atalay R (2012) A signal transduction score flow algorithm for cyclic cellular pathway analysis, which combines transcriptome and ChIP-seq data. *Mol Biosyst* 8: 3224-3231

Jaramillo M, Tibiche C (2010) Cancer Genomics to Cancer Biology. In *Cancer Systems Biology*, Wang E (ed), 12, pp 215-232. London: CRC Press

Jemal A, Bray F, Center M, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA Cancer J Clin* 61: 69-90

Jensen L, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C (2009) STRING 8 - - a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acides Res* 37: D412-D416

Jeong H, Mason S, Barabási A, Oltvai Z (2001) Lethality and centrality in protein networks. *Nature* 411: 41-42

Joerger A, Fersht A (2010) The tumor suppressor p53: From structures to drug discovery. *Cold Spring Harb Perspect Biol* 2: a000919

Jung M, Yun J, Chae H, Kim J, Kim S, Choi T, Shin D (2001) p53 and its homologues, p63 and p73, induce a replicative senescence through inactivation of NF-Y transcription factor. *Oncogene* 20: 5818-5825

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

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acides Res* 40: D109-D114

Karsan A, Yee E, Poirier G, Zhou P, Craig R, Harlan J (1997) Fibroblast growth factor-2 inhibits endothelial cell apoptosis by Bcl-2-dependent and independent mechanisms. *Am J Pathol* 151: 1775-1784

Katayama H, Sasai K, Kawai H, Yuan Z, Bondaruk J, Suzuki F, Fujii S, Arlinghaus R, Czerniak B, Sen S (2004) Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nat Genet* 36

Ke H, Pei J, Ni Z, Xia H, Qi H, Woods T, Kelekar A, Tao W (2004) Putative tumor suppressor Lats2 induces apoptosis through downregulation of Bcl-2 and Bcl-x(L). *Exp Cell Res* 298: 329-338

Khoury M, Bourdon J (2011) p53 Isoforms: An Intracellular Microprossor? *Genes Cancer* 2: 453-465

Kidokoro T, Tanikawa C, Furukawa Y, Katagiri T, Nakamura Y, Matsuda K (2008) CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene* 27: 1562-1571

Kier L, Cheng C, Testa B, Carrupt P (1996) A cellular automata model of enzyme kinetics. *J Mol Graph* 14: 227-231

Kim M, Kim C, Jung H, Seo M, Juhnn Y, Shin H, Ahn H, Thiele C, Chi J (2004) Fibroblast growth factor 2 induces differentiation and apoptosis of Askin tumour cells. *J Pathol* 202: 103-112

Klamt S, Saez-Rodriguez J, Gilles E (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol* 1: 2

Klamt S, Saez-Rodriguez J, Lindquist J, Simeoni L, Gilles E (2006) A methodology for the structural and functional analysis of signalling and regulatory networks. *BMC Bioinformatics* 7: 56

Klamt S, von Kamp A (2009) Computing paths and cycles in biological interaction graphs. *BMC Bioinformatics* 10: 1-11

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwing R (2009) Introduction. In *Systems Bilogy: A Textbook*, 1. Germany: Wiley-VCH Verlag GmbH & Co. KGaA

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwing R (2009) Modeling of Biochemical Systems. In *Systems Bilogy: A Textbook*, 2. Germany: Wiley-VCH Verlag GmbH & Co. KGaA

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwing R (2009) Molecular Biology of Apoptosis. In *Systems Bilogy: A Textbook*. Germany: Wiley-VCH Verlag GmbH & Co. KGaA

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwing R (2009) Network Sturctures, Dynamics, and Function. In *Systems Bilogy: A Textbook*, 8. Germany: Wiley-VCH Verlag GmbH & Co. KGaA

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwing R (2009) Specific Biochemical Systems. In *Systems Bilogy: A Textbook*, 3. Germany: Wiley-VCH Verlag GmbH & Co. KGaA

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwing R (2009) Stochastic Systems and Variability. In *Systems Bilogy: A Textbook*, 7. Germany: Wiley-VCH Verlag GmbH & Co. KGaA

