INTEGRATED MOLECULAR PROFILING FOR ANALYZING AND PREDICTING THERAPEUTIC MECHANISM, RESPONSE, BIOMARKER AND TARGET

Jia Jia

(B. Sci & M. Sci, Zhejiang University)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHARMACY

NATIONAL UNIVERSITY OF SINGAPORE

2010

ACKNOWLEDGEMENTS

I would like to deeply thank Professor Chen Yu Zong, for his constant encouragement and advice during the entire period of my postgraduate studies. In particular, he has guided me to make my research applicable to the real world problem. This work would not have been possible without his kindness in supporting me to shape up the manuscript for publication. I am also tremendously benefited from his profound knowledge, expertise in scientific research, as well as his enormous support, which will inspire and motivate me to go further in my future professional career.

I am also grateful to our BIDD group members for their insight suggestions and collaborations in my research work: Dr. Tang Zhiqun, Ms. Ma Xiaohua, Mr. Zhu Feng, Ms. Liu Xin, Ms. Shi Zhe, Dr. Cui Juan, Mr. Tu Weimin, Dr. Zhang Hailei, Dr. Lin Honghuang, Dr. Liu Xianghui, Dr. Pankaj Kumar, Dr Yap Chun wei, Ms. Wei Xiaona, Ms. Huang Lu, Mr. Zhang Jinxian, Mr. Han Bucong, Mr. Tao Lin, Dr. Wang Rong, Dr. Yan Kun. I thank them for their valuable support and encouragement in my work.

Finally, I owe my gratitude to my parents for their forever love, constant support, understanding, encouragement and strength throughout my life.

A special appreciation goes to all for love and support.

Jia Jia

August 2010

TABLE OF CONTENTS

SUMMARY

Tailored cancer diagnosis and treatment has been challenged over a century. The opportunities as well as the challenges facing disease “omics” are formidable. Taken microarray technology as an example, the invention of this technology and its ability to simultaneously interrogate thousands of genes has significantly changed the paradigm of molecular targeting and classification of human cancers as well as shifting clinical prognosis models to a broader prospect. Curretnly, much effort has been done for disease mechanism, treatment development and biomarker identification from the perspective of different molecular profiling. Particularly promising areas of research include: the identification of new targets for therapeutics and the potential for accelerating drug development through more effective strategies to evaluate therapeutic effect and toxicity; the development of novel biomarkers for disease classification and outcome prediction; and delineation of altered gene expression relevant to the course of disease.

The main objective of this dissertation is to investigate the therapeutic mechanism, drug efficacy, novel therapeutic targets and biomarkers for cancer prevention and treatment by collectively considering the mutational, amplification and microarray gene expression profiles. A collected tumor specific antigens prediction approach and a cancer marker discovery system have been further developed from microarray data for tumor marker and specific antigens prediction.

Combination therapies are now standard in therapeutic areas of multifactorial disease, such as cancer, diabetes and infectious disease, which have been proved to enhance the efficacy of agents that were initially developed as single-target drugs and reduce counter-

target activities and toxicities in cancer treatment. Unfortunately, the standard approach of combining monotherapies at the clinical stage limits the number of drug pairs that can be tested and bypasses the opportunity to find therapeutically relevant interactions between novel targets. It is essential to understand the molecular mechanisms underlying the successful drug combinations. The knowledge could facilitate the discovery of novel efficacious combinations and multi-targeted agents. In this study, we describe an extensive investigation of the published literatures on drug combinations for which the combination effect has been evaluated by rigorous analysis methods and for which relevant molecular interaction profiles of the drugs involved are available. Analysis of the 117 identified cases reveals general and specific modes of action of rationale combined drug, and highlights the potential value of molecular interaction profiles in the discovery of novel multicomponent therapies.

In the second study, a particular focus has been given to investigate the correlation between the integrative molecular profiles and the sensitivities of anticancer receptor tyrosine kinase inhibitor drugs (TKIs) in lung and breast cancer cell-lines. TKIs elicit markedly different response rates in clinical setting. These rates have nowadays been linked to mutation and amplification of drug target, activating-mutation of downstream signaling genes RAS, BRAF and PIK3CA, and loss-of-function of signaling regulator PTEN. Compensatory, alternative and redundant signaling that bypass target inhibition also influence drug response. Unlike the traditional cancer diagnostic and prognostic indices, which may group molecularly distinct patients into similar clinical classes based mainly on the morphology of diseases, we collectively considered the profiles of the bypass genes together with the profiles of the drug targets and the relevant downstream genes may improve the prediction of clinical response to the relevant drugs. We

retrospectively analyzed mutational, amplification and microarray gene expression profiles of the drug target and known bypass and downstream drug-resistant genes in non-small cell lung cancer (NSCLC) and breast cancer cell-lines sensitive or resistant to TKIs including gefitinib, erlotinib, lapatinib, and trastuzumab. Drug sensitivity prediction has been potentiated by collective analyzing mutation, amplification and expression profiles of target, bypass genes, and drug-resistant downstream signaling and regulatory genes.

Recent progresses in profiling somatic mutations and expressions of human cancer genomes, and in predicting T-cell epitopes enable genome-scale tumor-specifi antigen (TSA), a class of potential source of disease-targeting molecules, search by collectively analyzing these profiles. Such a collective approach has not been explored in spite of the availability and usage of individual methods. In this study, genome-scale TSA search was conducted by genome-scale search of tumor-specific mutations in differentially over-expressed genes of specific cancers based on tumor-specific somatic mutation, microarray gene expression data, and *in silico* T-cell recognition analysis. The performance of our method was tested against known T-cell defined melanoma and lung cancer TSAs and archieved a fairly good perdiction performance. It is suggested that noises in expression data of small sample sizes appear to be a major factor for misidentification of known TSAs. With improved data quality and analysis methods, the collective approach is potentially useful for facilitating genome-scale TSA search.

Cancer markers are useful in following the course of cancer and evaluating which therapeutic regimes are most effective for a particular type of cancer, as well as determining long-term susceptibility to cancer or recurrence. The case is clear for development of biomarkers for early detection and screening tests for diseases such as

lung, breast, colon, and ovarian cancer. In addition, diagnostic measurement of cancer disease progression is essential to successful disease management. For these reasons, development of new and effective biomarkers for cancer detection and diagnosis is crucial for efficiouse cancer prevention and treatment. In the last study, a particular focus has been given to develop marker discovery system that may benefit early disease diagnosis and correct prediction of prognosis. The expression level of such markers presents potential therapeutic drug targets and may give suggestions to proper treatment regime. We developed a novel gene selection method based on support vector machines (SVMs), recursive feature elimination (RFE), multiple random sampling strategies and multi-step evaluation of gene-ranking consistency to overcome the unstable and disease irrelevant nature of currently biomarker identification approaches. The as-developed program can be utilized to derive disease markers which present both good prediction performance and high levels of consistency with different microarray dataset combinations.

The biomarker discovery system has been test in lung adenocarcinoma for survival marker discovery by using an 86-sample lung adenocarcinoma dataset. In this case, 21 lung adenocarcinoma survival markers were identified with a farily stability across 10 sampleing-set, suggesting the effectiveness of our system on deriving stable disease markers and discovering therapeutic target.

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

ALL: acute lymphoblastic leukemia
AML: acute myeloid leukemia
ANN: artificial neural networks
cDNA: complementary DNA
CTLs: cytotoxic T lymphocytes
DNA: deoxyribonucleic acid
EGFR-I: EGFR inhibitors
EST: expressed sequence tag
FDA: food and drug administration
FN: false negative
FP: false positive
HER2-I: HER2 inhibitors
HLA: human leukocyte antigen
IG: immunoglobulin
KEGG: Kyoto encyclopedia of genes and genomes database
KNN: k-nearest neighbors
LS: least square method
MHC: major histocompatibility complex
MI: molecular interaction
MIAME: minimum information about a microarray experiment
ML: machine learning
NCBI: national center for biotechnology information
NSCLC: non-small cell lung cancer
NPV: negative predictive value
NSP: the number of non-survivable patients
PCA: principal component analysis
PDB: protein databank
Pfam: protein family
PNN: probabilistic neural networks
PPV: positive predictive value
Q: overall accuracy
RBF: radian basis kernel
RFE: recursive feature elimination
RNA: ribonucleic acid
RTKs: receptor tyrosine kinases
SAGE: serial analysis of gene expression
SCLC: small cell lung cancer
SE: sensitivity
SMD: Stanford Microarray Database
SMO: sequential minimal optimization
SP: specificity
SP: the number of survivable patients
SQL: structured query language
STDEV: standard deviation
SV: support vector
SVM: support vector machines
TCR: T-cell receptor

TN:	true negative
TP:	true positive
TTD:	therapeutic target database

1 Introduction

Global gene-expression profiling using microarray technologies has improved our understanding of the histological heterogeneity of cancer and has been increasingly used to discover potential biomarkers for patient classification and promising targets for disease prevention and treatment. However, gene expression profile alone may not reflect the full story of the cancer due to cancer result from a variety of factors, including individual variations in patients and somatic cell genetic differences in tumors, even those from the same tissue of origin. The performance and general applicability of published biomarkers are highly unstable and have difficulties in generalized cancer discrimination because of small numbers of subjects examined and inclusion of heterogeneous tumor types. Anti-cancer agents developed to direct at a single altered target frequently show reduced efficacies and poor safety and resistance issues. To address these concerns and to facilitate a stable marker-selection method, collective approaches have been used for analyzing and predicting the drug mechanism, response, novel targets and biomarkers in this study. With a focus of efficacious targeted cancer therapy and disease classification by molecular indicators, the first section of this chapter introduced the increasing role of molecular targeted therapies as well as combination strategies in cancer prevention and treatment (Section 1.1). The second section (Section 1.2) of this chapter gives an overview of therapeutic molecules, particularly tumor-specific antigen, and cancer biomarkers as well. The motivation of this study and outline of the structure of this dissertation are presented in Section 1.3.

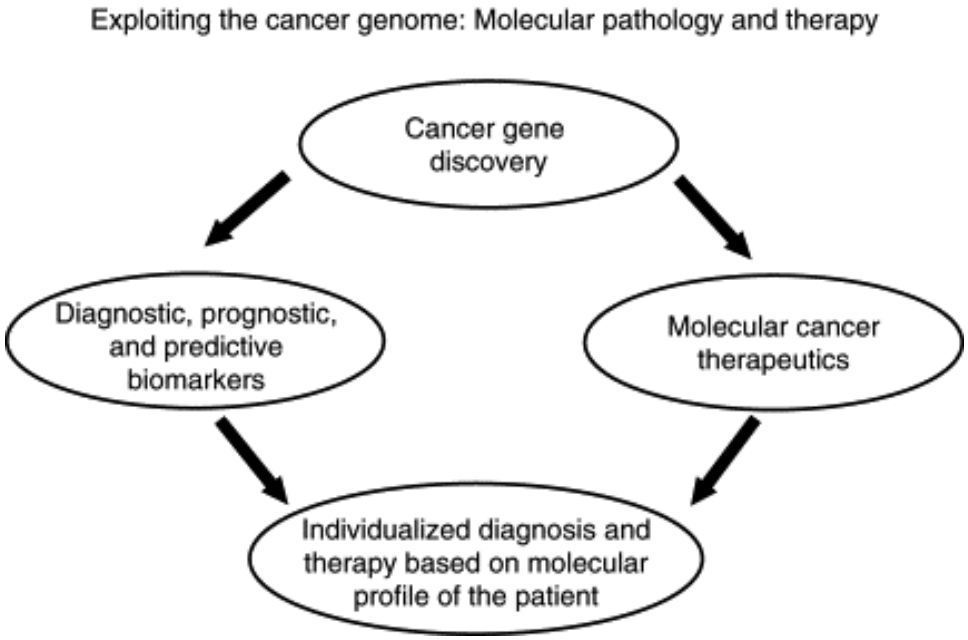
1.1 Overview of mechanism and strategies of molecular-targeted therapeutics

Knowing the origin of a disease is the first step in understanding the entire abnormal course of disease and helping the treatment of the disease. Sometimes it is very easy to determine the cause of certain diseases, such as infectious diseases which are generally caused by virus, bacteria or parasites. However, sources of some diseases may not be easily identified, especially some multifactorial diseases resulting from an accumulation of inherited and environmentally-induced changes or mutations in the genome, such as cancer [1-5], diabetes [6,7], cardiovascular disorders [8,9] and obesity [10].

For accurate disease treatment, it is very important to identify genes responsible for disease initiation, development and progress. As such, proper treatment regime can be applied and the survivability of the patients can be ultimately extended [11]. The completion of human genome project [12,13], and the new, economical, and reliable methods in functional genomics such as gene expression analysis present the potential for disease classification, target identification and drug mechanism study (Figure 1-1). Often, the potential disease targets are the molecules that show significantly different expression profiles between healthy people and patients, or among the patients with different progress stages/subtypes/outcomes, playing key roles in disease initiation [14] or disease progress [15,16]. The disease targets can be used in developing disease targeting molecules such as small molecules, antibody, and vaccines based on the protein-small molecule interaction and antibody-antigen interaction [17].

An ever-increasing number of molecular-targeted therapeutic agents are nowadays approved by US Food and Drug Administration (FDA), which have been summarized in Table 1-1 and Table 1-2. For cancer treatment, some promising targeted therapies are being studied for use alone, in combination with other targeted therapies, and in combination with other cancer treatments, such as radiotherapy and chemotherapy.

Figure 1-1 The discovery and exploitation of cancer genes has the potential to usher in a new era of individualized diagnosis and therapy.



Notes: The two critical steps in this process are: (1) the successful development of diagnostic, prognostic predictive and pharmacodynamic biomarkers; and (2) effective molecularly targeted therapeutics. The close integration of the discovery, development, and application of the molecular biomarkers and molecular therapeutics is key to future success. The figure is adapted from [18].

1.1.1 Current progress of molecular-targeted cancer therapeutics

The principle of targeted cancer therapy is certainly not new: by focusing on molecular and cellular changes that are specific to cancer, targeted cancer therapies may be more effective than other types of treatment, including chemotherapy and radiotherapy, and less

harmful to normal cells. This kind of targeted cancer therapies give clinicians a better way to tailor cancer treatment, especially when a target is present in some but not all tumors of a particular type, as is the case for HER2 in breast cancer. Eventually, treatments can be individualized based on the unique set of molecular targets produced by the patient's tumor. Moreover, targeted cancer therapies also hold the promise of being more selective for cancer cells than normal cells, thus harming fewer normal cells, reducing side effects, and improving quality of life.

The success of Imatinib (Gleevec) in the treatment of chronic myelogenous leukaemia (CML) has provided evidence that cancer can be effectively treated by the identification of underlying molecular defects. With the progress of genomics and proteomics technologies and the campaigns of studying cancer mechanism by these "omics" technology, increasing number of gene signatures closely related to cancer initiation and perpetuation have been identified, thus provide rich therapeutic candidate for targeted therapies. A new generation of drugs, that targeted on the specific molecular targets, such as receptor tyrosin kinases (RTKs) and some enzymes, are playing critical roles today in cancer treatment. The promising results in trials with RTKs targeted drugs, such as gefitinib (Iressa) and trastuzumab (Herceptin) (Table 1-3) have given encouragement results for such approaches. In addition, the identified disease genes also greatly stimulated the development of antibodies and tumor vaccines that used to activate the both arms of the immune system for cancer immunoprevention and treatment.

1.1.2 Challenges of targeted cancer therapy, receptor tyrosine kinase as a case study

The ErbB (Erythroblastic Leukemia Viral Oncogene Homolog) family, a class of transmembrane RTKs, regulates various signaling pathways that are critical in the development and progression of many cancers. It consists of four receptors (ErbB 1-4) whose ligands are Neuregulins and EGFs (Epidermal Growth Factors). Upon ligand binding, they form homo- and/or heterodimers to activate the downstream signaling pathways[19], including phosphatidylinositol-3 kinase (PI3K) and Erk mitogen-activated protein kinase (MAPK). Both of these pathways are critical to regulate cell proliferation and survival [20]. The key role of ErbB family in regulating signal transduction in the context of multiple cellular processes and environments, and the regulatory approval in clinical applications makes kinase as a readily accepted druggable protein [16]. In cancer, ErbB family members, especially EGFR and ErbB2, are frequently dysregulated through gene mutation and gene amplification, resulting in receptor overexpression. For instance, EGFR is overexpressed in almost 80% of head and neck cancer whereas overexpression of ErbB2 is found in around 30% breast cancer[19].

Over the past few years, these receptors have emerged as promising anticancer targets and good prognostic indicators. Many molecular-targeted drugs are developed to target their overexpressed proteins. Due to the specific and selective action, the therapeutic effect could be highly achieved with much less toxicity, compared to the traditional cytotoxic agents. Some successfully marketed drugs are gefitinib (Iressa) for lung cancer, Imatinib (Gleevec) for chronic myeloid leukemia and trastuzumab (Herceptin) for breast cancer. Table 1-3 summarized the clinical response rate of gefitinib, trastuzumab, and nine other kinase inhibitor drugs (cetuximab, erlotinib, canertinib, vandetanib, neratinib, lapatinib,

imatinib, nilotinib and dasatinib) in along with targeted cancer (non-small cell lung cancer (NSCLC), breast cancer, and acute lymphoblastic leukemia (ALL)), clinical setting (phase I/II, II, and III trial), and kinase target or targets of each drug directly responsible for its anticancer therapeutic efficacy. To maintain statistical significance of our analysis and in consideration of the typical sizes of the available trial data, only those trials with larger than 60 NSCLC, larger than 60 breast cancer and larger than 40 ALL patients are included. It is noticed that these drugs elicit markedly different clinical response rates and clinical response rates so far have not been as high as expected, for example, that of 15~26% ErbB2-positive patients was reported to response trastuzumab single-drug treatment (Table 1-3). Thus, agents directed at RTKs, with a few notable exceptions, such as imatinib, are frequently showed reduced efficacies, poor safety and resistant problems.

The efficacy of anti-RTK agents is affected by many factors. Diversity in patients' expression and mutation of a core set of disease-causing genes could be a reason for the failure of drug targeted only on ErbB families. Activation of parallel receptors, whose pathways could bypass ErbB signaling pathways, also contributes greatly to the resistance to ErbB-targeting drugs. Some of the common parallel receptors, which also belong to classes of RTKs, are platelet-derived growth factor receptors (PDGFR), vascular endothelial growth factor receptors (VEGFR), insulin-like growth factor receptors (IGFR), and hepatocyte growth factor receptor (HGFR or c-MET). Besides, activating mutations of downstream signaling molecules of PI3K and MAPK pathways, such as Ras, PTEN (phosphatase and tensin homolog), PIK3CA (encoding PI3K), have been highly associated with drug resistance, too. These are generally attributed to their ability to activate the signaling pathways by themselves [21-23]. Although possible reasons of some drug resistance have been postulated [24-33], practical solutions remain limited due to the lack

of insights into detailed resistance profiles as well as a systemic analysis for it. To address these questions, the drug response rate of 4 well-studied anticancer anti-RTK agents were investigated by collectively analyzing the mutation and expression profiles of molecular targets, expression of bypass genes, and activating-mutation of downstream signaling genes, including RAS, BRAF and PIK3CA, indictaed in many of RTK resistant research studies[31,34-47].

1.1.3 Systematic discovery of multicomponent therapies

The progress of systems biology has revealed that human cells and tissues are composed of complex, networked systems with redundant, convergent and divergent signaling pathways[48-51]. For example, the redundant function of proteins involved in cell-cycle regulation[52] has inspired efforts to intervene simultaneously at multiple points in these signaling pathways[53]. Drug combinations thus have been used for treating diseases and reducing suffering ever since the earliest days of recorded history. The traditional Chinese medicines are vivid examples. In some cases, single-target drugs cannot fully correct a complex disease condition such as cancer. The poor efficacy of these agents directed against individual molecular targets can be attributed to due to network robustness[24-26], redundancy[27], crosstalk[28-30], compensatory and neutralizing actions[31,32], and anti-target and counter-target activities[33]. The limitations of many monotherapies can be overcome by attacking the disease system on multiple points. Multicomponent therapeutics can be more efficacious and less vulnerable to adaptive resistance because the biological system is less able to compensate for the action of two or more drugs simultaneously. Several categories of multicomponent therapeutics have been proposed on the basis of target relationship. In the first category, the therapeutic effect occurs at separate molecular targets that can reside within individual signaling pathways, between

pathways within a cell or at separate tissues in the body. In the second category, modulation of one target facilitates action at a second target, for example by altering compound metabolism, inhibiting efflux pumps or blocking other resistance mechanisms. The third relationship is based on a coordinated action at multiple sites on a single target or macromolecular complex, which yields the improved therapeutic effect [54].

Attempts have been made during the past century to quantitatively measure the dose-effect relationships of each drug alone and its combinations. The concepts of synergy, additivism, and antagonism have been explored extensively, particularly in the fields of pharmacology and toxicology [55-57]. Traditionally, combinations can be efficiently discovered by dose–response matrix screening and systematically analyzing for drug synergies in various cell-based models of disease [58]. The cell-based phenotypic assays are employed because they maintain reasonable experimental efficiency while preserving disease-relevant molecular-pathway interactions [58]. *In vivo* screening using a whole organism model such as the zebrafish[59] could identify multi-target therapeutics that integrates their effect at the level of the organism. Often, limited combination testing samples can be used in these systems and is unlikely to have resulted in the selection of optimal combinations among the very large number of possibilities.

The deliberate mixing of drugs in a clinical setting requires an understanding of the mechanism of action of each drug class. [58]. Tremendous progress has been made over the past decade in the development and refinement of genomic and proteomic technologies and lead to uncover many novel functional pathways, therapeutic targets, and molecular interaction (MI) profiles between disease targets and therapeutic molecules. Despite of the importance of these progress towards the system-oriented drug combinations, knowledge

of disease biology, drug mechanism or intuition are still required to guide a large, agnostic surveys of molecular mechanisms that can combine to produce synergistic combination effects. Knowledge of the molecular mechanisms of currently explored multicomponent therapies is a particular crucial starting point for investigating and developing rational drug combinations and multi-targeting agents. In this study, some key characteristics of the modes of multicomponent therapies have been discussed by using the knowledge of molecular interaction profiles of individual drugs, disease network crosstalks and regulations. The identified modes of actions of drug combinations reveal some important categories of multicomponents therapeutics of current successful drug-combinations and multi-targeting agents. Rational multicomponents therapeutics thus can be realized from different perspectives, such as selectively modulating the elements of counter-target and toxicity activities[60-63].