Komarova E, Kondratov R, Wang K, Christov K, Golovkina T, Goldblum J, Gudkov A (2004) Dual effect of p53 on radiation sensitivity in vivo: p53 promotes hematopoietic injury, but protects from gastro-intestinal syndrome in mice. *Oncogene* 23: 3265-3271

Konovalov S, Scotch M, Post L, Brandt C (2010) Biomedical informatics techniques for processing and analyzing web blogs of military service members. *J Med Internet Res* 12: e45

Koshiji M, Kageyama Y, Pete E, Horikawa I, Barrett J, Huang L (2004) HIF-1alpha induces cell cycle arrest by functionally counteracting Myc. *EMBO J* 23: 1949-1956

Krallinger M, Leitner F, Vazquez M, Salgado D, Marcelle C, Tyers M, Valencia A, Chatr-aryamontri A (2012) How to link ontologies and protein-protein interactions to literature: text-mining approaches and the BioCreative experience. *Database (Oxford)* 21

Lane D (1992) Cancer. p53, guardian of the genome. *Nature* 358: 15-16

Lane D, Benchimol S (1990) p53: oncogene or anti-oncogene? *Genes Dev* 4: 1-8

Langley E, Pearson M, Faretta M, Bauer U, Frye R, Minucci S, Pelicci P, T. K (2002) Human SIR2 deacetylates p53 and antagonizes PML/p53-induced cellular senescence. *EMBO J* 21: 2383-2396

Laurora S, Tamagno E, Briatore F, Bardini P, Pizzimenti S, Toaldo C, Reffo P, Costelli P, Dianzani M, Danni O, G B (2005) 4-Hydroxynonenal modulation of p53 family gene expression in the SK-N-BE neuroblastoma cell line. *Free Radic Biol Med* 38: 215-225

Ledgerwood E, Morison I (2009) Targeting the apoptosome for cancer therapy. *Clin Cancer Res* 15: 420-424

Lee T, Rinaldi N, Robert F, Odom D, Bar-Joseph Z, Gerber G, Hannett N, Harbison C, Thompson C, Simon I, Zeitlinger J, Jennings E, Murray H, Gordon D, Ren B, Wyrick J, Tagne J, Volkert T, Fraenkel E, Gifford D, Young R (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* 298: 799-804

Levine A (1997) p53, the cellular gatekeeper for growth and division. *Cell* 88: 323-331

Levine A, Oren M (2009) The first 30 years of p53: growing ever more complex. *Nat Rev Cancer* 9: 749-758

Levine AJ, Bargonetti J, Bond GL, Hoh J, Onel K, Overholtzer M, Stoffel A, Teresky AK, Walsh CA, Shengkan J. (2005) The p53 Network. In Zambetti G (ed.), *The p53 tumor suppressor pathway and cancer.*, Vol. 2, pp. 1-24.

Li Z, Hann S (2009) The Myc-nucleophosmin-ARF network: a complex web unveiled. *Cell Cycle* 8: 2703-2707

Lin A, Lowe S (2001) Oncogenic ras activates the ARF-p53 pathway to suppress epithelial cell transformation. . *Proc Natl Acad Sci U S A* 98: 5025-5030

Lin N, Beyer C, Gießl A, Kireva T, Scholtysek C, Uderhardt S, Munoz L, Dees C, Distler A, Wirtz S, Krönke G, Spencer B, Distler O, Schett G, Distler J (2012) Autophagy regulates TNFα-mediated joint destruction in experimental arthritis. *Ann Rheum Dis* [Epub ahead of print]

Linke S, Clarkin K, Di Leonardo A, Tsou A, Wahl G (1996) A reversible, p53-dependent G0/G1 cell cycle arrest induced by ribonucleotide depletion in the absence of detectable DNA damage. *Genes Dev* 10: 934-947

Liu D, Xu Y (2011) p53, Oxidative Stress, and Aging. *Antioxid Redox Signal* 15: 1669-1678

Liu K, Bellam N, Lin H, Wang B, Stockard C, Grizzle W, Lin W (2009) Regulation of p53 by TopBP1: a potential mechanism for p53 in cancer. *Mol Cell Biol* 29: 2673-2693

Liu L, Scolnick D, Trievel R, Zhang H, Marmorstein R, Halazonetis T, Berger S (1999) p53 Sites Acetylated In Vitro by PCAF and p300 Are Acetylated In Vivo in Response to DNA Damage. *Mol Cell Biol* 19: 1202-1209

Liu M, Li C, Pazgier M, Li C, Mao Y, Lv Y, Gu B, Wei G, Yuan W, Zhan C, Lu W, Lu W (2010) D-peptide inhibitors of the p53-MDM2 interaction for targeted molecular therapy of malignant neoplasms. *Proc Natl Acad Sci U S A* 107: 14321-14326

Lu X (2010) Tied up in loops: positive and negative autoregulation of p53. *Cold Spring Harb Perspect Biol* 2: a000984

Macip S, Kosoy A, Lee S, O'Connell M, Aaronson S (2006) Oxidative stress induces a prolonged but reversible arrest in p53-null cancer cells, involving a Chk1-dependent G2 checkpoint. *Oncogene* 25: 6037-6047

Mai Z, Liu H (2009) Boolean network-based analysis of the apoptosis network: irreversible apoptosis and stable surviving. *J Theor Biol* 259: 760-769

Manna S, Gangadharan C, Edupalli D, Raviprakash N, Navneetha T, Mahali S, Thoh M (2011) Ras puts the brake on doxorubicin-mediated cell death in p53-expressing cells. *J Biol Chem* 286: 7339-7347

Marciniak-Czochra A, Kimmel A (2007) Reaction-diffusion approach to modeling of the spread of early tumors along linear or tubular structures. *J Theor Biol* 244: 375-387

Marcotte E, Xenarios I, Eisenberg D (2001) Mining literature for protein-protein interactions. *Bioinformatics* 17: 359-263

Martinez J, Georgoff I, Martinez J, Levine A (1991) Cellular Localization and cell cycle regulation by a temperature sensitive p53 protein. *Genes Dev* 5: 151-159

Materi W, Wishart D (2007) Computational systems biology in cancer: modeling methods and applications. *Gene Regul Syst Bio* 17: 91-110

McConnell B, Gregory F, Stott F, Hara E, Peters G (1999) Induced expression of p16(INK4a) inhibits both CDK4- and CDK2-associated kinase activity by reassortment of cyclin-CDK-inhibitor complexes. *Mol Cell Biol* 19: 1981-1989

Melino G, Bermassola F, Ranalli M, Yee K, Zong W, Corazzari M, Knight R, Green D, Thompson C, Vousden K (2004) p73 Induces apoptosis via PUMA transactivation and Bax mitochondrial translocation. *J Biol Chem* 279: 8076-8083

Mirzayans R, Andrais B, Scott A, Murray D (2012) New insights into p53 signaling and cancer cell response to DNA damage: implications for cancer therapy. *J Biol Biotechnol*

Moll U, Concin N. (2005) p53 in Human Cancer –Somatic and Inherited Mutations and Mutation-independent Mechanisms. In Zambetti G (ed.), *The p53 tumor suppressor pathway and cancer*, Vol. 2, pp. 115-154.

Momand J, Aspuria PJ, Furuta S. (2005) MDM2 and MDMX Regulators of p53 Activity. In Zambetti G (ed.), *The p53 tumor suppressor pathway and cancer.*, Vol. 2.