Table 1-1 US FDA-approved molecule targeting drugs (kinase inhibitor) between 2001 to 2010 [64,65].

Year	Drugs	Pubchem ID	Drug Types	Molecular Target	Disease Indication	Therapeutic Application	Company
2010	Votrient (Pazopanib)	CID: 10113978	Tyrosine kinase inhibitor	VEGF receptors, c-kit, and PDGFR	Renal cell carcinoma	Advanced renal cell carcinoma	GlaxoSmithKline
2007	Tasigna(Nilotinib)	CID: 644241	Tyrosine kinase inhibitor	BCR-ABL	Chronic myeloid leukemia (CML)	Imatinib-resistant chronic myeloid leukemia	Novartis
2006	Sprycel (Dasatinib)	CID: 3062316	Tyrosine kinase inhibitor	BCR-ABL, SRC	Chronic myeloid leukemia (CML)	Treatment of imatinib-resistant chronic myeloid leukemia	Bristol-Myers Squibb
	Sutent (Sunitinib)	CID: 5329102	Tyrosine kinase inhibitor	PDGFR, VEGFR, KIT, FLT3, CSF-1R, RET	Kidney Cancer; Gastrointestinal Stromal Tumors	Treatment of kidney cancer and gastrointestinal stromal tumors	Pfizer
2005	Nexavar (Sorafenib)	CID: 216239	Multikinase inhibitor	VEGFR, PDGFR, c-KIT	Renal Cell Carcinoma	Treatment of Renal Cell Carcinoma	Bayer/ Onyx
2004	Tarceva (Erlotinib,OSI 774)	CID: 176870	Tyrosine kinase inhibitor	EGFR	Non-small cell lung cancer (NSCLC)	Treatment of advanced refractory metastatic non-small cell lung cancer	Genentech, OSI Pharmaceuticals
2003	Iressa (Gefitinib)	CID: 123631	Tyrosine kinase inhibitor	EGFR	Non-small cell lung cancer (NSCLC)	The second-line treatment of non-small-cell lung cancer	AstraZeneca
2002	Gleevec (Imatinib mesylate)	CID: 123596	Protein-tyrosine kinase inhibitor	PDGF, SCF, c-kit,	Positive inoperable and/or metastatic malignant gastrointestinal stromal tumors (GISTs)	Treatment of gastrointestinal stromal tumors (GISTs)	Novartis
2001	Gleevec (Imatinib mesylate)	CID: 123596	Protein-tyrosine kinase inhibitor	c-kit, PDGFR	Chronic myeloid leukemia (CML)	Oral therapy for the treatment of chronic myeloid leukemia	Novartis

Table 1-2 US FDA-approved therapeutic antibody drugs.

Year	Drugs	Target Antigen	Type of Antibody	Isotype	Kd (nM)	FDA-Approved Indication(s)	Company	Reference
2006	Vectibix (panitumumab)	EGFR	Human antibody	IgG2, kappa	0.05	Treatment of colorectal cancer	Amgen	[64,66,67]
	Herceptin* (trastuzumab)	ERBB2	Humanized antibody	IgG1	0.1	A second- or third-line therapy for patients with metastatic breast cancer	Genentech	[64,68,69]
	Lucentis (ranibizumab)	VEGF	Humanized antibody fragment	IgG1 kappa		treat the "wet" type of age-related macular degeneration (ARMD), a common form of age-related vision loss	Genentech	[64,70]
2004	Avastin (bevacizumab)	VEGF	Humanized antibody	IgG1	1.1	Treatment of metastatic carcinoma of the colon or rectum	Genentech	[64,71,72]
	Erbitux (cetuximab)	EGFR	Chimeric antibody	IgG1, kappa	0.2	Treatment of EGFR-expressing metastatic colorectal cancer	Imclone, Bristol -Myers Squibb	[64,71,72]
2002	Humira (adalimumab)	TNF-alpha	Human antibody	IgG1, kappa	0.1	For treatment of adults with rheumatoid arthritis and psoriatic arthritis.	Abbott Laboratories	[64,71,72]
1998	Remicade (infliximab)	TNF-alpha	Chimeric antibody	IgG1, kappa	0.1	For treatment of rheumatoid arthritis, Crohn's disease, ankylosing spondylitis, psoriatic arthritis, and ulcerative colitis.	Johnson & Johnson	[64,71,72]
	Herceptin (trastuzumab)	HER2 protein	Humanized antibody	IgG1, kappa	5	For treatment of metastatic breast cancer	Genentech and Roche	[64,71,72]

*First approved October 1998, used extended 2006

Table 1-3 Clinical response rates, targeted cancer in clinical test, clinical test setting, and kinase target / targets of 11 anticancer kinase inhibitor drugs approved or in clinical trial for the treatment of non–small cell lung cancer (NSCLC), breast cancer, and acute lymphoblastic leukemia (ALL)

Drug	**Kinase Target / Targets Directly Responsible for Anticancer Effect (Potency) [Reference in Pubmed ID or as Specified]**	**Targeted Cancer in Clinical Test**	**Clinical Test Setting**			**Clinical Response Rate**	**Reference of Clinical Test in Pubmed ID or as Specified**
			Test Platform	No of Patients	Patient Status		
Cetuximab	EGFR (Kd 0.39nM) [19216623]	NSCLC	Phase II	60	Not found	3.3%	16472704
		NSCLC	Phase II	66	2nd line, recurrent or progressive NSCLC received chemotherapy	4.5%	17114658
Gefitinib	EGFR (IC50 33nM), HER4 (IC50 476nM) [9578319, 12384534, 18089823]	NSCLC	Phase II	210	Advanced NSCLC received 1 or 2 chemotherapy, at least one with platinum	18.4%	12748244
		NSCLC	Phase II	216	Locally advanced or metastatic NSCLC failed 2 or more chemotherapy with platinum, docetaxel	8%	Proc Am Soc Clin Oncol 21: 2002 (abstr 1166)
		NSCLC	Phase II	100	Chemotherapy naïve stage IIIB/IV NSCLC unfit for chemotherapy	6%	19289623
		Breast Cancer	Phase II	63	Pretreated metastatic breast cancer (HER2+ subgroup)	1.6%	16947082
Erlotinib	EGFR (IC50 1nM), HER4 (IC50 230-790nM)[15711537, 18183025, 18089823]	NSCLC	Phase II	1,949	2nd/3rd-line advanced NSCLC	12%	J Clin Oncol 27: 15s, 2009 (suppl abstr 8001)
		NSCLC	Phase III	731	Stage IIIB or IV NSCLC received 1 or 2 chemotherapy	8.9%	16014882
		NSCLC	Phase II	66	Chemotherapy-naïve NSCLC aged ≥70	10%	17228019
Canertinib	EGFR (IC50 0.8nM), HER2 (IC50 19nM), HER4 (IC50 4~7nM) [12138393]	NSCLC	Phase II	166	Advanced-stage NSCLC failed or refractory to platinum-based chemotherapy	4%	17761977
		Breast Cancer	Phase II	194	Progressive or recurrent metastatic breast cancer expressed 1 or more ErbB	7.3%	19294387
Vandetanib	EGFR (IC50 500nM), VEGFR2 (IC50 40nM) , RET (IC50 100nM), HER4 (IC50 480nM), SRC (IC50 70nM) [12183421, 12499271, 18183025]	NSCLC	Phase II	83	Locally advanced or metastatic stage IIIB/IV NSCLC failed platinum-based chemotherapy	8%	J Clin Oncol 24, No. 18S (Suppl, 2006: 7000
		NSCLC	Phase II	73	NSCLC histologies and pretreated CNS metastases	7%	18936474

Trastuzumab	HER2 (Kd 0.1nM) [19216623]	Breast Cancer	Phase II	222	Pretreated HER2+ metastatic breast cancer	15%	19707416
		Breast Cancer	Phase II	114	Untreated HER2+ metastatic breast cancer	26%	19707416
Neratinib	HER2 (IC50 59nM),EGFR (IC50 92nM) [19780706]	Breast Cancer	Phase II	124	HER2+ metastatic breast cancer received trastuzumab	26%	19733440
		Breast Cancer	Phase II	124	Trastuzumab-naïve HER2+ metastatic breast cancer	55%	19733440
Lapatinib	HER2 (IC50 7nM), EGFR (IC50 0.6-8.9nM), HER4 (IC50 54nM) [18183025]	Breast Cancer	Phase II	138	First-line ErbB2-amplified locally advanced or metastatic breast cancer	24%	18458039
		Breast Cancer	Phase II	78	Stage IIIB/IV HER2+ advanced or metastatic breast cancer received trastuzumab	13%	19179558
		Breast Cancer	Phase II	126	Relapsed or refractory HER2+ inflammatory breast cancer	39%	19394894
		Breast Cancer	Phase II	141	First-line HER2+ metastatic breast cancer	22.2%	19525314
Imatinib	ABL1 (IC50 12nM), c-KIT (IC50 14nM), PDGFRB (IC50 14nM), SRC (IC50 352nM) [18183025, 19890374, 16105974]	ALL	Phase II	48	Ph+ALL without response or relapse to chemotherapy or SCT	19%	12200353
		ALL	Phase II	56	Recurrent and refractory Ph+ ALL aged>18	19%	12200353
Nilotinib	ABL1 (IC50 56nM), PDGFR (IC50 22nM), c-KIT (IC50 18nM) [19922818]	ALL	Phase II	41	Ph+ALL resistant to imatinib or intolerant/ relapse to chemotherapy or SCT	24%	Ottmann et al(2007c)
Dasatinib	ABL1 (IC50 0.53nM), SRC (IC50 0.21nM), FGR (IC50 0.5nM), FYN (IC50 0.7nM), HCK (IC50 0.35nM), LYN (IC50 0.57nM), PDGFRB (IC50 0.63nM), PDGFRA (IC50 0.45nM), c-KIT (IC50 0.62nM) [18183025,19039322]	ALL	Phase II	46	Ph+ALL resistant to imatinib or intolerant/ relapse to chemotherapy or SCT	35%	17496201

1.2 Current progress in cancer biomarker discovery

1.2.1 Introduction to biomarker in cancer diagnosis and prediction

Generally tumors are differentiated according to their gross morphological appearance of the cells and the surrounding tissues. However, such a differentiation criterion has limitations. First, it relies on a subjective review of the tissue, which depends on the knowledge and experience of a pathologist, which may not be consistent or reproducible [73,74]. Second, this method provides discrete, rather than continuous classification of disease into broad groups with limited ability to determine the treatment regime of individual patients[75]. Third, disease with identical pathology may have different origins and respond differently to treatment [76]. Last but not the least, current pathology reports offer little information about the potential treatment regime which a disease will respond to. The accurate diagnosis and differentiation of tumor subtypes remains a challenge and the efforts to combat cancer remain extremely disappointing. One main reason for the lack of desired success in cancer diagnosis and differentiation is that in many cases, cancer is diagnosed too late and treated with improper regimens. Therefore, factors that can accurately predict response/toxicity to systemic treatments are urgently needed.

Fortunately, disease differentiation based on molecular profile of diseases can overcome those limitations [5,77-80]. Spectacular advances in molecular medicine, genomics, and proteomics are nowadays leading to the search for new biomarkers in cancer research. Microarray technology, for example, has become a very important component of disease molecular differentiation. The gene expression profiles can be used to identify markers which are closely associated with early detection/differentiation of disease, or disease

behavior (disease progression, response to therapy), and could serve as disease targets for drug design [81].

1.2.2 Types of cancer biomakers

A cancer biomarker is a substance that is objectively measured and evaluated as an indicator of pathogenic processes or pharmacological response to a therapeutic intervention in oncology[82]. Broadly speaking, cancer biomarkers can be divided into three categories[83] based on the recent advances in genomic and genetic research:

(a) Diagnostic (screening) biomarkers are used to detect and identify a given type of cancer in an individual. This type of biomarkers is expected to possess high levels of diagnostic sensitivity and specificity, especially if it is used in large-screening trials;

(b) Prognostic biomarkers are commonly used in clinical outcome prediction once the disease status has been established, independent of the therapy that is used. They are expected to predict the likely course of the disease, reflecting the metastatic, recurrence potential and/or growth rate of the tumor and thus they have an important influence on the aggressiveness of the therapy[84]. Prognostic biomarkers can be further divided into biological progression markers and risk biomarkers[84]. Biological progression markers are measures of tumor burden and are commonly circulating cellular proteins that are associated with tumor progression. Among the most commonly used of these "tumor markers" are CA-125 for ovarian cancer and prostate-specific antigen (PSA) for prostate cancer [85-87]. Risk biomarkers are usually implicated in the mechanisms of disease causality or neoplastic progression, and are increasingly used in drug development to identify populations likely to be responsive to a given drug treatment. The foremost

example of this is ERBB2/HER2 gene amplification in 25% of patients with invasive breast cancer, which correlates with inferior patient survival [88];

(c) Predictive biomarkers or pharmacodynamic biomarkers serve to predict the likely clinical response to a specific treatment that are most relevant for monitoring effects of a drug or other intervention, and classifying individuals as "responders" or "non-responders" by using molecular, cellular, histopathological, and imaging parameters [83,84]. Such predictive classification is of a major importance in designing clinical drug trails to define an intended use for the drug under investigation. It is a indicator of drug effect, they may not necessarily correlate with or predict a therapeutic impact on the disease[84]. Examples of typical pharmacodynamic biomarkers include changes in proliferation using Ki67 expression, apoptosis using the TUNNEL assay, alterations in gene expression profiles, and functional or molecular imaging changes[89].

Although prognostic and predictive biomarkers have different focus, some factors, such as EGFR mutations, are used as both prognostic and predictive in clinical studies[90].

1.2.3 Approaches of cancer biomarker discovery

1.2.3.1 **Traditional gene discovery method**

The early approaches to discover and identify cancer biomarkers were mainly based on preliminary clinical or pathological observations. Two methods, candidate gene approach and positional cloning approach, have traditionally been used to discover physiologically and pathologically natural history markers markers underlying human diseases. Candidate gene method is based on prior biochemical knowledge about the genes, such as putative

functional protein domain of genes and tissues in which genes are expressed [91,92]. Li-Fraumeni syndrome [93], hereditary prostate cancer risk [92], metastasis of hepatocellular carcinoma [94], and breast cancer risk [95] were discovered in this manner. However very limited well-characterized genes are currently available [91], and most genes cannot be analyzed in this manner due to the limitation of biochemical knowledge.

In contrast to candidate gene method, positional cloning identifies genes without any prior knowledge about gene function[8]. This method is performed in patients and their family members using DNA polymorphisms. Alleles of markers that are in close proximity to the chromosome location of the disease genes can be determined by genetic linkage analysis, and critical region can be defined by haplotype analysis. The candidate genes residing in the critical regions can be identified [8,91]. This method was applied in identifying genes related with asthma [96], cardiovascular disorders [8,9], and diabetes mellitus [7]. However, the nature of positional cloning limits its resolution to relatively large regions of the genome [91]. The candidate genes within a certain critical region need to be filtered from the relatively large regions of the genome by identifying mutations in genes that segregate with the disease [91].

These traditional approaches lead to the development of some most reliable and validated markers in daily clinic, which include carcinoembryonic antigen for a variety of cancers, such as calcitonine for medulary thyroid carcinoma,prostate-specific antigen for prostate carcinoma, thyroglobulin for papillary or follicular thyroid carcinoma, human chorionic gonadotropin or alpha-fetoprotein for germ cell tumors, CA-125 for ovarian carcinoma, CA 15-3 for breast carcinoma, and SCC for squamous cell carcinoma of the cervix (Table 1-4) [97]. The optimum management of patients with several types of malignancy also

requires the use of tumor biomarkers correlating with clinical response to a specific treatment and most relevant for monitoring drug activity. The foremost example of these biomarkers includes mutations of EGFR and K-ras for Non-small cell lung cancer and Colorectal respectively(Table 1-4)[97].

Table 1-4 Tumor markers used in clinical practice. Simplifed list of predictive tumor markers in breast and other solid tumors [97]

Cancer type	**Marker(s)**	**Main use(s)**	**Clinical scenario**
Colorectal	CEA	Prognosis, postoperative sruvelliance, monitoring therapy	
Germ cell	AFP, HCG, LDH (prognosis only)	Prognosis, postoperative surveillance, monitoring therapy	
Trophoblastic	HCG	Prognosis, postoperative surveillance, monitoring therapy	
Ovarian	CA-125	Monitoring therapy, differential diagnosis of benign and malignant masses in postmenopausal women	
Prostate	PSA	Screening, prognosis, postoperative surveillance, monitoring therapy	
Breast	ER, PR	Predicting response to hormone therapy, prognosis	Endocrine agents
	HER-2	Predicting response to trastuzumab and lapatinib, prognosis	Anti-HER2 agnets
	uPA, PAI-1 CA15-3, CEA	Prognoiss in node-negative patients postoperative surveillance, monitoring therapy	Chemotherapy
Hepatocellular	AFP	Diagnostic aid, prognosis, postoperative surveillance, monitoring therapy	
Thyroid(differentiated)	Thyroglobulin	Postoperative surveillance, monitoring therapy	
Colorectal	K-ras mutations		Getuximab
Non-small cell lung cancer	EGFR mutations		Gefitinib, erlotinib
Gastrointestinal stomal tumors	C-kit or PDGFRA mutations		Imatinib mesulate

Abbreviations: CEA,carcinoembryonic antigen; AFP, alpha-fetoprotein; HGC, human choriogonadotrophin; LDH, lactate dehydrogenase; PSA, prostate-specific antigen; ER, estrogen receptor; PR, progesterone receptor; uPA, urokinase plasminogen activator; PAI, plasminogen activator inhibitor.

1.2.3.2 **New approaches of cancer biomarker discovery**

Currently, there are a number of newly emerged platforms leading to the search for new biomarkers in cancer research. On the proteomic side, we have a number of emerging technologies that are applied in the area of biomarkers discovery, including surface enhanced laser desorption ionization (SELDI) [98,99], mass sepectrometry combined with two-dimensional liquid chromatography [100-102] or two-dimensional gel electrophoresis [103-106], protein microarrays [107-109], and imaging mass spectrometry[110-112]. On the genomic side, there are equally powerful platforms for biomarkers discovery, which use polymerase chain reaction (PCR) [113,114], serial analysis of gene expression (SAGE) [115], and DNA microarrays [4,5,116]. These technique is widely used in cancer research for the identification of cancer biomarkers, and provide new insights into tumorigenesis, tumor progression and invasiveness [4,5,117-120].

Among these well developed techniques, DNA microarrays has emerged as being the most clinically useful[121]. Computational analysis of microarray data can be used to identify certain sets of cancer-associated marker genes based on gene expression patterns. Furthermore, gene expression profiling provides high levels of specificity and sensitivity in cancer process where classical histo- or immunopathological approaches are unsatisfactory. Currently, powerful microarray technology has provided several new molecular classifications[4,117,122,123] of different solid tumors but also new prognostic and predictive tools in breast cancer [97], colorectal tumors [124], prostate cancer[125], non-Hodgkin's lymphoma [126], acute myeloid leukemia [127]. In breast cancer, multigene predictors of response to chemotherapy, endocrine therapy, or targeted agents are in earlier stages of development [97]. Two profiles, in particular, have undergone

detailed studies in this malignancy: MammaPrint® and Oncotype DX®. Similar multigene signatures are under development in other solid tumors[97].

1.2.4 Brief introduction of microarray technology

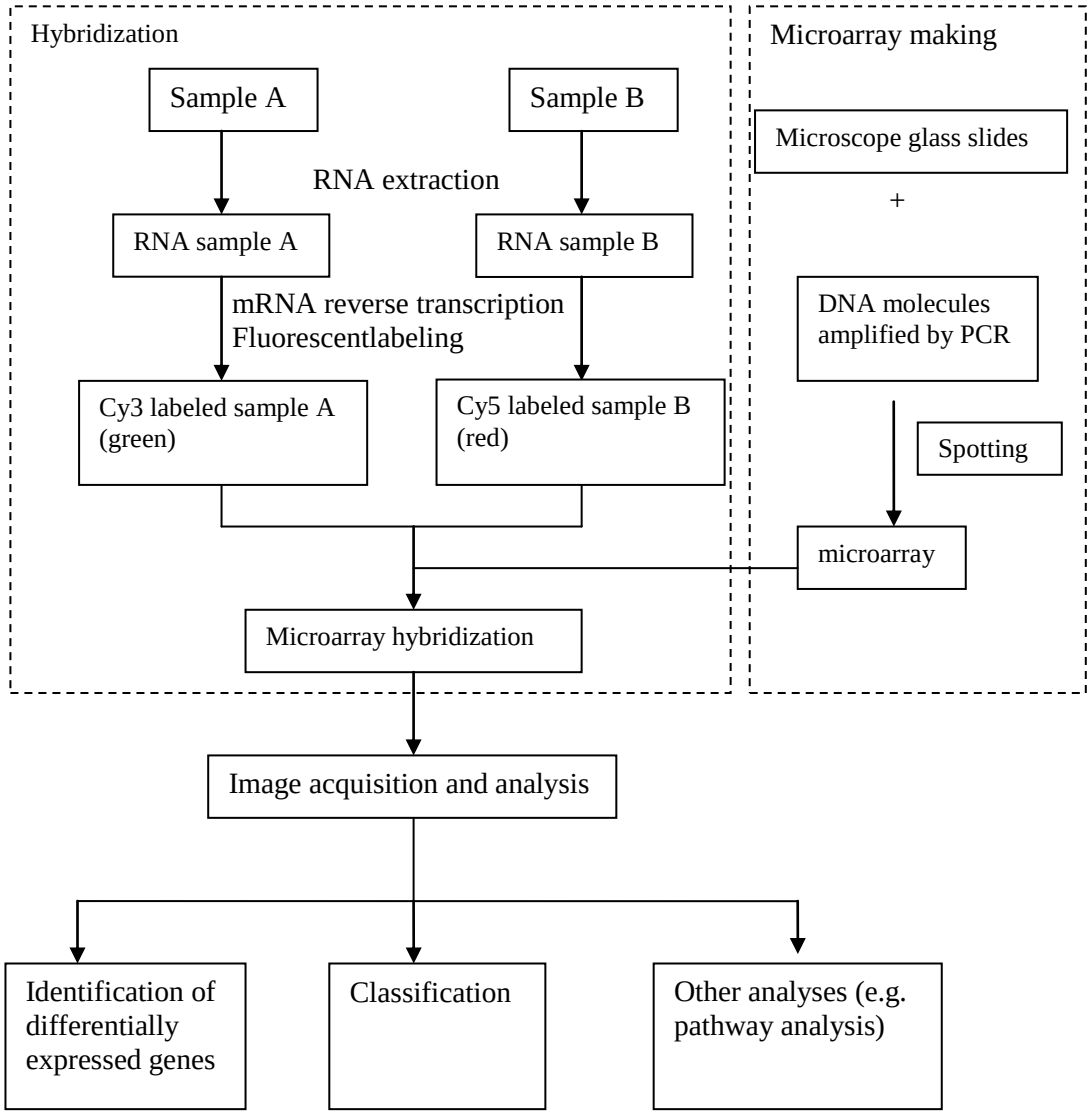
1.2.4.1 **Introduction to microarray experiments**

Microarray technology, also known as DNA chip, gene ship or biochip, is one of the indispensable tools in monitoring genome wide expression levels of genes in a given organism. Microarrays measure gene expression in many ways, one of which is to compare expression of a set of genes from cells maintained in a particular condition A (such as disease status) with the same set of genes from reference cells maintained under conditions B (such as normal status).