Mundt H, Stremmel W, Melino G, Krammer P, Schilling T, M M (2010) Dominant negative (DeltaN) p63alpha induces drug resistance in hepatocellular carcinoma by interference with apoptosis signaling pathways. *Biochem Biophys Res Commun* 396: 335-341

Murray-Zmijewski F, Slee E, Lu X (2008) A complex barcode underlies the heterogeneous response of p53 to stress. *Nat Rev Mol Cell Biol* 9: 702-712

Nadkarni P, Ohno-Machado L, Chapman W (2011) Natural language processing: an introduction. *J Am Med Inform Assoc* 18: 544-551

Nikulenkov F, Spinnler C, Li H, Tonelli C, Shi Y, Turunen M, Kivioja T, Ignatiev I, Kel A, Taipale J, Selivanova G (2012) Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. *Cell Death Differ* 19: 1992-2002

Nita M, Alves V, Carriho F, Ono-Nita S, Mello E, Gama-Rodrigues J (2002) Molecular aspects of hepatic carcinogenesis. *Rev Inst Med Trop Sao Paulo* 44: 39-48

Oda K, Arakawa H, Tanaka T, Matsuda K, Tanikawa C, Mori T, Nishimori H, Tamai K, Tokino T, Nakamura Y, Taya Y (2000) p53AIP1, a potential mediator of p53-dependent apoptosis,and its regulation by Ser-46-phosphorylated p53. *Cell* 102: 849-862

Okoshi R, Ozaki T, Yamamoto H, Ando K, Koida N, Ono S, Koda T, Kamijo T, Nakagawara A, Kizaki H (2008) Activation of AMP-activated protein kinase induces p53-dependent apoptotic cell death in response to energetic stress. *J Biol Chem* 283: 3979-3987

Osborne J, Walter A, Kershaw S, Mirams G, Fletcher A, Pathmanathan P, Gavaghan D, Jensen O, Maini P, Byrne H (2010) A hybrid approach to multi-scale modelling of cancer. *Philos Transact A Math Phys Eng Sci* 368: 5013-5028

Pecorino L (2005) DNA structure and stability:mutations vs. repair. In *Molecular Biology of Cancer, Mechanisms, Targets, and Therapeutics.*, 2, pp 36-39. Oxford University Press.

Pecorino L (2005) What is cancer? In *Molecular Biology of Cancer, Mechanisms, Targets, and Therapeutics.*, 1, pp 1-19. Oxford University Press.

Peng Z (2005) Current status of gendicine in China: recombinant human Ad-p53 agent for treatment of cancers. *Hum Gene Ther* 16: 1016-1027

Perfettini J, Castedo M, Nardacci R, Ciccosanti F, Boya P, Roumier T, Larochette N, Piacentini M, Kroemer G (2005) Essential role of p53 phosphorylation by p38 MAPK in apoptosis induction by the HIV-1 envelope. *J Exp Med* 201: 279-289

Plotz R, Naumann M (2012) Dynamics of p53 and NF-κB regulation in response to DNA damage and identification of target proteins suitable for therapeutic intervention. *BMC Syst Biol* 6: 125

Preto A, Singhrao S, Haughton M, Kipling D, Wynford-Thomas D, Jones C (2004) Telomere erosion triggers growth arrest but not cell death in human cancer cells retaining wild-type p53: implications for antitelomerase therapy. *Oncogene* 23: 4136-4145

Prives C, Hall P (1999) The p53 pathway. *J Pathol* 187: 112-126

Regev A, Silverman W, Shapiro E (2001) Representation and simulation of biochemical processes using the pi-calculus process algebra. *Pac Symp Biocomput*: 459-470

Rehm M, Huber H, Dussmann H, Prehn J (2006) System analysis of effector caspase activation and its control by X-linked inhibitor of apoptosis protein. *EMBO J* 25: 4338-4349

Reisman D, Elkind N, Roy B, Beamon J, Rotter V (1993) c-Myc trans-activates the p53 promoter through a required downstream CACGTG motif. *Cell Growth Differ* 4: 57-65

Ren B, Yee K, Lawler J, Khosravi-Far R (2006) Regulation of tumour angiogenesis by thrombospondin-1. *Biochim Biophys Acta* 1765: 178-188

Ries S, Biederer C, Woods D, Shifman O, Shirasawa S, Sasazuki T, McMahon M, Oren M, McCormick F