Figure 1-2 shows a typical procedure of microarray experiments [128,129]. A microarray is a glass substrate surface on which DNA molecules are fixed in an orderly manner at specific locations called spots (or features). A microarray may contain thousands of spots, and each spot may contain a few million copies of identical DNA molecules (probes) that uniquely correspond to a gene. The DNA in a spot may either be genomic DNA [130], or synthesized oligo-nucleotide strands that correspond to a gene [131-133]. This microarray can be made by the experimenters themselves (such as cDNA array) or purchased from some suppliers (such as Affymetrix GeneChip). The actual microarray experiment starts from the RNA extraction from cells. These RNA molecules are reverse transcribed into cDNA, labeled with fluorescent reporter molecules, and hybridized to the probes formatted on the microarray slides. At this step, any cDNA sequence in the sample will hybridize to specific spots on the glass slide containing its complementary sequence. The amount of cDNA bound to a spot will be directly proportional to the initial number of

RNA molecules present for that gene in both samples. Following, an instrument is used to read the reporter molecules and create microarray image. In this image, each spot, which corresponds to a gene, has an associated fluorescence value, representing the relative expression level of that gene. Then the obtained image is processed, transformed and normalized. And the analysis, such as differentially expressed gene identification, classification of disease/normal status, and pathway analysis, can be conducted.

Figure 1-2 Procedure of microarray experiment



1.2.4.2 ***Public repository for microarray data***

Reword to the variety of journals and funding agencies which have established and enforced microarray data submission standards, currently, a wealth of microarray data is now available in different databases such as the Stanford Microarray Database (SMD) [134], Gene Expression Omnibus (GEO) [135], and Array Express (EBI) [136]. Table 1-5 gives a list of public available microarray databases. Many of those databases require a minimum information about a microarray experiment (MIAME)-compliant manner in order to interpret the experiment results unambiguously and potentially be able to reproduce the experiment [137]. As a public resource, these expression databases are valuable substrates for statistical analysis, which can detect gene properties that are more subtle than simple tissue-specific expression patterns.

1.2.4.3 ***Statistical analysis of microarray data***

Since microarray contains the expression level of several thousands of genes, it requires sophisticated statistical analysis to extract useful information such as drug responsive markers. Theoretically, one would compare a group of samples of different conditions and identify good candidate genes by analysis of the gene expression pattern. However, a typical microarray data set is extremely sparse compared to traditional classification data sets[138]. Microarray data set may also contain some noises arising from measurement variability and biological differences [139,140]. The gene-gene interaction also affects the gene-expression level. Furthermore, the high dimensional microarray data can lead to some mathematical problems such as the curse of dimensionality and singularity problems in matrix computations, causing data analysis difficult. Therefore, choosing a suitable statistical method for gene selection is very important.

Table 1-5 A list of public available microarray databases

Database	Website*	Description	Organism	Ref
ArrayExpress	http://www.ebi.ac.uk/arrayexpress/	A public repository for microarray based gene expression data	European Bioinformatics Institute	[136]
ChipDB	http://chipdb.wi.mit.edu/chipdb/public/	A searchable database of gene expression	Massachusetts Institute of Technology	[141]
ExpressDB	http://twod.med.harvard.edu/ExpressDB/	A relational database containing yeast and E. coli RNA expression data	Harvard Medical School	[142]
Gene Expression Atlas	http://symatlas.gnf.org/SymAtlas/	A database for gene expression profile from 91 normal human and mouse samples across a diverse array of tissues, organs, and cell lines	Novartis Research Foundation	[143]
Mouse Gene Expression Database (GXD)	http://www.informatics.jax.org/menus/expression_menu.shtml	An extensive and easily searchable database of gene expression information about the mouse	The Jackson Laboratory, Bar Harbor, Maine	[144]
Gene Expression Omnibus (GEO)	http://www.ncbi.nlm.nih.gov/geo/	Microarray database containing tens of millions of expression profiles	National Center for Biotechnology Information	[135]
GermOnline	http://www.germonline.org/index.html	Information and microarray expression data for genes involved in mitosis and meiosis, gamete formation and germ line development across species	Biozentrum and Swiss Institute of Bioinformatics	[145]
Human Gene Expression (HuGE) Index database	http://www.biotechnologycenter.org/hio/	A comprehensive database to understand the expression of human genes in normal human tissues	Boston University	[146]
MUSC DNA Microarray Database	http://proteogenomics.musc.edu/ma/musc_madb.php?page=home&act=manage	A web-accessible archive of DNA microarray data	Medical University of South Carolina	[147]
RIKEN Expression Array Database (READ)	http://read.gsc.riken.go.jp/	A database of expression profile data from the RIKEN mouse cDNA microarray	RIKEN Yokohama Institute	[148]
Rice Expression Database (RED)	http://red.dna.affrc.go.jp/RED/	Expression profiles obtained by the Rice Microarray Project and other research groups	National Institute of Agrobiological Sciences, Japan	[149]
RNA Abundance Database (RAD)	http://www.cbil.upenn.edu/RAD/php/index.php	A public gene expression database designed to hold data from array-based and nonarray-based (SAGE) experiments	University of Pennsylvania	[150]
Saccharomyces Genome Database (SGD): Expression Connection	http://db.yeastgenome.org/cgi-bin/expression/expressionConnection.pl	A gene expression database of Saccharomyces genome	Stanford University	[151]

Stanford Microarray Database (SMD)	http://genome-www5.stanford.edu/	Raw and normalized data from microarray experiments, as well as their corresponding image files	Stanford University	[134]
Yale Microarray Database (YMD)	http://info.med.yale.edu/microarray/	A microarray database for large-scale gene expression analysis.	Yale University	[152]
yeast Microarray Global Viewer (yMGV)	http://www.transcriptome.ens.fr/ymgv/	A database for yeast gene expression	Ecole Normale Superieure, Paris, France	[153]

The statistical methods in microarray data analysis can be classified into two groups: unsupervised learning methods and supervised learning methods. Unsupervised analysis of microarray data aims to group relative genes without knowledge of the clinical features of each sample [154]. A commonly-used unsupervised method is hierarchical clustering method. This method groups genes together on the basis of shared expression similarity across different conditions, under the assumption that genes are likely to share the same function if they exhibit similar expression profiles [155-158]. Hierarchical clustering creates phylogenetics trees to reflect higher-order relationship between genes with similar expression patterns by either merging smaller clusters into larger ones, or by splitting larger clusters into smaller ones. A dendogram is constructed, in which the branch lengths among genes also reflect the degree of similarity of expression [159,160]. By cutting the dendogram at a desired level, a clustering of the data items into the disjoint groups can be obtained. Hierarchical clustering of gene expression profiles in rheumatoid synovium identified 121 genes associated with Rheumatoid arthritis I and 39 genes associated with Rheumatoid arthritis II [161]. Unsupervised methods have some merits such as good implementations available online and the possibility of obtaining biological meaningful results, but they also possess some limitations. First, unsupervised methods require no prior knowledge and are based on the understanding of the whole data set, making the clusters difficult to be maintained and analyzed. Second, genes are grouped based on the similarity which can be affected by input data with poor similarity measures. Third, some of the unsupervised methods require the predefinition of one or more user-defined parameters that are hard to be estimated (e.g. the number of clusters). Changing these parameters often have a strong impact on the final results [162].

In contrast to the unsupervised methods, supervised methods require a priori knowledge of the samples. Supervised methods generate a signature which contains genes associated with the clinical response variable. The number of significant genes is determined by the choice of significance level. Support vector machines (SVM) [163] and artificial neural networks (ANN) [164] are two important supervised methods. Both methods can be trained to recognize and characterize complex pattern by adjusting the parameters of the models fitting the data by a process of error (for example, mis-classification) minimization through learning from experience (using training samples). SVM separates one class from the other in a set of binary training data with the hyperplane that is maximally distant from the training examples. This method has been used to rank the genes according to their contribution to defining the decision hyperplane, which is according to their importance in classifying the samples. Ramaswamy et al. used this method to identify genes related to multiple common adult malignancies [5]. ANN consists of a set of layers of perceptrons to model the structure and behavior of neutrons in the human brain. ANN ranks the genes according to how sensitive the output is with respect to each gene's expression level. Khan et al identified genes expressed in rhabdomyosarcoma from such strategy [118].

In classification of microarray datasets, it has been found that supervised machine learning methods generally yield better results [165], particularly for smaller sample sizes [140]. In particular, SVM consistently shows outstanding performance, is less penalized by sample redundancy, and has lower risk for over-fitting [166,167]. Furthermore, some studies demonstrated that SVM-based prediction system was consistently superior to other supervised learning methods in microarray data analysis [168-170]. As such, SVM for identifying marker genes related to the survival and death lung cancer patients from microarray data analysis were used in this study. The lung cancer patients discrepancy

capability of identified disease markers were futher evaluated and compared by supervised methods including SVM and Probabilstic Neural Network, which is an extension of ANN; whereas the unsupervised approach, hierarchical clustering, were used to generate clusters of patients bearing the different survivalbilities.

*1.2.4.4 **Feature selection in microarray data analysis***

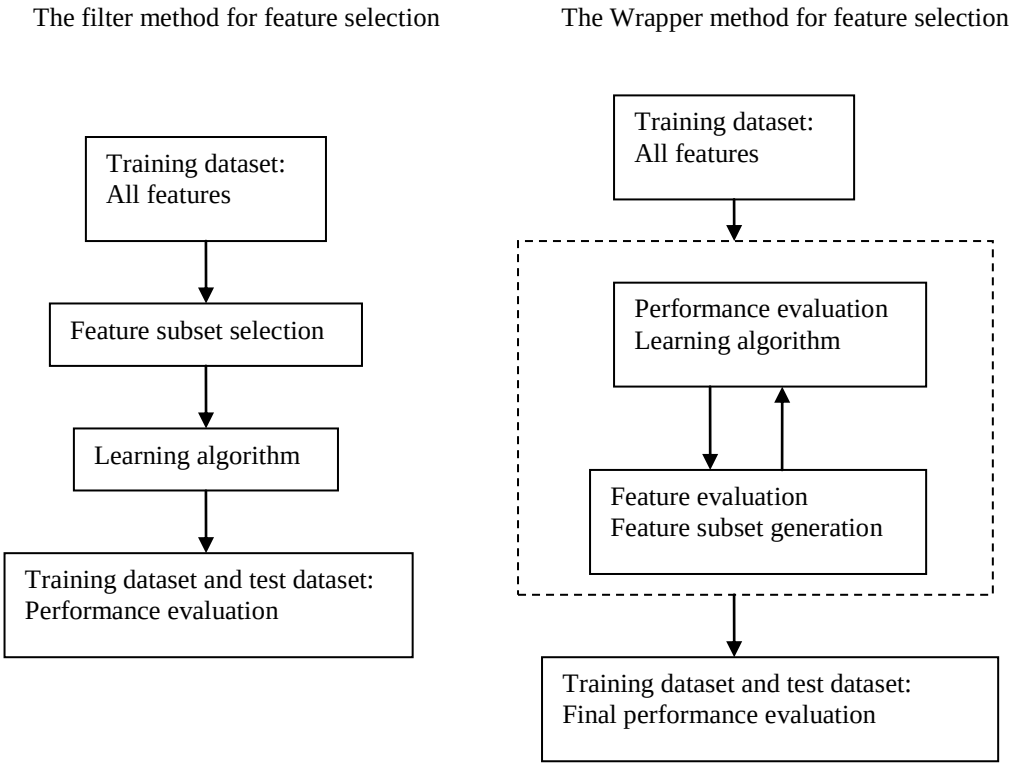
No matter whether the supervised or unsupervised methods are used, one critical problem encountered in both methods is feature selection, which has become a crucial challenge of microarray data analysis. The challenge comes from the presence of thousands of genes and only a few dozens of samples in currently available data. From the mathematical view, thousands of genes are thousands of dimensions. Such a large number of dimensions leads microarray data analysis to problems such as the curse of dimensionality [171,172] and singularity problems in matrix computations. Therefore, there is a need of robust techniques capable of selecting the subsets of genes relevant to a particular problem from the entire set of microarray data both for the disease classification and for the disease target discovery.

The goal of gene selection from microarray data is to search through the space of gene subsets in order to identify the optimal or near-optimal one with respect to the performance measure of the classifier. Many gene selection methods have been developed, and they generally fall into two categories: the filter method and the wrapper method [173]. Figure 1-3 shows how these two methods work.

In brief, the filter method selects genes independent of the learning algorithms [174-176]. It evaluates the goodness of the genes from simple statistics computed from the empirical distribution with the class label [177]. Filter method has some pre-defined criteria. Mutual

information and statistical testing (e.g. T-test and F-test) are two typical examples of filter method [4,174,178-182]. Filter method can be easily understood and implemented, and needs little computational time. But the pitfall of this method is that it is based on the assumption that genes are not connected to each other, which is not true in real biological process.

Figure 1-3 Filter method versus wrapper method for feature selection



Wrapper method generates genes from the evaluation of a learning algorithm. It is conducted in the space of genes, evaluating the goodness of each gene or gene subsets by such criteria as cross-validation error rate or accuracy from the validation dataset [183]. The wrapper method is very popular among machine learning methods for gene discovery [173,184,185]. Although the wrapper method needs extensive computational resources and time, it considers the gene-gene interaction and its accuracy is normally higher than

the filter method [173,184,185]. Recursive feature elimination (RFE) is a good example of the wrapper method for disease gene discovery. The RFE method uses the prediction accuracy from SVM to determine the goodness of a selected subset. This thesis will employ RFE for disease gene discovery from microarray data.

1.2.5 The problems of current marker selection methods

The methodology of SVM and RFE will be discussed in Chapter 2 in details. Here, some problems encountered in current marker discovery from microarray data analysis are discussed. One problem is to specify the number of genes for differentiating disease. The number of derived colon cancer genes and leukemia genes ranges from 1 to 200 [4,186-191]. 50 genes were arbitrarily chosen for differentiating AML from ALL by Golub et al, since they supposed that 50 genes might reflect the difference between AML and ALL [4]. In most cases, the gene number was decided by the classification performance of different gene combinations. The gene combination which produced the highest classification accuracy constituted the gene signature. This strategy might produce small sets of genes (one or two genes) that formed accurate classifier [189-191]. For example, Slonim et al reported that the classifier consisting of one gene (HOXA9) outperformed all of other classifiers consisting of other gene combinations for recurrence prediction in AML patients [191]. Li and Yang showed that one gene (Zyxin) constituted the best classifiers for AML/ALL differentiation [189]. Nevertheless these results were only obtained and tested of one dataset. Considering that the number of genes should correlate with the disease situation, the selected genes should be large enough to be robust against noise and small enough to be readily applied in clinical settings. Therefore, it is not appropriate to use the arbitrary gene number. Similarly, to use just one dataset to decide the optimal gene

number may not be satisfactory, because the optimal gene number varies with the different sample sizes and sample combinations [139,192,193].

Another problem in gene discovery is that gene signatures are highly unstable and strongly depended on the selection of patients in the training sets [4,118,139,140,194-197], despite the use of sophisticated class differentiation and gene selection methods by various groups. The unstable signatures were observed in most microarray datasets including colon cancer, lung adenocarcinoma, non-Hodgkin lymphoma, acute lymphocytic leukemia, acute myeloid leukemia, breast cancer, medulloblastoma, and hepatocellular carcinoma [139,159,168,173,176,194,198-201]. While these signatures display high predictive accuracies, the highly unstable and patient-dependent nature of these signatures diminishes their application potential for diagnosis and prognosis [139]. Moreover, the complex and heterogenic nature of disease such as cancer may not be adequately described by the few cancer-related genes in some of these signatures. The unstable nature of these signatures and their lack of disease-relevant genes also limit their potential for target discovery. The instability of derived signatures is likely caused by the noises in the microarray data arising from such factors as the precision of measured absolute expression levels, capability for detecting low abundance genes, quality of design and probes, annotation accuracy and coverage, and biological differences of expression profiles [140,202]. Apart from enhancing the quality of measurement and annotation, strategies for improving signature selection have also been proposed. These strategies include the use of multiple random validation [139], large sample size [203], known mechanisms [204], and robust signature-selection methods which is insensitive to noises [140,205,206].

This thesis explored a new gene signature selection method aiming to reduce the chances of erroneous elimination of predictor-genes due to the noises contained in microarray dataset. Multiple random sampling and gene-ranking consistency evaluation procedures were incorporated into RFE gene signature selection method. The consistent genes obtained from the multiple random sampling method may give us a better understanding to the disease initiation and progress, and may provide potential disease targets.

1.3 Current progress in tumor antigen discovery

1.3.1 Overview of tumor vaccine for cancer immunotherapy

The major role of the immune system is to destroy cells expressing non-self or mutation proteins, which is carried by cytotoxic T lymphocytes (CTLs) by recognizing short peptides in association with major histocompatibility complex (MHC) class I molecules. Tumors have been known to express aberrant levels of mutated or modified forms of proteins that are associated with malignant growth. Such proteins can be immunogenic and stimulate cellular and humoral immune responses[207-209]. Inducing T-cell immunity by peptide vaccines derived from these alternated proteins holds great potential of effectively destroying cells carrying viral invaders (by recognizing the antigenic viral peptides) or against tumor cells (by recognizing peptides from tumor antigens) [210].

The interests in peptide-based cancer vaccines have been steadily growing since it has been indeed shown that peptide immunization can elicit specific CTL responses and confer protective immunity against tumor cells. Furthermore, the identification of tumor-associated and -specific T cell epitopes has contributed significantly to the understanding of the interrelationship of tumor and immune system and is instrumental in the

development of attractive biomarkers and therapeutic vaccines to treat patients [211]. The discovery and identification of tumor antigens, which now number in the hundreds (Table 1-6) [212]; however, very few of tumor antigens have been successfully identified and the immune response they provoke in cancer treatment [212,213]. The main reason is that tumors cancer escape an immune response in many ways[214]. Moreover, Tumors are generally genetically unstable, and they can lose their antigens by mutation. Some tumors even lose expression of a particular MHC molecule, totally blocking antigen presentation[215]. There is a need to search for new cancer immunotherapies, such as cancer vaccines, from more diverse sources [216-219] that takes into consideration not only tumor-specific mutations and MHC-binding but also the expression profiles of the antigens, processing and transport of the epitopes, and availability of T-cell repertoire in specific tumors [220].