(2000) Opposing Effects of Ras on p53: transcriptional activation of mdm2 and induction of p19ARF. *Cell* 103: 321-330

Rodríguez A, Sosa D, Torres L, Molina B, Frías S, Mendoza L (2012) A Boolean network model of the FA/BRCA pathway. *Bioinformatics* 28: 858-866

Roos W, Kaina B (2013) DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. *Cancer Lett* 332: 237-248

Rossiter H, Scadeng M, Tang K, Wagner P, Breen E (2008) Doxycycline treatment prevents alveolar destruction in VEGF-deficient mouse lung. *J Cell Biochem* 104: 525-535

Rousseau D, Cannella D, Boulaire J, Fltzgerald P, Fitzgerald P, Fotedar A, Fotedar R (1999) Growth inhibition by CDK-cyclin and PCNA binding domains of p21 occurs by distinct mechanismsand is regulated by ubiquitin-proteasome pathway. *Oncogene* 18: 3290-3302

Roy S, Tenniswood M (2007) Site-specific acetylation of p53 directs selective transcription complex assembly. *J Biol Chem* 282: 4765-4771

Sakaguchi K, Herrera J, Saito S, Miki T, Bustin M, YVassilev A, Anderson C, Appella E (1998) DNA damage activates p53 through a phosphorylation–acetylation cascade. *Genes Dev* 12: 2831-2841

Samuels-Lev Y, O'Connor D, Bergamaschi D, Trigiante G, Hsieh J, Zhong S, Campargue I, Naumovski L, Crook T, Lu X (2001) ASPP proteins specifically stimulate the apoptotic function of p53. *Mol Cell* 8: 781-794

Santos C, Eggle D, States D (2005) Wnt pathway curation using automated natural language processing: combining statistical methods with partial and full parse for knowledge extraction. *Bioinformatics* 21: 1653-1658

Satyanarayana A, Hilton M, Kaldis P (2008) p21 Inhibits Cdk1 in the absence of Cdk2 to maintain the G1/S phase DNA damage checkpoint. *Mol Biol Cell* 19: 65-77

Schlatter R, Schmich K, Avalos VI, Scheurich P, Sauter T, Borner C, Ederer M, Merfort I, Sawodny O (2009) ON/OFF and beyond--a boolean model of apoptosis. *PLoS Comput Biol* 5: e1000595

Schwartz J, Gaugain C, Nacher J, de Daruvar A, Kanehisa M (2007) Observing metabolic functions at the genome scale. *Genome Biol* 8: R123

Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, T. I (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504

Shieh S, Ahn J, Tamai K, Taya Y, C P (2000) The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev* 14: 289-300

Shieh S, Ikeda M, Taya Y, C P (1997) DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell* 91: 325-334

Shieh S, Ikeda M, Taya Y, Prives C (1997) DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell* 91: 325-334

Shiheido H, Takashima H, Doi N, Yanagawa H (2011) mRNA display selection of an optimized MDM2-binding peptide that potently inhibits MDM2-p53 interaction. *PLoS One* 6: e17898

Siegel A, Radulescu O, Le Borgne M, Veber P, Ouy J, S. L (2006) Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks. *Biosystems* 84: 153-174

Sigal A, Rotter V (2005) The Oncogenic Activity of p53 Mutants. In *The p53 Tumor Suppressor Pathway and Cancer.*, Zambetti G (ed), Vol. 2, 9, pp 199-223.

Smeenk L, van Heeringen S, Koeppel M, Gilbert B, Janssen-Megens E, Stunnenberg H, Lohrum M (2011) Role of p53 serine 46 in p53 target regulation. *PLoS One* 6: 1-14

Smith M, Ford J, Hollander M, Bortnick R, Amundson S, Seo Y, Deng C, Hanawalt P, Fornace AJ (2000) p53-Mediated DNA Repair Responses to UV Radiation: Studies of Mouse Cells Lacking p53, p21, and/or gadd45 Genes. *Mol Cell Biol* 20: 3705-3714