Table 1-6 Selected cancer vaccines in early and late stages of clinical trials [212]

Product Composition	Company (location)	Product	Description	Indication	Trial phase
	Antigen Express (Worcester, MA, USA; a subsid-iary of Generex Biotechnology, Toronto)	Her-2/neu breast cancer vaccine	Her-2/neu epitope peptide con-jugated at N terminus to the C terminus of the key moiety of the MHC class II ¡Vassociated invari-ant chain (Ii protein) containing a four ¡Vamino-acid (LRMK) modi-fication	Breast cancer	Phase 2
	Apthera (Scottsdale, AZ, USA)	NeuVax	Immunopeptide (E25) from Her-2/neu administered together with GM-CSF	Early-stage breast cancer	Phase 1/ 2
	Argos Therapeutics (Durham, NC, USA)	AGS-003	Autologous dendritic cells loaded with total RNA from resected tumors	Renal cancer	Phase 2
	Immunocellular Therapeutics (Los Angeles, CA, USA)	ICT-107	Autologous dendritic cells treated with tumor-specific peptides from 6 antigens expressed on glioblastomas	Brain cancer	Phase 1
	Immunotope (Doylestown, PA, USA)	IMT-1012	Peptide vaccine containing 12 tumor-associated peptides discovered through proteom-ics, including A-kinase anchor protein 9, midasin (MIDAS-containing protein RAD50), talin 1, vinculin vimentin and cen-trosome-associated protein 350	Advanced ovarian and breast cancer	Phase 1
	Pevion Biotech (Bern, Switzerland)	Pevi-Pro	Influenza virosomes expressing three Her2/neu epitopes	Breast cancer	Phase 1
	Vaxon Biotech (Paris)	Vx-001	A peptide vaccine comprising the cryptic peptide human telomerase reverse tran-scriptase (TERT572) and its HLA-A*0201-restricted modified variant (TERT572Y)	NSCLC	Phase 1
Whole-cell-based autologous cells (personalized)	Avax Technologies (Philadelphia)	M-Vax	Autologous cell vaccine in which patient tumor cells are treated with the hapten dinitrophenyl	Metastatic melanoma with at least one tumor to create vaccine	Phase 3
Whole-cell-based autologous cells (personalized)	Dendreon	Provenge	Autologous dendritic cells exposed ex vivo to fusion protein combining prostate alkaline phosphatase and GM-CSF	Asymptomatic, metastatic hor-mone-refractory prostate cancer	Phase 3
Whole-cell-based autologous cells (personalized)	Geron (Menlo Park, CA, USA)	GRNVAC1	Autologous dendritic cells trans-fected with mRNA for human telomerase and a portion of lysosome-associated membrane protein (enhances antigen pre-sentation)	AML in remission	Phase 2
Whole-cell-based autologous cells (personalized)	IDM Pharma	Bexidem Uvidem Collidem	Autologous interferon- £^-activated macrophages (monocyte-derived activated NK cells). Autologous dendritic cell vaccine loaded ex vivo with tumor antigens derived from resected tumor	Superficial bladder cancer Melanoma with M1a or M1b stage disease and/or in-transit lesions; stage III and IV melanoma Colorectal cancer	Phase 2/3 Phase 2 Phase 1/2
Whole-cell-based autologous cells (personalized)	Introgen Therapeutics (Austin, TX, USA)	INGN 225	Dendritic cells treated with an adenovector carrying the human p53 gene	Advanced metastatic SCLC Breast	Phase 2

Whole-cell-based autologous cells (personalized)	MolMed (Milan)	M3TK	T cells bioengineered to express MAGE 3 tumor antigen	Metastatic melanoma	Phase 2 (enrollment halted)
Whole-cell-based autologous cells (personalized)	Northwest Biotherapeutics (Bethesda, MD, USA)	DC-Vax Prostate DC-Vax Brain	Dendritic cells loaded with recombinant prostate-specific membrane antigen (PSMA) Dendritic cells loaded with tumor extract	Hormone-dependent, nonmeta-static prostate cancer Newly diagnosed glioblastoma multiforma requiring surgery, radiation and chemotherapy	Phase 3 Phase 2
Whole-cell-based autologous cells (personalized)	Prima Biomed (Sydney, Australia)	CVac	Dendritic cells primed with a mucin-1 and a mannan-fusion protein adjuvant	Late-stage ovarian cancer	Phase 2
Whole-cell-based autologous cells (personalized)	Cell Genesys	GVAX pancreatic GVAX leukemia	Two allogeneic cultured cancer lines, irradiated and bioengi-neered to secrete GM-CSF. One allogeneic leukemia cell line irradiated and bioengineered to secrete GM-CSF	Metastatic pancreatic cancer Newly diagnosed AML, chronic CML and myelodysplastic syndrome	Phase 2 Phase 2
Whole-cell-based autologous cells (personalized)	NovaRx (San Diego)	Lucanix	Four non-small cell lung cancer cell lines carrying antisense oli-gonucleotides against transform-ing growth factor £]-2	Advanced NSCLC	Phase 3
Whole-cell-based allogeneic tumor cells (off-the-shelf)	Company (location)	Product	Description	Indication	Trial phase
Whole-cell-based allogeneic tumor cells (off-the-shelf)	Onyvax (London)	Onyvax-P protein	Three human cell lines repre-senting different stages of pros-tate cancer	Hormone-resistant prostate cancer	Phase 2
Unique-antigen-based (personalized): purified peptide	Antigenics	HSPPC-96 Oncophage	Heat shock protein vaccine puri-fied from autologous tumor cells	Recurrent glioma Resected renal-cell carcinoma (RCC)	Phase 2 (investigator-initiated trial) Phase 3 (completed)
Unique-antigen-based (personalized): purified peptide	Biovest International	BiovaxID	Tumor-specific idiotype conju-gated to keyhole limpet hemo-cyanin, plus GM-CSF	Mantle cell lymphoma Indolent follicular B-cell non-Hodgkin ¡$ lymphoma	Phase 2 Phase 3
Shared antigen (off-the-shelf): purified protein or peptide	Apthera (Scottsdale, AZ, USA)	NeuVax	Immunogenic peptide derived from the Her-2/neu protein plus GM-CSF	Early-stage Her-2-positive breast cancer	Phase 2/3
Shared antigen (off-the-shelf): purified protein or peptide	CellDex	CDX-110	A 14-amino-acid segment of a mutated EGFR	Glioblastoma multiforme	Phase 2/3
Shared antigen (off-the-shelf): purified protein or peptide	Cytos Biotechnology (Schlieren, Switzerland)	CYT004-MelQbG10	Modified fragment of the Melan-A/MART-1 protein coupled to the carrier QbG10	Advanced-stage melanoma	Phase 2
Shared antigen (off-the-shelf): purified protein or peptide	Generex Biotechnology	Ii-Key/HER2/neu cancer vaccine	Peptide vaccine containing Ii-Key modified Her-2/neu pro-tein fragment	Node-negative breast cancer	Phase 2
Shared antigen (off-the-shelf): purified protein or peptide	GlaxoSmithKline Biologicals (Brussels, Belgium)	MAGE-A3 antigen-specific can-cer immunotherapeu tic	Liposomally packaged cancer vaccine against MAGE-3 antigen	Metastatic MAGE-A3-positive melanoma NSCLC following surgery	Phase 3 Phase 3
Shared antigen (off-the-shelf): purified protein or peptide	IDM Pharma	IDM-2101	Nine CTL epitopes from four tumor-associated antigens, including two proprietary native epitopes and	NSCLC	Phase 2

			seven modified epitopes and one universal epitope (a source of T-cell help)		
Shared antigen (off-the-shelf): purified protein or peptide	Immatics Biotechnologies (Tuebingen, Germany)	IMA901 IMA910	Peptide vaccine comprising multiple fully synthetic tumor-associated peptides	Renal cancer Colorectal cancer	Phase 2 Phase 1/2
Shared antigen (off-the-shelf): purified protein or peptide	Norwood Immunology (Chelsea Heights, Australia)	Melanoma cancer vaccine	Melanoma-specific peptides gp100 and MAGE-3	Melanoma	Phase 2
Shared antigen (off-the-shelf): purified protein or peptide	Oncothyreon	Stimuvax	Liposomal vaccine containing a synthetic 25 ¡Vamino-acid-peptide sequence from MUC-1	Stage lll NSCLC	Phase 3
Shared antigen (off-the-shelf): purified protein or peptide	Pharmexa (Hoersholm, Denmark)	GV1001	Recombinant protein vaccine tar-geting human telomerase reverse transcriptase, plus GM-CSF	Pancreatic Liver Lung	Phase 3 Phase 2 Phase 2
Antigenics		OncoPhage	Heat shock protein vaccine puri-fied from autologous tumor cells	Renal cell carcinoma	Approved in Russia Granted fast track status by US FDA
Biovest International		BiovaxID	Tumor-specific idiotype conju-gated to keyhole limpet hemocya-nin, plus GM-CSF	Various B-cell ¡Vrelated cancers	Compassionate use in France, Germany, Italy, Greece, Spain and the UK. Granted fast track status by US FDA
Corixa (acquired by GSK in 2005)		Melacrine	Lysate from two melanoma cell lines, Detox adjuvant (proprietary) with monophosphoryl lipid A and mycobacterial cell wall skeleton	Melanoma	Approved in Canada
CreaGene (Seoul)		CreaVaxRCC	Autologous monocytes treated with GM-CSF and IL-4 to create immature dendritic cells acti-vated with tumor extracts plus cytokines	Metastatic renal cell carcinoma	Approved in Korea
Genoa Biotechnologia (Brazil)		Hybricell	Autologous monocytes treated with cytokines and converted to dendritic cells that are fused with patient-derived tumor cells	Various cancers	Approved in Brazil
Vaccinogen (Frederick, MD, USA)		OncoVax	Metabolically active, irradiated, autologous tumor cells with BCG	Colon cancer	Approved in Europe, available in Switzerland Granted Fast Track status by FDA
Mologen (Berlin)		dSlim/Midge	Genetically modified allogeneic (human) tumor cells for the expression of IL-7, GM-CSF, CD80 and CD154, in fixed combination with a DNA-based double stem loop immunomodulator (dSLIM).	Kidney cancer	Orphan drug status granted by EMEA in 2006
Center of Molecular Immunology (Cuba)		CimaVax EGF	EGF conjugated to rP64k	Lung cancer	Cuba, Peru

.

1.3.2 Introduction toT cell-defined tumor antigens

As discussed previously, the most dominant mechanism of immune response to foreign or self protein antigens is the activation of T-cells by the recognition of T-cell receptors of specific peptides degraded from these proteins and transported to the surface of antigen presenting cells [221]. Theoretically, every sub-sequence along the protein could be antigenic. However, T cell immunity is limited to a small number of immunodominant peptides [210]. Such recognition requires the binding of antigen-peptides to major histocompatibility complex (MHC) molecules. Peptide epitopes recognized by T-cells are potential tools for diagnosis and vaccines for immunotherapy of infectious, autoimmune, and cancer diseases [222]. Currently, reverse immunogenetic approaches attempt to optimize the selection of candidate epitopes, and thus minimize the experimental effort needed to identify new epitopes [223]. When predicting cytotoxic T cell epitopes, the main focus has been on the highly specific MHC binding event [223].

1.3.2.1 ***The role of MHC-peptides biding in T-cell epitopes pathway***

The MHC pathway of antigen processing and presentation is highly complex and involve many steps that select the peptides to be presented on the cell surface. The first step in this pathway is the digestion of native proteins. Peptides (epitopes) are initially generated in this step by proteasome, a multi-subunit protease that is responsible for the majority of intra-cellular protein degradation and representing about 1% of total cellular proteins [224]. It shows some degree of specificity, as some sites in the proteins are preferentially cleaved [225]. Furthermore, those peptides having escaped from proteolytic attack of various cytosolic proteases are transported into the endoplasmic reticulum (ER) by TAP.

Within the ER, peptides may undergo N-terminal trimming, whereas their C terminus is kept intact [226,227]. The ER aminopeptidase associated with Ag processing responsible

for this trimming was recently identified [228-230]. Peptides with correct sizes and proper amino acid sequence motifs bind to MHC, and the receptor-peptide complexes are transferred via the Golgi to the cell surface. The MHC binding is the most essential and selective step since only a minor fraction of the peptide repertoire will bind to a given MHC molecule [231]. Peptides in the ER with less efficient MHC binding are either degraded there or exported for rapid degradation in the cytosol [232]. For several areas in immunology, including the identification of CTL epitopes and vaccine design, reliable prediction of MHC binders is important [233].

1.3.2.2 ***Types of tumor antigens***

Broadly speaking, tumor antigens can be classified into unique antigens and shared antigens based on their pattern of expression [234-237]. Examples of these two groups of antigens are listed in Table 1-7. Unique antigens result from point mutations in genes that are expressed ubiquitously and some of these mutations may be implicated in tumoral transformation by affecting the coding region of the gene [235]. Such tumor-specific antigens (TSAs), which are unique to the tumor of an individual patient or restricted to very few patients, play an important role in the natural anti-tumor immune response of individual patients.

On the other hand, shared antigens are present on many independent tumors and may also be expressed by normal tissues. Such tumor-associated antigens (TAAs) can be further divided into three groups [237]. One group corresponds to peptides encoded by "cancer-germline" genes, such as MAGE, which are expressed in many tumors but not in normal tissues. The only normal cells in which significant expression of such genes has been detected are placental trophoblasts and testicular germ cells. Because these cells do not express MHC class I molecules, gene expression should not result in the expression of the

antigenic peptides and such antigens can therefore be considered as strictly tumor-specific. The genes encoding such antigens have also been referred to as "cancer-testis" (CT) genes. A second group of shared tumor antigens, named differentiation antigens, are also expressed in the normal tissue of origin of the malignancy. The paradigm is tyrosinase, which is expressed in normal melanocytes and in most melanomas. Antigens of this group are not tumor-specific, and their use as targets for cancer immunotherapy may result in autoimmunity towards the corresponding normal tissue. More serious concerns about autoimmune side effects apply to carcinoembryonic antigen (CEA), an oncofetal protein expressed in normal colon epithelium and in most gut carcinomas. Autoimmune toxicity should not be an issue, however, in situations where the tissue expressing the antigen is dispensable or even rejected by the surgeon in the course of cancer therapy, as would be the case for prostate specific antigen (PSA). The last group of shared antigens refers antigens that are expressed in a wide variety of normal tissues and overexpressed in tumors. Full list of T-cell defined TSAs and TAAs are available at Cancer Immunity database (http://www.cancerimmunity.org/ peptidedatabase/Tcellepitopes.htm).

Table 1-7 Examples of tumor-specific antigens and shared antigens (Source: GSK)

Expressed in cancer only	
	MAGE-3
	NY-ESO-1
	TRAG-3
Expressed in some normal tissues	
	WT-1
	PRAME
	SURVIVIN-2B
Overexpressed in cancer	
	Her-2
	MUC-1
	Survivin
Mutated, unique	
	p53
	a-actinin-4
	Malic enzymes

1.3.3 Application of computational methods for MHC-binding peptides and epitopes prediction

To facilitate the discovery of T-cell epitopes, computational methods for predicting MHC-binding peptides [238-244] and T-cell epitopes [223,242,245-249] have been developed. These methods predict MHC-binding peptides and T-cell epitopes by using binding motifs [240,241,244], quantitative matrices of structure-affinity relationships [242], structure-based methods [239], and statistical learning methods such as ANN [238,246] and SVM [246,250]. These methods achieve impressive prediction accuracies of 70%~90% for the binders and 40%~80% for the non-binders of selected MHC alleles [238-244]. T-cell epitope prediction accuracy has been further improved by integrating MHC-binding peptide prediction with proteasomal cleavage and TAP transport [223,231,247-249,251].

The application range and accuracy of *in silico* T-cell epitope prediction approaches can be affected by several factors. Most prediction systems have been developed for peptides of a few fixed lengths, 8, 9 and 10mers for MHC Class I and 9, 13 and 15mers for MHC Class II alleles, with the majority of these focusing on 9mers only [238-244]. While the longer peptides have been studied by using all of their constituent sequential segments of fixed lengths [238], this introduces some non-binder segments as binders. Some binders of longer lengths form mini-hairpin-like structure with the central section unattached and the two ends attached to the binding groove [252]. By fixing a shorter length, these peptides may not be adequately represented. Moreover, most studies have used a relatively small number of non-binders to train the respective prediction system, typically no more than 100 peptides [238-244], which may not be enough to cover the vast sequence space of 20^n possible sequences for an n-mers peptide. This inadequate representation of non-binders likely leads to a higher false binder rate. Some of the methods use sequence

straightforwardly for representing peptides, instead of their structural and physicochemical properties [243]. Consequently, the derived prediction systems tend to be less effective for peptides non-homologous to the known binders. Moreover, most studies cover a limited number of MHC alleles partly due to the lack of statistically significant number of known peptides in commonly studied length ranges.

These problems may be partially alleviated if the prediction algorithm is based on peptides of flexible lengths and sequence-derived structural and physicochemical properties, and the training is conducted by using a sufficiently diverse set of non-binders. A suitable method for accommodating these features is support vector machines [253]. SVM has shown promising capability for predicting proteins of varying lengths that belong to a specific functional class from sequence-derived structural and physicochemical properties [254,255]. Improved performance has been archieved while predicting novel proteins non-homologous to other proteins [255,256]. MHC-binding peptides also possess similar characteristics in that they also share some structural and physicochemical features to facilitate MHC binding [257-259]. Therefore, SVM is expected to be equally applicable for predicting MHC-binding peptides.

Based on this assumption, our research group developed a SVM prediction system (MHC-BPS) of 18 MHC Class I and 12 Class II alleles by using 4208~3252 binders and 234333~168793 non-binders, and evaluated by an independent set of 545~476 binders and 110564~84430 non-binders[260]. The comparison between MHC-BPS and other popular computational methods such as MHCBN, BIAS and AFIPHITH suggested that this tool works more efficiently than other well-known methods for the same HIV test dataset with improved sensitivity and specificity. Currently the reported sensitivities (true hit rates) of

three T-cell recognition software NetChop3.0, NetCTL1.2, and MHC-BPS (for the studied alleles in this work) are 81%, 98.5%, and 75−93.8%, and the reported specificities (true non-hit rates) are 48%, 32%, and 96.2−99.8% respectively [231,261,262]. In this work, MHC-BPS was therefore used to facilitate the in silico prediction of T-cell recognistion of newly derivied tumor antigens.

1.4 Scope and research objective

The main purpose of this study is to explore the usefulness of cancer molecular profiling toward "personalized" cancer therapies. Collected molecular profiling is used to analyze the general and specific mode of actions of combination treatment and prediction, to evaluate the efficacy of molecular-targeted cancer therapy, and to design bioinformatics tools for tumor biomarker and antigens discovery.

There are three main objectives of this work. The first objective of this study is to understand the mechanism that underlies the rational combination disease treatment (Chapter 3). Due to the lower response rate observed in a single molecular targeted therapy, systems-oriented drug design, such as development of multi-component therapies, has been increasingly emphasized as a potentially more productive strategy in multi-factorial diseases treatment. Understanding the molecular mechanisms underlying synergistic, potentiative and antagonistic effects of drug combinations is crucial to discover of novel efficacious combinations and multi-targeted agents from systemically perspective. Analysis of 117 selected drug combinations revealed 7 general and specific modes of action from the cases of synergistic, potentiative, additive, antagonistic and reductive combinations. It is suggested that knowledge of MI profiles of individual drugs, network crosstalk and regulation, and modes of actions of drug combinations are useful

starting points for investigating the effects of drug combinations. The molecular clues derived from the pharmacodynamic, toxicological and pharmacokinetic effects should be highlighted in the discovery of novel multicomponent therapies.

The second objective of this study, as shown in Chapter 4, is to study the drug efficacy of anticancer tyrosine kinase inhibitors by using genetic and expression profiles of target and bypass gene in selected cancer types. In this chapter, we retrospectively analyzed mutation, amplification and gene expression profiles of targets and known bypass genes of 4 well studied anti-cancer drugs, namely gefitinib, erlotinib, lapatinib, and trastuzumab. The clinical relevance of the correlation analysis against cell-line data was evaluated by comparing the levels of correlation between the individual profiles and the sensitive/resistant cell-lines with the reported level of correlations. The study suggested that the individual-profile tends to show good performance for sensitivity prediction and it is capable of predicting EGFR inhibitors sensitivity from NSCLC cell-lines at performance levels that reflect the sensitivity of real patients. Comparing to the individual profiles, the collective profiles showed a more balanced and improved correlation with sensitive and resistant cell-lines. This study suggests that an accurate patient differentiation, better safety profile, improved response rate and personalized treatment can be achieved by system-oriented molecular profiling approaches. Moerover, this study also indicates that disease molecular profiles are useful sources of discover new diagnostic and effective therapeutic targets with targeted disease population. In Chapter 5, we applied the similar strategy of genome-scale tumor-specific antigens (TSAs) search by collectively considering the tumor-specific mutations in differentially over-expressed genes in specific cancers. Virtual T-cell recognition analysis, including proteasomal cleavage, TAP mediated transport and MHC-I binding affinity prediction, were performed to ensure the

tumorgenic and epitopable of identified peptides. While the results show fair agreement in identifying know TSAs from melanoma and lung cancer, the known TSA hit rates (1.9% and 0.8%) are enriched by 29-fold and 35-fold over those of mutation analysis. The numbers of predicted TSAs are within the testing range of typical screening campaigns.

The third objective of this study is to design bioinformatics tools for cancer biomarker discovery from microarray data. A cancer biomarker discovery system is developed by using gene selection strategies from microarray data. This system aims to identify stable and cancer relevant biomarkers with good prediction performance for disease diagnosis and differentiation. The strategies include the incorporation of multiple random sampling methods and the evaluation of gene-consistency into RFE gene selection procedure. The stable gene signatures may help us understand the mechanism of disease initiation and process, and may provide an insight for diagnosing disease, predicting disease types, prognosis of the outcome of a specific therapeutic strategy, and drug resistance before drug treatment. In Chapter 6, survival marker selection from lung cancer is used as a case study to evaluate the performance of the system. The stable gene signatures provide the biologists an opportunity to further investigate the role of derived genes in the initiation and progress of a disease, and give suggestions about potential disease targets for therapeutic molecule design.

This thesis is divided into 7 chapters. Chapter 1 provides the background, current progress, and challenges of molecular-targeted therapy and cancer biomarker discovery. Chapter 2 describes the methodology of this study. Methods for machine learning approach, especially the support vector machine, and microarray data analysis are described in detail. The rationale of current combination therapy is discussed in Chapter 3. Chapter 4 and Chapter 5 present the application of integrated molecular profiling for drug sensitivity and

novel target identification. The development and application of bioinformatics tools for cancer biomarker discovery is described in Chapter 6. Chapter 7 presents the conclusion and future work.

2 Methodology

In this Chapter, a specific introduction of machine learning algorithms related to this study is described in the first Section of this chapter (Section 2.1), which includes Support vector machines, Probabilistic neural network, and Hierarchical clustering. Strategies of microarray data processing used for drug sensitivity, cancer biomarker, and tumor-specific antigen are presented in section 2.2. Other important methodologies used for biomarker discovery are further introduced in Section 2.3. The last section introduces the methodologies are required for tumor-specific antigen prediction.

2.1 Introduction to machine learning methods

Machine learning is the study of computer algorithms capable of learning to improve their performance automatically through experiences [263]. The goal of machine learning is to extract useful information from data by building good probabilistic modes, which should be simple enough to be understood by human [264]. Machine learning is closely related to statistics and pattern recognition, since they all study the analysis of data. However, unlike statistics and pattern recognition, machine learning is primarily concerned with the algorithmic complexity of computational implementations [265]. These methods for classification employ computational and statistical methods to construct mathematical models from training samples which is used to classify independent test sample. The training samples are represented by vectors which can binary, categorical or continuous.

Over the past 50 years, a variety of machine learning methods have been developed for solving real-life problems, for examples, Decision Tree (DT), Hidden Markov Model (HMM), Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Self

organized map (SOM). All these methods could be classified along the distinction between supervised and unsupervised learning. In supervised learning, a classification system is given some inputs along with their answers. The purpose of the learning process is to specify a relationship between the sample vector input and the answers. Algorithms under this category such as decision trees, neural networks and support vector machines were widely applied in the field of pattern recognition. On the contrary, answer set would not be given to unsupervised learning approach, so it is basically left on its own to classify its inputs. Well-known clustering methods and self-organized maps are implemented in the unsupervised manner. Websites for the freely downloadable codes of some methods are given in Table 2-1. In the following section, three machine learning algorithms will be introduced, including support vector machines, neural network and hierarchical clustering. The specific properties of each method will be discussed briefly.