Smith M, Kumar M (2010) The "Two faces" of Tumor Suppressor p53-revisited. *Mol Cell Pharmacol* 2: 117-119

Smolinski T (2010) Computer literacy for life sciences: helping the digital-era biology undergraduates face today's research. *CBE Life Sci Educ* 9: 357-363

Smoot M, Ono K, Ruscheinski J, Wang P, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431-432

Sorger P (2005) A reductionist's systems biology: opinion. *Curr Opin Cell Biol* 17: 9-11

Soussi T (2000) The p53 tumor suppressor gene: from molecular biology to clinical investigation. *Ann N Y Acad Sci* 910: 121-137

Stapley B, Benoit G (2000) Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts. *Pac Symp Biocomput*: 529-540

Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data.

Bioinformatics 18: 207-208

Sui X, Jin L, Huang X, Geng S, He C, Hu X (2011) p53 signaling and autophagy in cancer: a revolutionary strategy could be developed for cancer treatment. *Autophagy* 7: 565-571

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen L, von MC (2011) The STRING database in 2011: functional interaction networks for proteins, globally integrated and scored. *Nucleic Acids Res* 39: D561-D568

Taira N, Nihira K, Yamaguchi T, Miki Y, Yoshida K (2007) DYRK2 is targeted to the nucleus and controls p53 via Ser46 phosphorylation in the apoptotic response to DNA damage. *Mol Cell* 25: 725-738

Taira N, Yamamoto H, Yamaguchi T, Miki Y, Yoshida K (2010) ATM augments nuclear stabilization of DYRK2 by inhibiting MDM2 in the apoptotic response to DNA damage. *J Biol Chem* 285: 4909-4919

Tallack M, Whitington T, Yuen W, Wainwright E, Keys J, Gardiner B, Nourbakhsh E, Cloonan N, Grimmond S, Bailey T, Perkins A (2010) A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* 20: 1052-1063

Tang X, Zhu Y, Han L, Kim A, Kopelovich L, Bickers D, Athar M (2007) CP-31398 restores mutant p53 tumor suppressor function and inhibits UVB-induced skin carcinogenesis in mice. *J Clin Invest* 117: 3753-3764

Taura M, Fukuda R, Suico M, Eguma A, Koga T, Shuto T, Sato T, Morino-Koga S, Kai H (2010) TLR3 induction by anticancer drugs potentiates poly I:C-induced tumor cell apoptosis. *Cancer Sci* 101: 1610-1617

Teodoro J, Evans S, Green M (2007) Inhibition of tumor angiogenesis by p53: a new role for the guardian of the genome. *J Mol Med (Berl)* 85: 1175-1186

Thessen A, Cui H, Mozzherin D (2012) Applications of natural language processing in biodiversity science. *Adv Bioinformatics* 2012: 391574

Tian K, Rajendran R, Doddananjaiah M, Krstic-Demonacos M, Schwartz J (2013) Dynamics of DNA damage induced pathways to cancer. *Plos One* 8: e72303

Tibbetts R, Brumbaugh K, Williams J, Sarkaria J, Cliby W, Shieh S, Taya Y, Prives C, RT. A (1999) A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev* 13: 152-157

Tsai R (2012) Improving Protein-Protein Interaction Pair Ranking with an Integrated Global Association Score. *IEEE/ACM Trans Comput Biol Bioinform* Epub ahead of print

Tsujii Laboratory. (2013) MEDIE.

Twycross J, Band L, Bennett M, King J, Krasnogor N (2010) Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. *BMC Syst Biol* 4: 34

U.S.National Library of Medicine. (2013) MEDLINE. National Institues of Health.

U.S.National Library of Medicine. (2013) PubMed Central Homepage., National Center for Biotechnology Information.