Table 2-1 Websites that contain freely downloadable codes of machine learning methods

Decision Tree	
PrecisionTree	http://www.palisade.com.au/precisiontree/
DecisionPro	http://www.vanguardsw.com/decisionpro/jdtree.htm
C4.5	http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/c4.5/tutorial.html
C5.0	http://www.rulequest.com/download.html
KNN	
k Nearest Neighbor	http://www.cs.cmu.edu/~zhuxj/courseproject/knndemo/KNN.html
PERL Module for KNN	http://aspn.activestate.com/ASPN/CodeDoc/AI-Categorize/AI/Categorize/kNN.html
Java class for KNN	http://nlp.stanford.edu/nlp/javadoc/javanlp/edu/stanford/nlp/classify/old/KNN.html
LDA	
DTREG	http://www.dtreg.com/lda.htm
LR	
Paul Komarek's Logistic Regression Software	http://komarix.org/ac/lr/lrtrirls
Web-based logistic regression calculator	http://statpages.org/logistic.html
Neural Network	
BrainMaker	http://www.calsci.com/
Libneural	http://pcrochat.online.fr/webus/tutorial/BPN_tutorial7.html
fann	http://leenissen.dk/fann/
NeuralWorks Predict	http://www.neuralware.com/products.jsp
NeuroShell Predictor	http://www.mbaware.com/neurpred.html
SVM	

SVM light	http://svmlight.joachims.org/
LIBSVM	http://www.csie.ntu.edu.tw/~cjlin/libsvm/
mySVM	http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/index.html
BSVM	http://www.csie.ntu.edu.tw/~cjlin/bsvm/
SVMTorch	http://www.idiap.ch/learning/SVMTorch.html
WinSVM	http://www.cs.ucl.ac.uk/staff/M.Sewell/winsvm/
LS-SVMlab	http://www.esat.kuleuven.ac.be/sista/lssvmlab/
GIST SVM Server	http://svm.sdsc.edu/svm-intro.html

2.1.1 Support Vector Machines

Support vector machines (SVM), a new machine learning method proposed by Vapnik in 1995 [163,266,267], is a supervised classification algorithm that provides state-of-the art performance in a variety of applications, including image classification and object detection [268,269], text categorization [270], prediction of protein solvent accessibility [271], microarray data analysis [169,170,187,200], protein fold recognition [272], protein secondary structure prediction [273], prediction of protein-protein interaction [274] and protein functional class classification [254].

Based on the structural risk minimization principle of statistical learning theory[275,276], SVM has shown outstanding classification performance due to the system can be less penalized by sample redundancy and has lower risk for over-fitting[277,278]. In linearly separable cases, SVM constructs a hyper-plane to separate two different groups of feature vectors with a maximum margin. For the cases of biomarker identification from cancer patients and healthy people, a feature vector is represented by $\mathbf{x}_i$, with gene expression descriptors of a patient as its components. The hyperplane is constructed by finding another vector $\mathbf{w}$ and a parameter b that minimizes $\|\mathbf{w}\|^2$ and satisfies the following conditions:

$$\mathbf{w}\cdot\mathbf{x}_i + b \geq +1, \text{ for } y_i = +1 \qquad \text{Class 1 (active)} \qquad (1)$$

$$\mathbf{w}\cdot\mathbf{x}_i + b \le -1, \text{ for } y_i = -1 \qquad \text{Class 2 (inactive)} \qquad (2)$$

where y_i is the class index, **w** is a vector normal to the hyperplane, $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin and $\|\mathbf{w}\|^2$ is the Euclidean norm of **w**. Base on **w** and *b*, a given vector ***x*** can be classified by

$$f(\mathbf{x}) = sign[(\mathbf{w}\cdot\mathbf{x}) + b] \qquad (3)$$

A positive or negative *f(**x**)* value indicates that the vector **x** belongs to the cancer patient or healthy people respectively.

In nonlinearly separable cases, which frequently occur in classifying gene expression profile of high dimension, non-linear SVM projects both positive and negative examples into a higher-dimensional feature space using a kernel function K($\mathbf{x}_i$, $\mathbf{x}_j$) and then separates them in that space.

An example of a kernel function is the Gaussian radian basis kernel (RBF), which has been extensively used and consistently shown better performance than other kernel functions in a number of classification studies [271-274,279-281]

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_j - \mathbf{x}_i\|^2 / 2\sigma^2} \qquad (4)$$

After project the examples into higher-dimensional feature space, the linear SVM procedure is then applied to the feature vectors in this feature space with the following decision function:

$$f(\mathbf{x}) = sign(\sum_{i=1}^{l} \alpha_i^0 y_i K(\mathbf{x}, \mathbf{x}_i) + b) \qquad (5)$$

Where the coefficients α_i^0 and b are determined by maximizing the following Langrangian expression:

$$\sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (6)$$

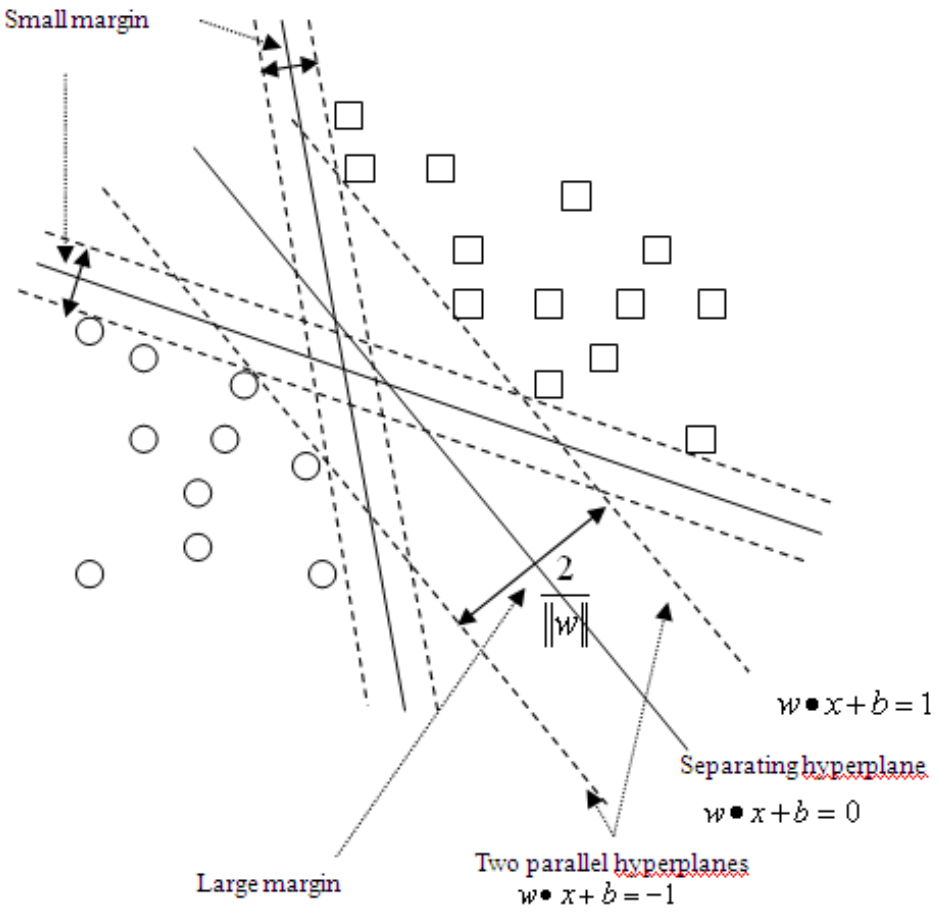
Under conditions:

$$a_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \quad (7)$$

A positive or negative value from Eq. (3) or Eq. (5) indicates that the vector **x** belongs to the positive or negative group respectively.

RBF kernel was chosen to construct the SVM model in cancer biomarker and tumor antigen discovery in this study. In practice, RBF kernel is the most widely used kernel function due to three reasons. First, linear kernel and sigmoid kernel can be treated as special cases since RBF kernel in certain parameters has the same performance as the linear kernel [282] or sigmoid kernel [283]. Second, comparing with polynomial kernel, RBF kernel has few parameters which influence the complexity of model selection. Third, RBF function has less computational cost compared with polynomial kernels in which kernel values may go to infinity or zero while the degree is large. Based on these reasons, we mainly applied RBF kernel SVM model in this study. We further choosed hard margin SVM and scanned σ from 1 to 200 for the best discrepancy model for survival and death lung cancer patient. An integrated SVM software, LibSVM[284], was chosen to do the machine learning in this work.

Figure 2-1 Margins and hyperplanes of a 2D SVM model



Linear SVM is the simplest form of SVM, in which the data represented as a p-dimensional vector (a list of p numbers) can be separated by a p-1 dimensional hyperplane. On each side of this p-1 hyperplane, two parallel hyperplanes can be constructed (Figure 2-1). The separating hyperplane is the one that maximizes the distance between these two parallel hyperplanes. Many linear hyperplanes (also called classifiers) can separate the data. However, only one can achieve the maximum separation. Under the assumption that the larger the margin or distance between these two parallel hyperplanes, the better of the generalization error of the classifier will be [285], the maximum separating hyperplane (also known as maximum-margin hyperplane) is clearly of interest.

2.1.2 Probabilistic neural network

Artificial neural network (ANN) is a statistical learning method inspired by the biological nervous system, which employs a multilayered network and uses its connected structures for pattern recognition and classification [286,287](Figure 2-2). Neural networks perform impressively in various applications such as pattern reorganization, association, and transformation to modeling in process control or expert system. It is suitable for both discrete and continuous data (especially better for the continuous domain). Probabilistic Neural Networks (PNN) is an improved form of ANN that classifies objects based on Bayes' optimal decision rule[288]:

$$h_i c_i f_i(\mathbf{x}) > h_j c_j f_j(\mathbf{x}) \quad (8)$$

where h_i and h_j are the prior probabilities, c_i and c_j are the costs of misclassification and $f_i(x)$ and $f_j(x)$ are the probability density function for class i and j respectively.

A compound $\boldsymbol{x}$ is classified into class i if the product of all the three terms is greater for class i than for any other class j (not equal to i). In most applications, the prior probabilities and costs of misclassifications are treated as being equal. The probability density function for each class for a univariate case can be estimated by using the Parzen's nonparametric estimator[289].

$$g(\mathbf{x}) = \frac{1}{n\sigma} \sum_{i=1}^{n} W\left(\frac{\mathbf{x} - \mathbf{x}_i}{\sigma}\right) \quad (9)$$

where n is the sample size, σ is a scaling parameter which defines the width of the bell curve that surrounds each sample point, $W(d)$ is a weight function which has its largest value at $d = 0$ and $(\boldsymbol{x} - \boldsymbol{x}_i)$ is the distance between the unknown vector and a vector in the training set. The Parzen's nonparametric estimator was later expanded by Cacoullos[290] for the multivariate case.

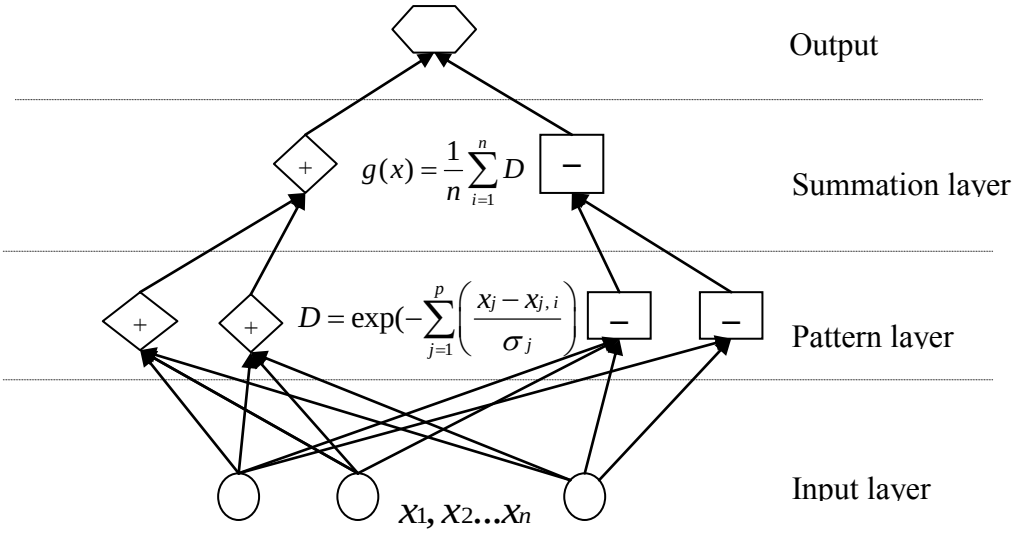
$$g(x_1,\ldots,x_p)=\frac{1}{n\sigma_1\ldots\sigma_p}\sum_{i=1}^{n}W(\frac{x_1-x_{1,i}}{\sigma_1},\ldots,\frac{x_p-x_{p,i}}{\sigma_p}) \qquad (10)$$

The Gaussian function is frequently used as the weight function because it is well behaved, easily calculated and satisfies the conditions required by Parzen's estimator. Thus the probability density function for the multivariate case becomes

$$g(\mathbf{x})=\frac{1}{n}\sum_{i=1}^{n}\exp(-\sum_{j=1}^{p}\left(\frac{x_j-x_{ij}}{\sigma_j}\right)^2) \qquad (11)$$

In this work, PNN was used to evaluation the classification performance of identified cancer biomarkers, and the programme is developed by our research group.

Figure 2-2 PNN architecture



The network architectures of PNN are determined by the number of input samples and descriptors in the training set. There are 4 layers in a PNN. The input layer provides input values to all neurons in the pattern layer and has as many neurons as the number of descriptors in the training set. The number of pattern neurons is determined by the total number of samples in the training set. Each pattern neuron computes a distance measure between the input and the training case represented by that neuron and then subjects the distance measure to the Parzen's nonparameteric estimator. The summation layer has a neuron for each class and the neurons sum all the pattern neurons' output corresponding to members of that summation neuron's class to obtain the estimated probability density function for that class. The single neuron in the output layer then estimates the class of the unknown sample x by comparing all the probability density function from the summation neurons and choosing the class with the highest probability density function.

2.1.3 Hierarchical clustering

Cluster analysis, which is one of the famous pattern recognition tools and has hierarchical and non-hierarchical methods, contributes to the objectivity in this sense. Instead of relying on any expertise or personal interpretations, clustering methods provide a mathematical approach with the multivariate data set.

Hierarchical clustering is often portrayed as the better quality clustering approach, but is limited because of its quadratic time complexity[291,292]. Its quality may be observed if an appropriate distance metric can be defined to obtain the similarity, in this case a *distance matrix*. Some of the best known distances metric for distance measures are Minkowski Metric, Euclidean Metric, Manhattan Metric, and Mahalanobis Distance. Minkowski metric is a more general form where some others can be extracted from.

$$d(x,x') = (\sum_{k=1}^{d} | x_k - x_k^{'} |^q)^{1/q} \quad (12)$$

where $d(x,x')$ is the distance between x and x'.

The Euclidean metric is a particular case of Minkowski metric. In this case we have the distance as following:

$$d(x,x') = (\sum_{k=1}^{d} | x_k - x_k^{'} |^2)^{1/2} \quad (13)$$

Manhattan metric which can also be intuitively seen from Minkowski metric

is

$$d(x,x') = \sum_{k=1}^{-1} | x_k - x_k^{'} | \quad (14)$$

It is also known as taxicab distance.

Mahalanobis distance is

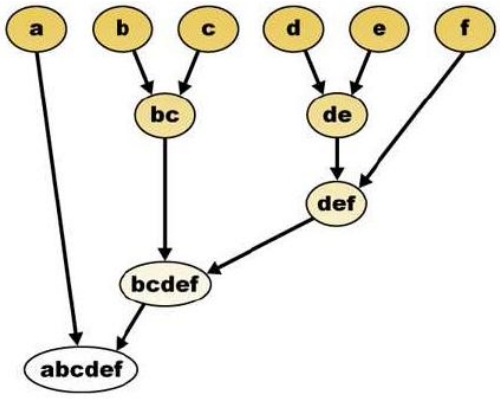
$$d(x, x') = (x - x')^t \sum^{-1} (x - x') \quad (15)$$

where $\sum^{-1}$ is the inverse of the covariance matrix [293]

There are two major types of hierarchical techniques: divisive and agglomerative. Agglomerative hierarchical techniques are used more commonly in microarray data analysis. The idea behind this set of techniques is to start with each cluster comprising of exactly one object and then progressively agglomerating (combining) the two nearest clusters until there is just one cluster left consisting of all the objects. Nearness of clusters is based on a measure of distance between clusters. All agglomerative methods require as input a distance measure between all the objects that are to be clustered. This measure of distance between objects is mapped into a metric for the distance between clusters (sets of objects) metrics for the distance between two clusters. The only difference between the various agglomerative techniques is the way in which this inter-cluster distance metric is defined[292]. Three graphical tools are often used to measure inter-clusters distances. These graphical methods are single linkage (distance between any two clusters is the minimum distance between two points such that one of the points is in each of the clusters.), complete linkage(distance between any two clusters is the minimum distance between two points such that one of the points is in each of the clusters) and average linkage methods (distance between any two clusters is the average distance between two points such that each pair has a point in both clusters)[294].

In this work, the hierarchical clustering was carried out by using a software, Cluster, developed by EisenLab, which provides variety of types of cluster analysis and data processing on large microarray datasets [160].

Figure 2-3 Example of agglomerative hierarchical clustering by using Single Linkage method



.

Agglomerative clustering starts from leaf which keeps on adding together untill it reaches to root. The Single Linkage method firstly determines and stores the distance between each pair of clusters. (Initially, each point is considered a cluster by itself) Also, for each cluster determine its nearest neighbor. The pair of clusters with the smallest distance between them can be determined and agglomerated. Finally, the system updates the pairwise distances and the new nearest neighbors. This process is repeated till only one cluster is left.

2.1.4 Parameters optimization and model validation

To each machine learning method, parameters that represent the best prediction performance must be optimized by using training data sets. In this work SVM is trained by using a Gaussian RBF function which has an adjustable parameter σ. The range of σ is estimated from the recommedate values of LibSVM, where σ=sqrt(number of features) [295].

For PNN, the only parameter to be optimized is a scaling parameter σ. Optimization of the parameter for each of these methods is conducted by scanning the parameter through a range of values. The set of parameters that produces the best prediction performance model can be determined by using test data sets or cross-validation methods, such as 5-fold cross-validation or leave-one-out cross-validation (LOOCV). Finally, an independent data can be used to further validate the generalization capability of selected models.

Overfitting, a major concern in machine learning, is recognized as a violation of Occam's razor[296]. In the process of overfitting, the performance on the training samples still increases, whereas the performance on the test data worsens. Overfitting occurs, for example, because the model is too complicated, or irrelevant descriptors are used, or the training sample is too small or its dimension is too high, and the learning process is too long [297]. It is somewhat suspicious that a learning method, without any mechanism to reduce the risk of overfitting, can achieve a robust performance. Various regularization methods such as penalization, selection and shrinkage can be employed to reduce the risk of overfitting [298]. In real life application, a simply but frequently used method of overfitting evaluation and prevention is to compare the prediction accuracies determined by using cross validation methods with those determined by using independent validation sets[297]. It is suggested that a model that is not overfitted should not have large differences in the estimation of its predictive capability between cross-validation methods and independent validation sets.

Microarray data are typically small size sample with high dimension descriptors, it is therefore overfitting need to be handled with caution during the period of gene selection. In this study, methods include LOOCV and validation with independent data set was used to parameters optimization and overfitting avoidance.

2.1.5 Performance evaluation

The performance of SVM can be measured as true positive TP (the number of positive examples which are correctly predicted as positive), false negative FN (the number of positive examples which are incorrectly predicted as negative), true negative TN (the number of negative examples which are correctly predicted as negative) and false positive

FP (the number of negative examples which are incorrectly predicted as positive) (Table 2-2).

The simplest way to evaluate the performance of a classification is overall accuracy (*Q*), which measures the proportion of the total number of the correctly predicted examples.

$$Q = \frac{TP + TN}{TP + FN + TN + FP} \tag{16}$$

Another three concepts, sensitivity (SE), specificity (SP) and Matthew's correlation coefficient MCC [299], which measure the positive and negative prediction performance respectively, are also frequently used in classification.

$$SE = \frac{TP}{TP + FN} \tag{17}$$

$$SP = \frac{TN}{TN + FP} \tag{18}$$

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \tag{19}$$

In some cases such as epidemiology and the evaluation of diagnostic tests [300], positive predictive value (PPV, also called precision rate) and negative predictive value (NPV) are very important.

$$PPV = \frac{TP}{TP + FP} \tag{20}$$

$$NPV = \frac{TN}{TN + FN} \tag{21}$$

Table 2-2 Relationships among terms of performance evaluation

		Condition		
		True	false	
Test outcome	Positive	True positive (TP)	False positive (FP)	→Positive predictive value (PPV)
	Negative	False negative (FN)	True negative (TN)	→Negative predictive value (NPV)
		↓ Sensitivity (SE)	↓ Specificity (SP)	

Typically, the performance of screening large data sets can be measured by the quantities of yield (similar to TP, the number of positive examples which are correctly predicted as positive), hit-rate (similar to PPV, percentage of virtual hits that are known positives), false-hit rate (similar to NPV, percentage of virtual hits that are known negatives), and enrichment factor (EF, magnitude of hit-rate improvement over random selection) [301].

$$EF = \frac{\text{hit rate}}{(TP + FN) / (TP + FN + TN + FP)} \tag{22}$$

2.2 Methodology for microarray data processing

2.2.1 Preprocessing of microarray data

The goal of data preparation in microarray data analysis is the same as for all data mining, which is to transform the data to make it suitable for analysis and to aid in producing the best possible models. Data preparation usually takes place in two stages. The first stage contains those aspects of data preparation which are independent of any class data; these are scaling normalization, thresholding, filtering, and missing data estimation. Thresholding and filtering are “low-level” forms of data cleaning; techniques of this kind

are broadly applicable, but the details will vary with the microarray device used to produce the data. The thresholding and filtering details given in this study are specific to the paper suggestions.

The second stage of microarray data processing contains aspects which make specific use of classification information of given data set, and is broadly termed as “feature selection”. Here we are performing a new data reduction method by narrowing the set of features to those relevant to the specific set of disease classes to be analyzed.

2.2.2 Normalization and scaling

Normalization is the attempt to compensate systematic experimental and/or technical variations affects the measured gene expression levels, so that biological difference can be easily distinguished and the comparison of expression levels across samples can be performed. Although normalization alone cannot control for all systematic variations, normalization plays an important role in the earlier stage of microarray data analysis because expression data can significantly vary from different normalization procedures. Subsequent analyses, such as differential expression testing would be more important such as clustering, and gene networks, though they are quite dependent on a choice of a normalization procedure[4,160]

Normalization methods differ for experimental platforms (eg. dual-labeled platforms versus single labeled platforms such as the Affymetrix GeneChip arrays) [302,303]. Yang et al.[303] summarized a number of normalization methods for dual labelled microarrays such as global normalization and locally weighted scatterplot smoothing (LOWESS) [304]. There have been some extensions for global and intensity-dependent normalizations. For example, Kepler et al. [305] considered a local regression to estimate

a normalized intensities as well as intensity dependent error variance. Wang et al.[306] proposed an iterative normalization of cDNA microarray data for estimating a normalized coefficients and identifying control genes. Workman *et al.*[307,308] proposed a roust non-linear method for normalization using array signal distribution analysis and cubic splines. Chen et al.[309] proposed a subset normalization to adjust for location biases combined with global normalization for intensity biases. Edwards [308] considered a non-linear LOWESS normalization in one channel cDNA microarrays mainly for correcting spatial heterogeneity. For affymetrix microarray data, many methods have been proposed in the literature for signal quantification, detection and normalization. Robust Multichip Average (RMA [310]) and Affymetrix Microarray Suite (MAS) 5.0 algorithms are two popular methods. MAS 5.0 yields scaled, background-subtracted, nonnormalized signal intensities, while RMA provides background-subtracted, log-transformed signal intensities. In this study, the approach used in microarray data normalization is conducted according to the platform properties. Data extraction was performed using affy library in BioConductor[311], an R-based package[312].