Unger T, Juven-Gershon T, Moallem E, Berger M, Vogt Sionov R, Lozano G, Oren M, Haupt Y (1999) Critical role for Ser20 of human p53 in the negative regulation of p53 by Mdm2. *EMBO J* 18: 1805-1814

Vogelstein B, Lane D, Levine A (2000) Surfing the p53 network. *Nature* 408: 307-310

Vousden K (2000) p53: Death Star. . *Cell* 103: 691-694

Vucic D, Fairbrother W (2007) The inhibitor of apoptosis proteins as therapeutic targets in cancer. *Clin Cancer Res* 13: 5995-6000

Wagner A, Fell D (2001) The small world inside large metabolic networks. *Proc Biol Sci* 268: 1803-1810

Wagner J, Ma L, Rice J, Hu W, Levine A, Stolovitzky G (2005) p53–Mdm2 loop controlled by a balance of itsfeedback strength and effective dampening usingATM and delayed feedback. *Syst Biol(Stevenage)* 152: 109-118

Wang E (2010) A Roadmap of Cancer Systems Biology. In *Cancer Systems Biology*, Wang E (ed), 1, pp 1-22. London: CRC PressINC, 2010

Wang Y, Tsay Y, Tan B, Lo W, Lee S (2003) Identification and Characterization of a Novel p300-mediated p53 Acetylation Site, Lysine 305. *J Biol Chem* 278: 25568-25576

Wanka C, Brucker D, Bähr O, Ronellenfitsch M, Weller M, Steinbach J, Rieger J (2012) Synthesis of cytochrome C oxidase 2: a p53-dependent metabolic regulator that promotes respiratory function and protects glioma and colon cancer cells from hypoxia-induced cell death. *Oncogene* 31: 3764-3776

Welford S, Giaccia A (2011) Hypoxia and senescence: the impact of oxygenation on tumor suppression. *Mol Cancer Res* 9: 538-544

Wilhelm E, Pellay F, Benecke A, Bell B (2008) TAF6delta controls apoptosis and gene expression in the absence of p53. *PLoS One* 3: 1-13

Wu G, Burns T, McDonald Er, Meng R, Kao G, Muschel R, Yen T, el-Deiry W (1999) Induction of the

TRAIL receptor KILLER/DR5 in p53-dependent apoptosis but not growth arrest. *Oncogene* 18: 6411-6418

Ying B, Yang T, Song X, Hu X, Fan H, Lu X, Chen L, Cheng D, Wang T, Liu D, Xu D, Wei Y, Wen F (2009) Quercetin inhibits IL-1 beta -induced ICAM-1 expression in pulmonary epithelial cell line A549 through the MAPK pathways. *Mol Biol Rep* 36: 1825-1832

Yoon H, Yang V (2004) Requirement of Krüppel-like factor 4 in preventing entry into mitosis following DNA damage. *J Biol Chem* 279: 5035-5041

Zevedei-Oancea I, Schuster S (2011) Topological analysis of metabolic networks based on petri net theory. *Stud Health Technol Inform* 162: 17-37

Zhang R, Shah M, Yang J, Nyland S, Liu X, Yun J, Albert R, Loughran TJ (2008) Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci U S A* 105: 16308-16313

Zhang X, Liu F, Wang W (2011) Two-phase dynamics of p53 in the DNA damage response. *Proc Natl Acad Sci U S A* 108: 8990-8995

Zhang X, Qin Z, Wang J (2010) The role of p53 in cell metabolism. *Acta Pharmacol Sin* 31: 1208-1212

Zhao C, Szekely L, Bao W, Selivanova G (2010) Rescue of p53 function by small-molecule RITA in cervical carcinoma by blocking E6-mediated degradation. *Cancer Res* 70: 3372-3381

Zhao H, Piwnica-Worms H (2001) ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol Cell Biol* 21: 4129-4139

Zhou D, He Y (2008) Extracting interactions between proteins from the literature. *J Biol Inform* 41: 393-407

Zhou J, Ahn J, Wilson S, Prives C (2001) A role for p53 in base excision repair. *EMBO J* 20: 914-923

Zhou R, Vitols S, Gruber A, Liliemark J, Wang Y, Liliemark E (1999) Etoposide-induced DNA strand breaks in relation to p-glycoprotein and topoisomerase II protein expression in leukaemic cells from patients with AML and CLL. *Br J Haematol* 105: 420-427

Additional Files

Additional File 1: Combined supporting information file containing Additional Tables 1-8.