Molecular descriptors are usually scaled before they are used in machine learning methods. This is to ensure that all descriptors have equal potential to affect the prediction model. The similar procedure is also applied via processing the microarray data for biomarker discovery. Currently, a typical scaling procedure for microarray data is (1) normalizing the expression levels of each sample to zero-mean and unit variance, and then (2) normalizing the expression levels of each gene to zero-mean and unit variance over all the samples. This scaling method have been shown to perform well [313,314] and is applied in this thesis.

2.2.3 Threshold filtering

Since many genes are not expressed at all or do not vary sufficiently to be useful, a threshold filtering operation is usually applied before subsequent analyses, such as differential expression testing and clustering. Typical filtering 1) eliminates genes showing abnormal expression intensities, and 2) excludes genes showing minimal variation across samples.

Studies of expression measurement error, which process the same sample several times, have shown that the measurements were reproducible above values of 100, and much less reproducible below 100 [315]. For data believed to contain noise, a lower threshold of 100 would be appropriate. For classification problems that are very robust (e.g. distinguishing different types of brain tumors), we used a threshold of 100 units because there was a sufficiently large number of genes correlated with the distinction that the threshold could be set high, thereby minimizing noise, and maximizing potential biological interpretation of the marker genes. For the more subtle distinctions (e.g. outcome prediction), few correlates of the distinction are found, and for this reason the threshold was set at a lower level (20 units) so as to avoid missing any potentially informative marker genes. The upper threshold is usually 16,000 because this level has been observed as the fluorescence saturation of the scanner; values above this cannot be reliably measured and image intensities above this level tend to have a non-linear correlation with the actual expression levels. [316,317]

Gene variation filtering be realized through variation filter tests for a fold-change and absolute variation over samples, for example

$MaxValue(G) / MinValue(G) < 5$

$MaxValue(G) – MinValue(G) < 500$

where MaxValue(G) and MinValue(G) are the maximum and minimum values of gene G across all samples. If maxValue/minValue is lower than the specified ratio (5 in this case) and maxValue - minValue is lower than the specified difference (500 in the above dialog), then the gene are excluded from the resulting data set.

In this study, threshold filtering was performed accordingly with the purpose of the analysis. The precise parameters of the variation filtering for data sets of cancer biomarker identification are provided in method section of Chapter 6 (In most cases the variation filter excluded those where max/min < 5 and max – min < 500. In some cases more or less stringent values were used.)

2.2.4 Missing data estimation

Missing values is a common issue existing in microarray data. The missing values arise from experimental errors due to spotting problems (cDNA), dust, poor hybridization, inadequate resolution, fabrication errors (e.g. scratch) and image corruption [318,319]. They could also come from the suspicious data with low expression (e.g. background is stronger than signal) or censored data [117]. Repeating experiments could be a solution but often not be a realistic option because of economic reasons or limitations in biological material [170,320]. However, many microarray data analysis methods, such as classification, clustering and gene selection methods, require complete data matrix. Therefore in many microarray projects, one needs to determine how to estimate missing values. Proper missing value estimation can significantly improve performance of the analysis methods [321-323]. The simplest way is to remove all genes and arrays with missing values, or to replace missing values with an arbitrary constant (usually zero), row (gene) average or column (array) average. The better approaches had also been proposed such as k-nearest neighbors method (KNN) [323], least square methods (LS) [320,324],

and principal component analysis (PCA) [325,326]. Among these estimation methods, KNN is the most widely used and is also a standard method for missing value estimation currently [134,321,323].

The KNN-based method for missing value estimation involves selecting *k* neighbor genes with similar expression profiles to the target gene (the gene with missing values in one or more arrays), and estimating the missing value of the target gene in specific array as the weighted mean of the expression levels of the *k* neighbor genes in this array. A popular KNN-based method is KNNimpute [323], which is the only imputation method available in many microarray data analysis tools for missing value estimation [116,327,328]. KNNimpute can be downloaded from Stanford Microarray Database [134,329]. In this thesis, KNNimpute is employed if the microarray data contains missing values.

2.3 Feature selection procedure

Feature (gene) selection is a necessary step in supervised learning methodologies as many classifier algorithms cannot deal with thousands of input variables and require some type of dimensionality reduction or prior selection. Meanwhile, the problem of choosing statistically significant molecular markers or differentially expressed genes provides a subset of candidate predictor-genes for further validation with traditioned experimental technique (e.g. RT-PCR, immunochemistry, etc.). In this work, a new REF based biomarker selection procedure was developed by incooperating random sampling, feature elimination and consistency evaluation is described in detail in the following setions.

2.3.1 REF based gene selection procedure

A novel gene selection procedure method based on Support Vector Machines classifier, recursive feature elimination, multiple random sampling strategies and multi-step evaluation of gene-ranking consisitency was established (Figure 2-4): (1) After preprocessing the original data, by using random sampling method, a large number of training-test sample combinations are generated from the original data set.

(2) The large number of sample combinations is divided into *n* groups, and each group contains 500 sample combinations.

(3) In each training-test sample combination of each group, SVM and RFE are used to classify the samples (SVM classifiers) and rank the genes (RFE gene rank criteria). Therefore 500 gene ranking sequences are formed.

(4) The consistency evaluation can be performed based on the 500 sequences and a certain number of genes (for example, *k* genes) can be eliminated.

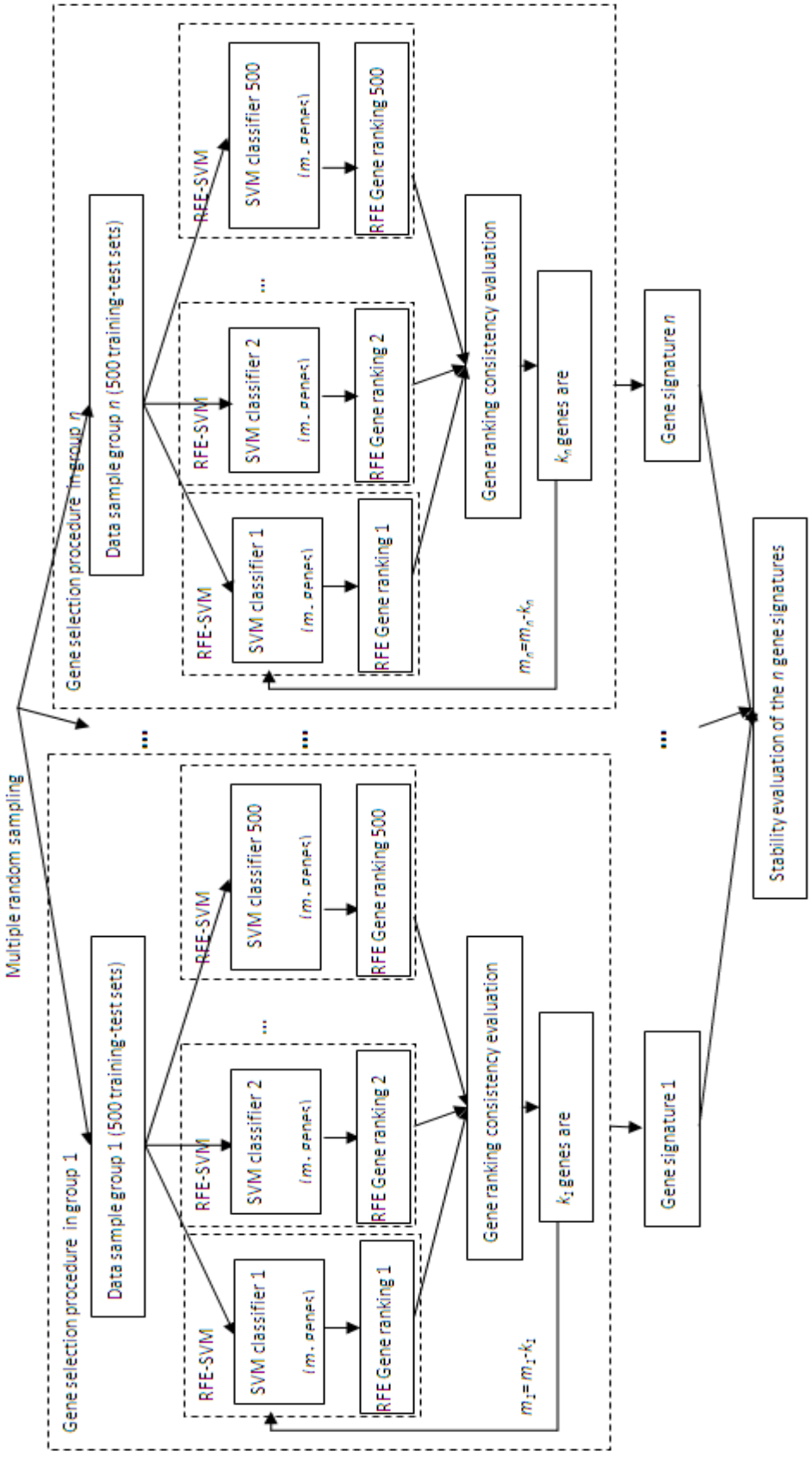
(5) Step (3) and (4) can be iteratively done until no gene can be eliminated.

(6) The gene subset which gives us the highest overall accuracies of the 500 test sample sets can be selected as gene signatures of this group. Using this method, we can obtain *n* gene signatures.

(7) The stability evaluation of the gene signatures can be performed by looking into the overlap gene rate of the *n* gene signatures.

In the following section Recursive feature elimination [330] is introduced first and followed by a detailed introduction of the whole feature selection procedure.

Figure 2-4 Overview of the gene selection procedure



2.3.2 Recursive feature elimination

During gene selection procedure, the genes are ranked according to their contribution to the SVM classifiers. The contributions of genes are calculated by Recursive feature elimination (RFE) procedure, which sort genes according to a gene-ranking function generated from SVM classifier. As described in Section 2.1, SVM training process needs to find the solution for the optimum problem (also known as objective function or cost function) shown in equation (14), which can be rewritten as

$$J = \frac{1}{2}\alpha^T H\alpha - \alpha^T 1 \tag{22}$$

Under the constraints $\sum_{i=1}^{n} a_i y_i = 0$ and $\alpha_i \geq 0$, i=1,2,…n.

Where $H(i, j) = y_i y_j K(x_i, x_j)$, K is the kernel function.

The gene-ranking function of RFE can be defined as the change in the objective function J upon removing a certain gene. When a given feature is removed or its weight w_k is reduced to zero, the change in the cost function $J(k)$ is

$$DJ(k) = \frac{1}{2}\frac{\partial^2 J}{\partial w_k^2}(Dw_k)^2 \tag{23}$$

where the change in weight $Dw_k = w_k - 0$ corresponds to the removal of feature k.

Under the assumption that the removal of one feature will not significantly influence the values of αs, the change of cost function can be estimated as

$$DJ(k) = \frac{1}{2}\alpha^T H\alpha - \frac{1}{2}\alpha^T H(-k)\alpha \tag{24}$$

Where *H* is the matrix with elements $y_i y_j K(x_i, x_j)$, and *H (-k)* is the matrix computed by using the same method as that of matrix *H* but with its *k*th component removed.

The change in the cost function indicates the contribution of the feature to the decision function, and serves as an indicator of gene ranking position [331].

2.3.3 Random sampling, feature elimination and consistency evaluation

In order to present statistical meaning, gene selection is conducted based on multiple random sampling. Each random sampling divide all microarray samples into a training set which contains half number of samples and an associates test set which contains another half number of samples. This sampling method can be treated as a special case of the bootstrap technique. Many researchers showed that bootstrap-related techniques present more accurate estimation than cross-validation on small sample sets [332,333]. By using this random sampling, thousands of training-test sets, each containing a unique combination of samples, are generated. These thousands of randomly generated training-test sets are randomly divided into several sampling groups, with equal number of training-test sets (such as 500 traing-test sets) in each group. Every sampling group is then used to derive a signature by RFE-SVM.

In every training-test sampling group generated by multiple random sampling, each training-set (totally 500 training-test sets) is used to train a SVM class-differentiation system and the genes are ranked by using Recursive feature elimination (RFE), according to the contribution of genes to the SVM classifier. In order to derive a gene-ranking criterion consistent for all iterations and all the 500 training-test sets in this group, a SVM class-differentiation system with a universal set of globally optimized parameters, which

give the best average class-differentiation accuracy over all of the 500 test sets in this group, is applied by RFE gene-ranking function at every iteration step and for every training-test set.

To further reduce the chance of erroneous elimination of predictor-genes, additional gene-ranking consistency evaluation steps are implemented on top of the normal RFE procedures in each group:

(1) For every training-set, subsets of genes ranked in the bottom (which give least contribution to the SVM classification procedure) with combined score lower than the first few top-ranked genes (which give highest contribution to the SVM classification procedure) are selected such that collective contribution of these genes less likely outweigh top-ranked ones;
(2) For every training-set, the step (1) selected genes are further evaluated to choose those not ranked in the upper 50% in previous iteration so as to ensure that these genes are consistently ranked lower;
(3) A consensus scoring scheme is applied to step (2) selected genes such that only those appearing in most of the 500 testing-sets were eliminated.

For each sampling group, different SVM parameters are scanned, various RFE iteration steps are evaluated to identify the globally optimal SVM parameters and RFE iteration steps that give the highest average class-differentiation accuracy for the 500 testing-sets.

The several signatures derived from these sampling-groups are then applied to evaluate the stability and performance.

2.4 Construction of the feature vector for peptide

2.4.1 Feature vector for peptide

Conventional computational methods, such as machine learning based approaches, have been widely applied in various protein studies. In protein classification problem, machine learning methods cannot accept the sequence information directly. It is essential to convert the sequences into numerical vectors as input. Various methods were proposed to construct a feature vector from protein sequence [238-244], most of which are straightforward by using original sequence directly or binary vector representation of the amino acid of the sequence. In this study, we constructed a digital feature vector from the encoded representations of tabulated residue properties including amino acid composition, hydrophobicity, normalized van der Waals volume, polarity, polarizability, charge, surface tension, secondary structure and solvent accessibility [254,274]. This well-formulated and high performance approach was firstly proposed by our research group [254,255]. Following are the detailed procedures of feature vector constructed for representing the structural and physicochemical properties of a given peptide.

Given the sequence of a protein, its amino acid composition and the properties of every constituent amino acid are computed according to the amino acid physical and chemical properties, and used to generate this vector. The computed amino acid properties include hydrophobicity, normalized Van der Waals volume, polarity, polarizability, charge, surface tension, secondary structure and solvent accessibility [254]. Amino acids are divided into three groups for each properties such that those in a particular group are regarded to have approximately the same property[254].

A typical peptide is mainly about 7~30 amino acids long in the case of MHC binding peptides study, which is quite shorter than protein sequence. For this reason, amino acids

are divided into three or even six groups in this study to improve the physical and chemical sensitivities of generated vectors. For instance, amino acids can be divided into hydrophobic (CVLIMFW), neutral (GASTPHY), and polar (RKEDQN) groups (Table 2-3). Three descriptors, composition (C), transition (T), and distribution (D), are introduced to describe global composition of each of these properties. C is the number of amino acids of a particular property (such as hydrophobicity) divided by the total number of amino acids in a protein sequence. T characterizes the percent frequency with which amino acids of a particular property is followed by amino acids of a different property. D measures the chain length within which the first, 25%, 50%, 75% and 100% of the amino acids of a particular property is located respectively. Overall, there are 51 elements representing these three descriptors: 6 for C, 15 for T and 30 for D, with a six-grouped property.

Overall, the feature vector of a peptide is constructed by combining the 51 elements of all of those 6-groups properties, 21 elements of those 3-groups properties and the 20 elements of amino acid composition in sequential order. A total of 308 elements are used as feature vector for each protein as shown in Table 2-4.

Table 2-5 gives the computed descriptors of the HLA-0201 binding peptide (VLFRGGPRGLLAVA [334]). The feature vector of a peptide is constructed by combining all of the descriptors in sequential order.

2.4.2 Scaling of feature vector

Molecular descriptors are usually scaled before they are used for machine learning methods. This is to ensure that all descriptors have equal potential to affect the prediction model. There are four main types of descriptor scaling: autoscaling, range scaling, feature

weighting and Pareto scaling (Livingstone 1995 and Eriksson 2001). Range scaling is the most common type of descriptor scaling methods used in various applications.

In range scaling, the minimum value of the descriptor is subtracted from the descriptor values and the resultant values are divided by the range:

$$A_i^{new} = \frac{2*(A_i - \overline{A_i})}{\max(A_i) - \min(A_i)} - 1 \qquad (23)$$

where $\min(A_i)$ and $\max(A_i)$ are the minimum and maximum value of descriptor *i* respectively. The range-scaled descriptors have a minimum and maximum value of -1 and 1 respectively. Range scale can be carried out over any preferred range by multiplication of the range-scaled values by a factor. The disadvantage of range scaling is that it mainly depends on the minimum and maximum values of the descriptors, which makes it very sensitive to outliers.

Table 2-3 Division of amino acids into different groups for different physicochemical properties.

Property		**Group 1**	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**
Hydro-phobicity	Type	0~0.043	0.165~0.359	0.45~0.501	0.616~0.738	0.825~0.88	0.943~1
	Amino Acids in Group	RDE	HNQKS	TG	ACPM	VWY	ILF
Van der Waals volume	Value	0~1.6	2.43~2.78	2.95~3	3.78~4.0	4.43~4.77	5.89~8.08
	Amino Acids in Group	GAS	CTPD	NV	EQIL	MHK	FRYW
Polarity	Value	0	0.352~0.456	0.6~0.608	0.648~0.696	0.792~0.8	0.864~1.0
	Amino Acids in Group	VI	LFWCMY	PA	TGS	HQR	KNED
Polarizability	Value	0~0.062	0.105~0.108	0.128~0.15	0.18~0.186	0.219~0.23	0.29~0.409
	Amino Acids in Group	GAS	DT	CPNVE	QIL	KMH	FRYW
		Group 1		Group 2		Group 3	
Charge	Type	Positive		Neutral		Negative	
	Amino Acids in Group	KRH		ANCQGHILMFPSTWYV		DE	
Surface tension	Value	-0.20~0.16		-0.3~ -0.52		-0.98~ -2.46	
	Amino Acids in Group	GQDNAHR		KTSEC		ILMFPWYV	
Secondary structure	Type	Helix		Strand		Coil	
	Amino Acids in Group	EALMQKRH		VIYCWFT		GNPSD	
Solvent accessibility	Type	Buried		Exposed		Intermediate	
	Amino Acids in Group	ALFCGIVW		RKQEND		MPSTHY	

Table 2-4 List of features for peptides

Feature Description	Number of group	Number of dimensions
Amino acid composition	-	20
Hydrophobicity	6	51
Van der Waals volume	6	51
Polarity	6	51
Polarizability	6	51
Charge	3	21
Surface tension	3	21
Secondary structure	3	21
Solvent accessibility	3	21
Total	-	308

Table 2-5 Computed descriptors of the HLA-0201 biding peptide (sequence: VLFRGGPRGLLAVA [334]). The feature vector of a peptide is constructed by combining all of the descriptors in sequential order.

Property	Elements of descriptors									
Amino acid composition	14.29	0.00	0.00	0.00	7.14	21.43	0.00	0.00	0.00	21.43
	0.00	0.00	7.14	0.00	14.29	0.00	0.00	14.29	0.00	0.00
Hydro-phobicity	14.29	0.00	21.43	21.43	14.29	28.57	0.00	15.38	7.69	0.00
	7.69	0.00	0.00	0.00	0.00	7.69	0.00	7.69	15.38	7.69
	7.69	28.57	28.57	28.57	28.57	57.14	0.00	0.00	0.00	0.00
	0.00	35.71	35.71	35.71	42.86	64.29	50.00	50.00	50.00	85.71
	100.00	7.14	7.14	7.14	7.14	92.86	14.29	14.29	21.43	71.43
	78.57									
Van der waals volume	35.71	7.14	14.29	21.43	0.00	21.43	7.69	15.38	15.38	0.00
	15.38	0.00	0.00	0.00	7.69	7.69	0.00	0.00	0.00	7.69
	0.00	35.71	35.71	42.86	64.29	100.0	50.00	50.00	50.00	50.00
	50.00	7.14	7.14	7.14	7.14	92.86	14.29	14.29	14.29	71.43
	78.57	0.00	0.00	0.00	0.00	0.00	21.43	21.43	21.43	28.57
	57.14									
Polarity	14.29	28.57	21.43	21.43	14.29	0.00	7.69	15.38	0.00	0.00
	0.00	7.69	7.69	7.69	0.00	7.69	7.69	0.00	15.38	0.00
	0.00	7.14	7.14	7.14	7.14	92.86	14.29	14.29	21.43	71.43
	78.57	50.00	50.00	50.00	85.71	100.00	35.71	35.71	35.71	42.86
	64.29	28.57	28.57	28.57	28.57	57.14	0.00	0.00	0.00	0.00
	0.00									
Polarizability	35.71	0.00	21.43	21.43	0.00	21.43	0.00	23.08	15.38	0.00
	15.38	0.00	0.00	0.00	0.00	7.69	0.00	7.69	0.00	7.69
	0.00	35.71	35.71	42.86	64.29	100.0	0.00	0.00	0.00	0.00
	0.00	7.14	7.14	7.14	50.00	92.86	14.29	14.29	14.29	71.43
	78.57	0.00	0.00	0.00	0.00	0.00	21.43	21.43	21.43	28.57
	57.14									
Charge	14.29	85.71	0.00	30.77	0.00	0.00	28.57	28.57	28.57	28.57
	57.14	7.14	21.43	50.00	78.57	100.0	0.00	0.00	0.00	0.00
	0.00									
Surface tension	50.00	0.00	50.00	0.00	53.85	0.00	28.57	28.57	42.86	64.29
	100.00	0.00	0.00	0.00	0.00	0.00	7.14	7.14	21.43	71.43
	92.86									
Secondary structure	50.00	21.43	28.57	38.46	30.77	0.00	14.29	14.29	57.14	78.57
	100.00	7.14	7.14	7.14	21.43	92.86	35.71	35.71	42.86	50.00
	64.29									

Solvent accessibility	78.57	14.29	7.14	23.08	7.69	7.69	7.14	14.29	42.86	78.57
	100.00	28.57	28.57	28.57	28.57	57.14	50.00	50.00	50.00	50.00
	50.00									

3 Mechanisms of drug combinations: interaction and network perspectives

Combination therapies are increasingly becoming an important part of modern disease therapy. By exploiting targeted, mechanism-based treatments through the use of rational combinations, the personalized treatment and increased treatment efficacy can be achieved. Understanding the molecular mechanisms underlying synergistic, potentiative and other contrasting combinations can greatly facilitate the discovery of efficacious combinations and multi-targeting agents as well. In this study, we performed an extensive investigation of 117 drug combinations for which the combination effect has been evaluated by rigorous analysis methods and for which relevant molecular interaction (MI) profiles of the drugs involved are available. Seven general and specific modes of action are summarized based on the MI profiles and pathway analyzing of identified drug combinations. The study also highlights the potential value of molecular interaction profiles and network regulation in the discovery of novel multicomponent therapies.