This file contains Additional Tables 1-8 as support information.

Additional File 2: Java programme code

This file includes the Java programme codes mentioned in Chapter 2 and are stored in a CD.

Interface 1(P53Extraction): this programme is the Java programme interfaces to automatically import STRING records into the CellNetAnalyzer (Figure 2.4). It includes the following classes:

Idsearcher.java: this class finds all genes which interact with p53 from the protein action file downloaded from STRING database. It produces a list of ids for all genes which interact with p53;

CompleteSearcher.java: this class searches all interaction records from the protein action file, which contains the name of Ids shown in the output of Idsearcher.java;

getActivation.java: this class filters all interactions whose interaction type (the value of “action” in Table 2.3) was “activation” or “inhibition”;

getActivationDirection.java: this class filtered all duplicate activation or inhibition records according to their direction (the value of “a_is_acting” in

Table 2.3);

getPtmod.java: this class filters all posttranslational modifications whose interaction type (the value of “mode” in Table 2.3) was “ptmod”;

getPtmodDirection.java: this class filters all duplicate posttranslational modification records according to their direction (the value of “a_is_acting” in Table 2.3);

getHighScore.java: this class classifies all remained activation, inhibition and posttranslational modification records in to two different groups: one lists records at high confidence score (≥700) and the other lists records at low confidence score (<700);

setMetabolite.java: This class translates all gene names into a text file which can be imported into the node input file for CellNetAnalyzer;

setReaction.java: This class translates all interaction names into a text file which can be imported into the edge input file for CellNetAnalyzer.

Interface 2(comparingmatrix): this programme is the Java programme interfaces to automatically compare dependency matrixes obtained by knock-out tests (Figure 2.5) and contains following classes:

dmsearcher.java: this class retrieves information of effect cells in the dependency matrix;

getid.java: this class returns a pair of gene names which are corresponding to

a dependency effect in the model;

getnewposition.java: this class finds the corresponding effect cells in the dependency matrix of the p53 wild type;

comparingmatrix.java: this class compares the effect cells between two node in those two dependency matrix;

getresult.java: this class merges previous results and lists both types of effect cells in the dependency matrix of the knock-out test and the p53 wild type model. Each line in the results corresponds to a dependency relationship between two genes;

finddifference.java: this class lists effects changes of dependency matrix from the p53 wild type model to the knock out tests. They are classified into 6 groups according to the type of effects in the dependency matrix of knock-out test.

Interface 3(microarrayAnalysis): this programme is the Java programme interface to automatically validate the model prediction using microarray data as a source. It contains flowing classes:

getlss.java: this class translates logical steady state analysis results produced by CellNetAnalyzer to a text table for comparison;

getYmod.java: this class compares gene states in two different scenario simulations and lists their change trend(up-regulated, no change or down-regulated);

Evaluation.java: this class compared the difference of gene state change trend between model simulation (Ymod.text) and microarray experimental data (ExpOutput.text) obtained from Additional file 4. The number of true predictions, predictions with small error and predictions with large errors are calculated.

Additional File 3: Interaction data for the PKT1377 model and the PKT2275 model

All interaction data retrieved for the PKT1377 model and the PKT2275 model from the STRING database are included in this file. They are stored in the CD due to large size.

Additional File 4: Gene expression data.

This file contains all microarray experimental data for validation in Chapter 6. The median values of gene expression levels in microarray experimental data are listed: the data for U2OS and SAOS2 cells are on Sheet 1 and the data for HCT116 cells on Sheet 2. Fold changes for different comparisons and their $\log_{10}$ are given. They are also stored in the CD.

Additional File 5: Kun Tian's article published for new results.

This paper (Tian et al, 2013) described a new model PKT206 model, which was extended from the PKT205/G3 model and further analysis results based on previous finding. The contributions made by authors were clearly explained in the end.