3.1 Introduction

In recent years, drug discovery efforts have primarily focused on identifying agents that modulate preselected individual targets[335-337]. While new drugs have continuously been discovered, there is a growing productivity gap, despite major spending on research and development and advances in technology development[338]. This problem arises partly because agents directed at an individual target frequently show limited efficacies and poor safety and resistance profiles, which are often due to factors such as network robustness[24-26], redundancy[27], crosstalk[28-30], compensatory and neutralizing actions[31,32], and anti-target and counter-target activities[33]. With such issues in mind,

systems-oriented drug design has been increasingly emphasized[60-63] and supported by several clinical successes of multicomponent therapies that use drug-combinations and multi-targeting agents[54,339-341].

Increasing interests and efforts have been directed at the discovery of new multicomponent therapies[26,54,60-62,342,343]. However, uncovering drug combinations by direct screening is quite challenging due to the large number of potential combinations. A recent high-throughput screen was able to systematically test about 120,000 different two-drugs combinations[58]. Yet, the campaign of NCI60 anticancer drug screen count with a stock of above 100,000 potential therapeutic agents[344] , resulting in more than 5×10^9 drugs pairs. The situation is worse when we address combinations that consist of more than two drugs. More important, assuming that most drug combinations will not improve significantly over single drugs, attempting such high-throughput screens is highly inefficient. Understanding the molecular mechanisms underlying the known synergistic, potentiative and other contrasting combinations is therefore important to provide general guideline for efficacious experimental design.

Knowledge of the molecular mechanisms of currently explored multicomponent therapies provides useful clues for discovering new drug combinations and multi-targeting agents. Some key characteristics of the modes of multicomponent therapies have been outlined [33,54,62,342]. The multiple targets can reside in the same or different pathways and tissues. Their modulation produces more-than-additive (synergistic) effects triggered by actions converging at a specific pathway site, and by negative regulation of network compensatory and neutralizing responses, drug resistance sources, and anti-target and

counter-target activities. However, the exact mechanism has been fully elucidated for few of the explored drug combinations[62,345-350].

Advances in the comprehension of biological systems, driven by genomics and proteomics, have recently provided a framework in which preclinical scientists can predict the biological responses resulting from the modulation of multiple independent targets in combination. These advances provide a change in extensive investigations of the molecular basis of drug actions and responses, yielding a substantial amount of information on experimentally determined drug-mediated molecular interaction (MI) profiles and regulatory activities of many drugs and compounds[65,335,336,351-355]. The MI profile of a drug describes its interactions with individual biomolecules, pathways or processes attributable to its pharmacodynamic, toxicological, pharmacokinetic, and combination effects. Apart from using them for guiding the development of target discovery technologies[356-362], MI profiles may be explored for gaining further insights into general modes of actions of multicomponent therapies and the mechanism of specific drug combinations. Such a task may be accomplished by analyzing the relevant MI profiles from the perspective of coordinated interactions and network regulations[29-31].

In this work we introduce a systematic framework to analyze effective drug combinations by comprehensively investigating literature reported synergistic and other types of drug combinations. The action modes of some of combination therapies and the multi-target therapeutics that are currently on the market or in development were evaluated based on the identified MI profiles and pathways analysis. This pathway-focused approach to target discovery could help lead to a greater understanding of disease biology, helping outline

some of the important aspects of the discovery of multi-targeting therapeutics using compounds and cell-based *in vitro* assays.

3.2 Materials and Methods

3.2.1 Mechanism of drug interactions

Often, multicomponent therapies are developed and evaluated from cell-based *in vivo* test via acting simultaneously at multiple molecular targets. Some terminology, such as synergism, additivity, antagonism, independence, inertism and coalism, are commonly used to describe the mechanism of joint action of two agents in these cell-based *in vivo* tests[363]. When two drugs produce the same broad therapeutic effect, their combination produces the same effect of various magnitudes compared with the summed effects of the individual drugs. A combination can be pharmacodynamically synergistic, additive or antagonistic if the effect is greater than, equal to, or less than the summed effects of the partner drugs [364]. Drug combinations may also produce pharmacokinetically potentiative or reductive effects such that the therapeutic activity of one drug is enhanced or reduced by another drug via regulation of its absorption, distribution, metabolism and excretion (ADME) [364]. A further type of drug combination is a coalistic combination, in which all of the drugs involved are inactive individually but are active in combination[363,365-367].

Synergistic and potentiative drug combinations have been explored to achieve one or more favourable outcomes: enhanced efficacy; decreased dosage at equal or increased level of efficacy; reduced or delayed development of drug resistance; and simultaneous enhancement of therapeutic actions and reduction of unwanted actions (efficacy synergism plus toxicity antagonism)[54,62,367]. The mechanisms underlying these activities can be

better understood by studying the mechanistically contrasting additive, antagonistic and reductive drug combinations.

3.2.2 Methods for drug-combination analysis

Attempts have been made during the past century to quantitatively measure the dose-effect relationships of each drug alone and its combinations and to determine whether or not a given combined treatment would gain a synergistic effect [363,364,368]. We have carefully chosen 7 rigorous drug-combination analysis methods, which have been developed and extensively used for analyzing combinations from experimental data[54,60,369]. These include checkerboard, combination index, fractional effect analysis, isobolographic analysis, interaction index, median drug effect analysis, and response surface approach[369-373].

3.2.3 Collection of literature-reported drug combinations

PubMed remains an optimal tool in biomedical electronic research, which is accessed for free and offers optimal update frequency and includes online early articles[374]. We therefore searched Pubmed[375] to select literature-reported drug combinations that had been evaluated by rigorous combination analysis methods and for which relevant MI profiles were retrievable from Pubmed. Combinations of the keywords “drug combination”, “drug interaction”, “multi-drug”, “additive”, “antagonism”, “antagonistic”, “infra-additive”, “potentiated”, “potentiative”, “potentiation”, “reductive”, “supra-additive”, “synergism”, “synergistic”, and “synergy” were used to search publications since 1999. Coalistic drug combinations were not evaluated because few of them are described in the literature. This is partly due to the focus on combinations of drugs that include at least one active drug; indeed, a Medline search using “coalistic” and “coalism”

returns only one abstract. In addition, a significantly higher percentage of the studies published before 1999 are based on non-rigorous drugcombination methods. It has been suggested that analysis without using a rigorous method may easily lead to errors in assessing synergism with respect to such effects as enhancement and potentiation[369]. Therefore, to maintain the level of reliability of our assessment without substantially losing statistical significance, we focused on studies published since 1999, which constitute approximately 50% of all abstract entries selected by using our search method.

3.3 Results and discussion

3.3.1 Statistics of collected drug combinations and MI profiles

We collected 315, 88 and 62 abstract entries describing pharmacodynamically synergistic, additive, and antagonistic combinations, respectively, and 56 and 33 abstract entries describing pharmacokinetically potentiative and reductive combinations, respectively. We then removed 158, 53, 32, 15 and 18 of these entries, respectively, that are redundant (for example, the same combination or the same paper selected by different keyword combinations); ambiguous (for example, synergistic in one report or condition, additive in another report or condition); and involving more than two drugs so as to focus on simpler cases. We further removed 45, 12, 1, 1 and 2 papers, respectively, that described studies using nonrigorous drug-combination methods. For the remaining 217 papers, we searched additional literature for experimentally determined MI profiles related to the mechanism of the claimed combination effects. Our analysis showed that the available literature-reported MI profiles are insufficient or irrelevant to substantiate the claimed combination effects in 110 (59 synergistic, 11 additive, 17 antagonistic, 20 potentiative and 3 reductive combinations) of the 217 remaining papers. This led to the identification of 107 combinations that can be substantiated by available literature-reported MI profiles. These

comprise 53, 12 and 12 sets of pharmacodynamically synergistic, additive and antagonistic combinations, and 20 and 10 sets of pharmacokinetically potentiative and reductive combinations, respectively. Examples of our evaluated drug combinations are shown in Table 3-1 to Table 3-7. Full data are summarized in Appendix Table S1 to Table S7, together with literature-reported mechanisms related to their therapeutic and combination effects. The statistical significance of our assessment can be roughly estimated as follows: for the 110 combination sets that are not substantiated by the available MI profiles, it is reasonable to assume a high percentage of these combinations may eventually be substantiated by additional experimental findings. If one further assumes that the reported combination effects substantiated by MI profiles are at least partly true, then the estimated ratio of truly and falsely reported combinations should be substantially larger than 107 out of 110. Hence, there seems to be a statistically significant number of combinations and sufficient percentages of true claims for supporting a fair assessment of general combination types and mechanisms of drug combinations from the information collected by our search methods.

3.3.2 Mechanism underlying the pharmacokinetic and pharmacodynamic drug interactions

The general and specific modes of action of 4 pharmacokinetic interactions and 3 pharmacodynamic interactions were studied based on the 117 drug combinations. These include pharmacodynamically synergistic drug combinations due to anti-counteractive actions, complementary actions, and facilitating actions, pharmacodynamically additive drug combinations, pharmacodynamically antagonistic, potentiative and reductive drug combinations, which are described separately in the following sections.

The use of multiple drugs with different mechanisms or modes of action may direct the different combined-actions and therapeutic benefits. The possible favorable or undesired outcomes can be explained by analyzing the relevant MI profiles of drugs from the perspective of coordinated interactions and network regulations. Many of the MI profiles directly point to a specific bimolecular as the inhibiting, activating or regulating target. Therefore, it is possible to determine the combination effects based on the expected therapeutic and pharmacokinetic consequences of these interactions. Although the molecular target is not exactly specified, some of the profiles identify a specific pathway or process as a target, and provide the pharmacodynamic or pharmacokinetic consequence of the interaction. For instance, in literature reports, arsenic trioxide produces anticancer activity by generating reactive oxygen species, which is partially counteracted by its activation of the AKT survival pathway[376]. The anticancer agent 17-(allylamino)-17-demethoxygeldanamycin (17-AAG) produces its effects by inhibiting the nuclear factor-κb (NF-κb), AP-1 (also known as Jun) and phosphatidylinositol 3-kinase(PI3K)–AKT pathways61. Therefore, when used in combination, 17-AAG abrogates arsenic trioxide's counteractive activation of AKT survival pathway[376].

3.3.2.1 ***Pharmacodynamically synergistic combinations***

It is known that pharmacodynamically synergistic may arise from anti-counteractive action, complementary action, and facilitating action[377]. Anti-counteractive actions reduce network's counteractive activities against a drug's therapeutic effect. Complementary actions positively regulate a target or process by interactions with multiple target/pathway sites, different target subtypes and states, and competing mechanisms[60]. Facilitating actions are secondary actions of one drug in enhancing the activity or level of another drug. The 53 sets of synergistic drug-combinations arise from anti-counteractive (21 sets), complementary (26 sets), and facilitating (6 sets) actions of

the drugs were involved in this stduy. The therapeutic and synergistic mechanism of the sets in these three combination groups are summarized in Appendix Table S1, Table S2, and Table S3 (examples are given in Table 3-1, Table 3-2 and Table 3-3) respectively.

Anti-**c**ounteractive actions may arise from interactions against anti-targets or counter-target[33] and from negative modulations of network's robustness[24-26], crosstalk[28-30], and compensatory and neutralizing actions[31,32]. These anti-counteractive synergistic combinations act on different targets of related pathways (8 sets), different targets of corss-talking pathways (4 sets), different targets of the same pathway that crosstalk to each other via another pathway (1 set) or regulate the same (5 sets) or different targets (2 sets), and different sites of the same target (1 set). An example of actions on different targets is the anticancer cisplatin-topotecan combination[378-380]. Cisplatin binds to the major groove of GG, AG and TACT sites in DNA[381], which is bypassed by network's counteractive activity of mutagenic translesional bypass replication across cisplatin-DNA adducts[382]. Topotecan inhibits topoisomerase I, interacts with DNA, and stabilizes covalent topoisomerase-DNA complex to block DNA replication forks[383]. The last function reduces the counteractive effect against cisplatin, resulting in synergism. An example of actions on the same target is the anticancer cisplatin-trabectedin combination[384]. Trabectedin interacts with DNA and its repair systems in a way different from cisplatin[384] via covalent binding to the 2-amino group of the central guanine of selected DNA pyrimidine-G-G and purine-G-C triplets[385], which induces the formation of unusual DNA replication intermediates that strongly inhibits DNA replication[386] which subsequently reduces the counteractive effect against cisplatin.

Complementary actions primarily involve positive regulation of a target or process by targeting multiple points of a pathway[387,388] and its crosstalk pathways[387-391], interacting with multiple sites[381,392], states[393], conformations[60], and mutant forms[60] of the target, collectively modulating target activity and expression[348], and simultaneously enhancing the positive and reducing the negative effects of the target[394,395]. These combinations act on different targest of the related pathways that regualated the same targets (8 sets) or the same target/process (5 sets), different targets of related pathways that regulate different targets (6 sets), different targets of the same pathway that regulate the same target (2 sets), different target subtypes in related pathways (1 set), and the same target at different sites (2 sets), overlapping sites (1 set), and different states (1 set). An example of actions on different targets is the elecoxib-emodin combination, which synergistically represses the growth of certain cancer cells[396]. Celecoxib, a COX-2 inhibitor, suppresses cancer growth by inactivating protein kinase Akt to stop its suppression of apoptosis[397]. Emodin suppresses cancer growth by inhibiting tyrosine kinase[398] and down-regulating Akt via inhibition of PI3K pathway to reduce Akt suppression of apoptosis[399]. The later complements celecoxib's inactivation of Akt[397] to reduce its suppression of apoptosis.

Facilitating actions can be illustrated by two examples. One is the gentamicin-vancomycin combination that produces synergistic antibacterial action against penicillin-resistant bacterial strains[400]. Gentamicin targets bacterial ribosome, causes misreading of the genetic code and inhibits translocation to disrupt protein synthesis[401]. Vancomycin inhibits bacterial cell wall peptidoglycan synthesis[402] and alters permeability of cell membrane and selectively inhibits ribonucleic acid synthesis[403]. The later enhances gentamicin penetration into bacterial cells thereby increasing its bioavailability. The

second example is the BQ-123 and enalapril combination that produces synergistic endothelium-dependent vasodilation enhancement[404]. BQ-123 is an ETA receptor antagonist that mediates vasodilatation[405], enalapril up-regulates ETB as well as inhibits ACE leading to vasodilation[406,407]. BQ-123 antagonism of ETA[405] displaces endogenous ET-1 from ETA onto upregulated ETB to enhance its activity by effectively increasing ETB agonist concentration[404].

Table 3-1 Examples of literature reported pharmacodynamically synergistic drug combinations due to anti-counteractive actions, in which synergy has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed

Combination target relationship	**Drug A (mechanism of actions related to synergy)**	**Drug B (mechanism of actions related to synergy)**	**Reported synergistic effect**	**Synergism determination method**	**Possible mechanism of synergism in anti-counteractive actions**
Different targets of the same pathway	Oxaliplatin (DNA adduct, preferably bind to major groove of GG, AG and TACT sites, complex conformation different from that of cisplatin[408], caused DNA strand break and non-DNA initiated apoptosis[409])	Irinotecan (DNA topoisomerase I inhibitor, increased EGFR phosphorylation in Lovo & WiDR cells[410])	Synergistic anticancer effect in AZ-521 and NUGC-4 cells, additive effect in MKN-45 cells[411]	Median drug effect analysis	Effect of oxaliplatin's DNA adduct formation[408] may be partially reduced by certain mutant DNA topoisomerase I acting on DNA adduct to generate different topoisomers[412]. Irinotecan inhibition of DNA topoisomerase I[410] partially off-sets this counteractive activity
Different targets of the same pathway that regulated the same target	DL-cycloserine (bacterial cell wall synthesis inhibitor[413])	Epigallocatechin gallate (disrupted integrity of bacterial cell wall via direct binding to peptidoglycan[413])	Synergistic effect on bacterial cell wall[413]	Fractional inhibitory concentration index	Cell wall alteration may induce counteractive cell wall synthesis to restore cell wall integrity[414], DL-cycloserine inhibition of cell wall synthesis hindered the restoration thereby enhanced Epigallocatechin gallate's cell wall disruption activity
Different targets of related pathways	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395] distorted mitosis to trigger apoptosis[415], induced p53 and CDK inhibitors[416], activated caspase-10, caspases-8, -6, and -3, leading to apoptosis[417], activated ERK [418] and CDK2[419], activated p38 MAP kinase and p53[420])	NU6140 (CDK inhibitor, down-regulated antiapoptotic protein survivin[421])	Synergistic apoptotic response[421]	Median drug effect analysis	Use of both drugs promoted complementary apoptosis activities via triple actions of surviving down-regulation by NU6140[421], microtubule stabilization[395] and caspase activation[417] by paclitaxel. Paclitaxel's promotion of apoptosis may be partially offset by its counteractive pro-growth activation of ERK[418] and CDK2[419], which may be partially reduced by NU6140's inhibition of CDK[421]
Different targets of cross-talking pathways	Gefitinib (EGFR tyrosine kinase inhibitor, induced cyclin-dependent kinase inhibitors p27 and p21, decreased MMP-2 and MMP-9 enzyme activity[422])	Taxane (disrupted microtubule by binding to beta-tubulin[423], induced tumor suppressor gene p53 and cyclin-dependent kinase inhibitors P21, down regulated Bcl-2, leading to apoptosis[416])	Strong synergistic effect in breast cancer MCF7/ADR cells[424]	Combination index	Taxane produced anticancer effect by inducing apoptosis[416] and microtubule disruption[423]. Crosstalk between EGFR and hypoxia-inducible factor-1alpha pathways increased resistance to apoptosis by up-regulating survivin[28]. Gefitinib produced anticancer effect via EFFR tyrosine kinase inhibition, which offsets the counteractive EGFR-hypoxia crosstalk in resisting taxane's pro-apoptosis activity
Different targets in the same pathway that crosstalks via	Gefitinib (EGFR tyrosine kinase inhibitor, induced cyclin-dependent kinase inhibitors p27 and p21, decreased MMP-2 and MMP-9 enzyme	PD98059 (MEK inhibiton[425])	Synergistic antitumor effect in breast cancer MDA-MB-361 cells186	Combination index, isobolographic analysis	An autocrine growth loop critical for tumor growth is formed in EGFR-Ras-Raf-MEK-ERK network such that activated MEK activates ERK which upregulates EGFR ligands which promotes the autocrine growth

other pathway	activity[422])				loop[426]. This loop produced counteractive activity against gefitinib or PD98059 by reducing the effect of MEK or EGFR tyrosine kinase inhibition. Simultaneous use of both drugs helps disrupting this autocrine growth loop, thereby enhancing each other's effect
Same target (different sites)	AZT (HIV-1 reverse transcriptase inhibitor[427])	Non-nucleoside HIV-1 reverse transcriptase inhibitor[428]	Antiviral synergism[429]	Isobolographic analysis, Yonetani & Theorell plot	AZT resistance is partly due to phosphorolytical removal of the AZT-terminated primer[430], NNRTI inhibited RT catalyzed phosphorolysis, thereby reduced AZT resistance[429]

Table 3-2 Examples of literature reported pharmacodynamically synergistic drug combinations due to complementary actions, in which synergy has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed

Combination target relation	**Drug A (mechanism of actions related to synergy)**	**Drug B (mechanism of actions related to synergy)**	**Reported synergistic effect**	**Synergism determination method**	**Possible mechanism of synergism in promoting complementary actions**
Different targets of the related pathways that regulate the same targets	Azithromycin (hindered bacterial protein synthesis by binding to 50S component of 70S ribosomal subunit[431])	Ceftazidime (blocked penicillin-binding proteins and thus bacterial cell wall synthesis[432])	Synergistic antibacterial effect[433]	Checkerboard method, fractional inhibitory concentration	Hindered protein synthesis by azithromycin[431] reduces penicillin-binding proteins to complement ceftazidime's blocking of penicillin-binding proteins[432]
Different targets of the related pathways that regulate the same process	Aplidin (induced apoptosis by activating and clustering death receptors of FasL[434], activating JNK, EGFR, Src, and p38MAPK[435], inhibited VEGF release and secretion[436])	Cytarabine (DNA binder[437], inhibited synthesome associated DNA polymerase alpha activity[438], inhibited RNA synthesis and DNA repair that lead to increased cellular stress and reduced survival protein Mcl-1 which subsequently activate caspase and apoptosis[439])	Aplidin synergizes with cytarabine in exhibiting anticancer activities in leukemia and lymphoma models in vitro and in vivo[440]	Chou-Talelay combination index (Calcusym Biosoft)	Both drugs complement each other's activity by inducing apoptosis via each of the two major cascades of apoptosis pathway, aplidin activated and clustered death receptors of FasL[434] which subsequently activates the receptor-mediated extrinsic cascade [441], cytarabine increased cellular stress and reduced survival protein Mcl-1 [439] which subsequently triggers the mitochondrial intrinsic cascade [441]
Different targets of the same pathway that regulate the same target	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	Tubacin (histone deacetylase 6 inhibitor, inhibited microtubule associated alpha-tubulin deacetylase activity[442])	Synergistically enhanced tubulin acetylation[394]	Combination index (Calcusym)	Both drugs complement each other's microtubule stabilization effects through enhanced acetylation activity of alpha-tubulin by paclitaxel[395] and reduced deacetylation activity of alpha-tubulin deacetylase by tubacin[442]
Different	Gefitinib (EGFR tyrosine kinase	ST1926 (activated MAP kinases p38	Synergistic	Combination	Gefitinib 's inhibition of EGFR is complemented by

targets of related pathways that regulate the same target	inhibitor, induced cyclin-dependent kinase inhibitors p27 and p21, decreased MMP-2 and MMP-9 enzyme activity[422])	and JNK, released cytochrome c, activated caspase proteolytic cascad[443])	modulation of survival signaling pathways[444]	Index	ST1926's activation of MAP kinases p38[443] that subsequently mediates internalization of EGFR[445], and by ST1926's activation of caspase proteolytic cascade[443]
Different targets of related pathways	Sildenafil (phosphodiesterase-5 inhibitor[446])	Iloprost (prostacyclin receptor agonist leading to vascular relaxation[447], activated phospholipase C [448] that promoted VEGF-induced skin vasorelaxation [449], self-regulated endothelial cell adhesion molecules[450])	Synergistic action to cause strong pulmonary vasodilatation[451]	Dose effect curve surpassed that of individual drug alone combined	Sildenafil produced vasodilation activity by inhibiting phosphodiresterase-5[446], iloprost produced vasodilation activity by agonizing prostacyclin receptor[447] and by activating phospholipase C[448]. Targeting of multiple vasodilatation regulation pathways NO/cGMP[452], MaxiK channel -mediated relaxation[453], and phospholipase C[448] contribute to the synergistic actions.
Different target subtypes of related pathways	Dexmedetomidine (alpha2A receptor agonist, produced antinociceptive effect via an endogenous sleep-promoting pathway[454])	ST-91 (agonist of alpha2 receptor of other subtypes, produced antinociceptive effect via upraspinal receptors and at both spinal and brainstem levels of the acoustic startle response pathway[455])	Synergistic antinociceptive action[345,456]	Isobolographic analysis	ST-91 modulated spinal and supraspinal pathways[455] that regulate pain[390] , dexmedetomidine promoted sleepiness[454] that sustains reduction in spike activity of spinoreticular tract neurons[389]
Same target (different states)	Mycophenolate mofetil (inosine monophosphate dehydrogenase inhibitor, drug metabolite mycophenolic acid binds to the site of nicotinamide adenine dinucleotide cofactor[393])	Mizoribine (inosine monophosphate dehydrogenase inhibitor, drug metabolite mizoribine monophosphate binds to the enzyme in transition state having a new conformation[457])	Mild synergistic suppression of T and B cell proliferation[458]	Median drug effect analysis, Combination index	Simultaneous inhibition of enzyme reactant-state and transition state have the advantage of covering more conformational space for the inhibitors to better compete with natural substrates for the binding sites.
Same target (overlapping binding sites)	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	Discodermolide (stabilized microtubule dynamics and enhanced microtubule polymer mass[459] resulting in aberrant mitosis that triggers apoptosis [415] and induced p53 and CDK inhibitors[416], retained antiproliferative activity against carcinoma cells resistant to paclitaxel due to beta-tubulin mutations[460])	Antiproliferative synergy[461]	Combination index	Explanation 1: Binding sites of both drugs overlapping, certain mutations resistant to one drug are ineffective against the other, thereby covering more diverse range of mutant types[60,340,462]. Explanation 2: One drug binds and induces conformational change in tubulin that increases the binding affinity of the other[60,463]. Explanation 3: These drugs may differentially bind to or affect different tubulin subtypes, microtubule architectures, or microtubule regulators, thereby covering more diverse range of microtubule dynamics[57,60,463,464]
Same target	Paclitaxel (stabilized microtubules	Peloruside A (binds at a different site	Peloruside A	Berenbaum’s	Explanation 1: Binding sites of both drugs

(different binding sites)	via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	from that of paclitaxel, stabilized microtubules via binding to a unique site on the tubulin alpha, beta heterodimer[465])	synergizes with paclitaxel to enhance the antimitotic action of the drugs[465]	combination index	overlapping, certain mutations resistant to one drug are ineffective against the other, thereby covering more diverse range of mutant types[60,340,462]. Explanation 2: One drug binds and induces conformational change in tubulin that increases the binding affinity of the other[60,463]. Explanation 3: These drugs may differentially bind to or affect different tubulin subtypes, microtubule architectures, or microtubule regulators, thereby covering more diverse range of microtubule dynamics[57,60,435,463]

Table 3-3 Examples of literature reported pharmacodynamically synergistic drug combinations due to facilitating actions, in which synergy has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed

Combination target relationship	**Drug A (mechanism of actions related to synergy)**	**Drug B (mechanism of actions related to synergy)**	**Reported synergistic effect [Ref]**	**Synergism determination method**	**Possible mechanism of synergism in promoting facilitating actions**
Different targets of related pathways	Ampicillin (blocked penicillin-binding protein 2A and thus bacterial cell wall synthesis[466])	Daptomycin (disrupted bacterial membrane structure[467])	Significant antibacterial synergy[347]	Checkerboard method, fractional inhibitory concentration	Most penicilling-binding proteins are associated with membrane[468], membrane disruption by daptomyci [467] likely hinders the functions of penicilling-binding proteins and further expose them to ampicillin binding
Different targets of related pathways that regulate the same target	Candesartan-cilexetil (angiotensin AT1 receptor antagonist[469])	Ramipril (angiotensin converting enzyme inhibitor[470], reduced angiotensin II formation[471])	Synergistically reduced systolic blood pressure[472]	Dose-response curve shifted 6.6-fold leftwards compared to hypothetic additive curve	Candesartan-cilexetil reduced systolic blood pressure by antagonizing angiotensin AT1 receptor[469], ramipril reduced systolic blood pressure by inhibiting angiotensin converting enzyme[470], ramipril inhibited AT1 receptor agonist formation[471] thereby facilitating the action of candesartan-cilexetil by reducing AT1 agonist-antagonist competition

3.3.2.2 ***Pharmacodynamically additive and antagonistic combinations***

Investigation of additive and antagonistic combinations provides contrasting perspectives for facilitating the study of synergistic combinations. Additive combinations (Appendix Table S4, examples in Table 3-4) result from equivalent or overlapping actions (9 sets) and independent actions (4 sets) of the drugs involved. Equivalent and overlapping actions involve interactions with different targets of the same pathways that equivalently regulate the same target (7 sets) or interactions that directly or indirectly affect the same site of the same target (2 sets). For instance, retinoic acid and trichostatin A additively inhibit cell proliferation by overlapping actions of up-regulation of retinoic acid receptor beta and reactivation of its messenger RNA (mRNA) expression[473]. Independent actions involve interactions at different targets of un-related pathways (3 sets) or different sites of the same target (1 set). For instance, doxorubicin and trabectedin produce additive anticancer effect via equivalent action of DNA intercalation and covalent guanine adduct formation at specific sites in DNA minor groove[474]. Both drugs bind to DNA in non-interfering manner, and doxorubicin prefers AT regions[475], while trabectedin prefer to alkylates guanines[476]. Recent progresses in designing dual platinum-intercalator conjugates[477] suggest that it is possible for both drugs to act without hindering each other's binding mode.

Antagonistic drug combinations (Appendix Table S5, examples in Table 3-5) involve interfering actions at the same target (2 sets) or different targets of related pathways that regulate the same target (2 sets). One possible mechanism for antagonistic drug combination against the same target is mutual interference at the same site, which can be illustrated by the aminophylline-theophylline combination[478]. Both aminophylline and

theophylline are adenosine receptor antagonist and phosphodiesterase inhibitor, and involved in the release of intracellular calcium[478]. Adenosine receptor antagonist binding may be associated with non-unique binding site conformations[479]. Therefore, aminophylline or theophylline binding likely locks the receptor into a unique conformation that hinders theophylline or aminophylline binding, leading to antagonism. Likewise, inhibitor-activator, antagonist-agonist, blocker-substrate, and other mutually interfering pairs of drugs that bind to the same site may also produce antagonism. One mechanism for antagonistic drug combination against different targets of related pathways is the counteractive actions that hinder a partner drug's normal actions, which can be illustrated by the cytarabine and 17-AAG combination[437]. Cytarabine is a DNA binder[437] and 17-AAG is a heat-shock protein antagonist that abrogates Akt survival pathway[376,480]. 17-AAG antagonizes the cytotoxic activity of cytarabine, due in part to the induction of G1 arrest, which subsequently prevent the incorporation of cytarabine into cellular DNA[437].

Table 3-4 Examples of literature reported pharmacodynamically additive drug combinations, in which additive action has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed

Action type	**Combination target relationship**	**Drug A (mechanism of action related to additive effect)**	**Drug B (mechanism of action related to additive effect)**	**Reported additive effect [Ref]**	**Additism determination method**	**Possible mechanism of additive effect**
Equivalent or overlapping actions	Different targets of the same pathways that regulate the same target	Diazoxide (ATP-sensitive K+ channel opener[481], enhanced ATPase activity of channel regulatory subunits sulphonylurea receptors[482])	Dibutyryl-cGMP (activated ATP-sensitive K+ channel[481], activated channel via a cGMP-dependent protein kinase[483, 484])	Additive antinociceptive effect[485]	ANOVA synergism & dose effect data analysis	Diazoxide enhanced ATPase activity of channel regulatory subunits[482], Dibutyryl-cGMP activated channel via a cGMP-dependent protein kinase [483,484]
	Same target (different sites with direct contact with agonist site)	Propofol (interacted with GABA A receptor, acting on at TM3 segment of the beta2 subunit[486])	Sevoflurane (interacted with GABA A receptor at Ser270 of the alpha1 and alpha2 subunits[487])	Additive action in producing consciousness and movement to skin incision during general anesthesia[488]	Dixon up-down method	Propofol binds to TM3 segment of the beta2 subunit[486], Sevoflurane binds to Ser270 of the alpha1 subunit[487]. As agonist binding site is located between alpha1 and beta2 subunits[489], both drugs likely hinder agonist activity, thereby producing mutually substitutable actions.
	Same target (same site):	Ampicillin (blocked penicillin-binding protein 2A and thus bacterial cell wall synthesis)[466]	Imipenem (inhibited penicillin-binding protein -1A, -1B, -2, -4 and -5 and thus bacterial cell wall synthesis)[490]	Additive antibacterial effect[347]	Checkerboard method, fractional inhibitory concentration	Both acted at the same active site of penicillin-binding protein 2A[491] but at relatively high MICs of ≥32μg/ml [466]. The relatively high MICs make it less likely for both drugs to saturate target sites, thereby maintaining additive antibacterial effect.
Independent actions	Different targets of unrelated pathways	Artemisinin (interfered with parasite transport proteins PfATP6, disrupted parasite mitochondrial function, modulated host immune function)[492]	Curcumin (generated ROS and down-regulated PfGCN5 histone acetyltransferase activity, producing cytotoxicity for malaria parasites)[261]	Additive antimalarial activities[493]	Fractional inhibitory concentrations	Artemisinin blocked calcium transport to endoplasmic reticulum[492], Curcumin induced DNA damage and histone hypoacetylation[261]. They act at different sites in non-interfering manner.
	Same target (different sites)	Doxorubicin (DNA intercalator[475], preferred AT regions[475])	Trabectedin (formed covalent guanine adduct at specific sites in DNA minor grove[476], interacted with DNA repair system)	Additive anticancer effect[474]	Isobolographic analysis	Both bind to DNA in non-interfering manner, one preferred AT regions[475], the other alkylated guanines[476]. Recent progresses in designing dual platinum- intercalator conjugates[477] suggested that it is possible for both drugs to act without hindering each

						other's binding mode
Independent actions at dosages significantly lower than MICs, complementary actions at higher dosages	Different targets of unrelated pathways	Azithromycin (hindered bacterial protein synthesis by binding to 50S component of 70S ribosomal subunit[431])	Imipenem (inhibited penicillin-binding protein -1A, -1B, -2, -4 and -5 and thus bacterial cell wall synthesis)[490]	Additive antibacterial effect[433]	Checkerboard method, fractional inhibitory concentration	Azithromycin hindered bacterial protein synthesis[431] at MIC of 0.12μg/ml[494]. Imipenem blocked bacterial cell wall formation[490] at MICs of ≥32μg/ml[466]. At dosages significantly lower than MICs for both drugs, azithromycin's reduction of penicillin-binding proteins[490] may be insufficient for imipenem to saturate these proteins, allowing its unhindered inhibition of these proteins[490], thereby these actions proceed in non-interfering manner

Table 3-5 Examples of literature reported pharmacodynamically antagonistic drug combinations in 2000-2006, in which antagonism has been determined by established methods and its molecular mechanism has been revealed. The antagonism of the listed drug combinations is due to interfering actions of the partner drugs in each combination.

Combination target relationship	**Drug A (mechanism of action related to antagonism)**	**Drug B (mechanism of action related to antagonism)**	**Reported antagonistic effect [Ref]**	**Antagonism determination method**	**Possible mechanism of antagonism of interfering actions**
Different targets of related pathways that regulate the same target	Amphotericin B (formed ion channels in fungal membranes[495])	Ravuconazole (inhibited biosynthesis of ergosterol, a component of fungal cell membrane[496])	Antagonism in experimental invasive pulmonary aspergillosis[497,498]	Loewe additivity-based drug-interaction model	Amphotericin B can form ion channels more easily in the presence of ergosterol [495], ravuconazole inhibition of ergosterol synthesis[496] can therefore reduce the activity of amphotericin B in forming ion channels[495]
Same target	Aminophylline (adenosine receptor antagonist, phosphodiesterase inhibitor, released intracellular calcium[478])	Theophylline (released intracellular calcium, adenosine receptor antagonist, phosphodiesterase inhibitor[478])	Antagonism of inhibitory adenosine autoreceptors and release of intracellular calcium[478]	Quantal release measurement	Adenosine receptor antagonist binding may be associated with non-unique binding site conformations [479]. Aminophylline binding may lock the receptor into a unique conformation that hinder theophylline binding, thereby producing antagonistic effect

3.3.2.3 ***Pharmacodynamically potentiative combinations***

Potentiative drug combinations (Appendix Table S6, examples in Table 3-6) involve positive modulation of drug transport or permeation (7 sets), distribution or localization (8 sets), and metabolism (3 sets). Potentiative modulation of drug transport or permeation enhances drug absorption via disruption of transport barrier, delay of barrier recovery, or reduction of first-pass excretion by inhibiting drug efflux. Potentiative modulation of drug distribution or localization increases drug concentration in plasma or specific tissue by blocking drug uptake and inhibiting metabolic processes that convert drugs into excretable forms. Potentiative metabolism modulation stimulates the metabolism of drugs into active forms, or inhibits the metabolism of drugs into inactive forms.

Typical potentiative effects can be illustrated by two examples. One is the enhanced absorption of antithrombotic low molecular weight heparin (LMWH) by chitosan[499]. LMWH is an antithrombin binder that inhibits activated coagulation factors. Chitosan reversibly interacts with components of tight junctions to widen paracellular routes, which increases permeability of LMWH across mucosal epithelia and thus enhances its absorption and improved therapeutic effect. The second example is 2'-deoxyinosine enhancement of antitumor activity of 5-fluorouracil in human colorectal cell lines and colon tumor xenografts[500]. 5-fluorouracil is metabolized by thymidine phosphorylase and others into a metabolite that stabilizes P53 due to RNA-directed effects. 2'-deoxyinosine enhances thymidine phosphorylase activity and thus the metabolism of 5-fluorouracil into active metabolite, which is subsequently incorporated into cellular ribonucleic acid (RNA) in place of the normal metabolite, yielding fraudulent RNA[501].

3.3.2.4 ***Pharmacokinetically reductive combinations***

Appendix Table S7 (examples in Table 3-7) summarized 2, 1 and 4 of the 7 sets of reductive drug combinations involve negative modulation of drug transport or permeation, distribution or localization, and metabolism respectively. Reductive modulation of drug transport or permeation typically blocks drug absorption or promotion of first-pass excretion by such actions as drug-drug aggregation to reduce permeability and inhibition of drug transport into plasma or target site. Reductive modulation of drug distribution/localization decreases drug concentration in plasma or specific tissue, which typically involves stimulation of metabolic processes for converting drugs into excretable forms and inhibition of metabolic processes for increasing drug concentration. Drug activity can also be reduced by metabolism modulation to covert drugs into inactive forms.

Table 3-6 Literature reported pharmacokinetically potentiative drug combinations, in which potentiative effect has been determined by established methods and its molecular mechanism has been revealed

Biochemical class of potentiative effect	**Drug A (therapeutic or toxic effects and mechanism of actions)**	**Drug B (mechanism of action related to potentiative effect)**	**Reported potentiative effect [Ref]**	**Possible mechanism of potentiative actions**
Positive regulation of drug transport or permeation	AZT (anti-HIV, HIV-1 reverse transcriptase inhibitor)	1,8-Cineole (formed hydrogen bonds with lipid head groups of stratum corneum lipids[502])	Enhanced cross-skin permeation of AZT[503]	Enabled drug transport across skin possibly by disrupting absorption barrier via binding to lipid head groups
Enhanced drug distribution or localization	Cerivastatin (cholesterol-lowering, HMG-CoA reductase inhibitor)	Gemfibrozil (inhibited CYP2C8 mediated metabolism of statins, inhibited OATP2 mediated uptake of cerivastatin[504])	Increased plasma concentration of statins by inhibiting their metabolism and uptake[504-506]	Enhanced level of drug in plasma by metabolism and uptake inhibition
Enhanced drug metabolism	Doxorubicin (anticancer by DNA intercalation, converted to doxorubicinol by NADPH-dependent aldo/keto or carbonyl reductases [507], which produced cardiotoxicity by mediating transition from reversible to irreversible damage)	Paclitaxel (stimulated enzymatic activity of NADPH-dependent aldo/keto or carbonyl reductases[507])	Enhanced cardiotoxicity by increasing metabolism of doxorubicin into toxic metabolit[507]	Enhanced metabolism of drug into toxic metabolite

Table 3-7 Examples of literature reported pharmacokinetically reductive drug combinations, in which reductive effect has been determined by established methods and its molecular mechanism has been revealed.

Biochemical class of reductive effect	**Drug A (therapeutic or toxic effects and mechanism of actions)**	**Drug B (mechanism of action related to reductive effect)**	**Reported reductive effect [Ref]**	**Possible mechanism of reductive actions**
Drug transport and permeation	Amphotericin B (antileishmanial, formed aggregate with miltefosine[508])	Miltefosine (antileishmanial, formed aggregate with amphotericin B[508])	Reduced miltefosine-induced paracellular permeability enhancement in Caco-2 cell monolayers, inhibited uptake of both drugs, decreased transepithelial transport of both drugs[509]	Reduced drug permeability and transport
Drug distribution and localization	Cisplatin (DNA inter- and intra- strand adduct, preferably bind to the major groove of GG, AG and TACT sites[381] thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis[510])	Procainamide hydrochloride (formed cisplatin-procainamide complex[511])	Reduced cisplatin-induced hepatotoxicity via formation of less toxic platinum complex, leading to inactivation of cisplatin or its highly toxic metabolites and to a different subcellular distribution of platinum[511]	Reduced level of toxic drug by formation of less toxic complex and rearrangement of its subcellular distribution
Drug metabolism	Warfarin (anticoagulant and antithrombotic, affected coagulation proteins that act sequentially to produce thrombin, metabolized by CYP3A4[512])	Quinidine (stimulated CYP3A4 mediated metabolism of warfarin[513])	Reduced anticoagulanet effect of warfarin by stimulating its metabolism[513]	Enhanced metabolism of active drug into inactive metabolite

3.3.2.5 ***Further assessment of popularly-used combinations***

A number of drug combinations have been heavily used in clinical applications for decades 1999[514-520]. For some of these classical drug combinations, the studies of their combination effects have been primarily conducted and published before 1999, and are frequently based on non-rigorous combination analysis methods. Thus, some of these classical combinations were not selected by our search procedure. Non-the-less, their popular use is a strong indication of their possible beneficiary combination effects in comparison to those of individual drugs. Therefore, it is of interest to assess the effects and mechanisms of these classical drug combinations.

We identified 10 sets of classical drug combinations that were missed by our search procedure and contain no drug of abuse and withdrawn drug. Table 3-8 summarizes literature-described modes of actions of individual drugs, suggested combination type and possible mechanism of these combinations. The 10 combinations include 5 synergistic[514, 515,519-521], 1 dual synergistic/additive[515], 1 non-synergistic[515, 522] combinations. This non-synergistic combination has been replaced by single drug therapy in clinical setting[515]. For the remaining 3 combinations, we were unable to find a literature report indicating their possible types of combination. It is also noted that 4 of the 10 combinations have been studied by rigorous drug combination analysis methods.

Table 3-8 Assessment of clinically heavily-used drug combinations not collected by our literature search procedure. These combinations have primarily been studied by less rigorous combination analysis methods and the relevant studies have been published before 1999.

Suggested combination type	**Combination target relationship**	**Drug A (mechanism of actions related to combination effect)**	**Drug B (mechanism of actions related to combination effect)**	**Reported combination effect [Ref]**	**Combination analysis method**	**Possible mechanism of combination actions**
Potentiative combination by enhanced drug distribution or localization		Amoxicillin (inhibited bacterial cell wall synthesis[523], destructed by beta-lactamase[524])	Clavulanate (beta-lactamase inhibitor[525])	Antibacterial synergy[514]	Comparison of inhibitory activity	Clavulanate maintained level of amoxicillin at bacterial cell wall by inhibiting its degradation enzyme beta-lactamase inhibitor[525], thereby potentiatived the antibacterial activity of amoxicillin.
Synergistic combination due to facilitating actions	Different targets of related pathways that regulate the same target	Salmeterol (beta 2-adrenoceptor agonist[526] that activated T-cell subtypes[189] and promoted apoptosis via adrenoreceptor- and cAMP-independent, Ca2+-dependent mechanism[527])	Fluticasone (glucocorticoid receptor binder[528] that induced apoptosis[529], up-regulated beta2-adrenoceptor[530])	Synergistic in vitro T-cell activation and apoptosis induction in asthma[521]	Comparison of activity and protein levels	Salmeterol's agonistic activity on beta 2-adrenoceptor[526] is facilitated by fluticasone's up-regulation of beta2-adrenoceptor[530], leading to synergistic T-cell activation and apoptosis induction
Redundent combination in targeting upstream and downstream targets of the same single-route pathway	Different targets of the same pathway (upstream – downstream relationship)	Suphamethoxazole (dihydropteroate synthase inhibitor[531], metabolite covalently haptenated human serum proteins[532])	Trimethoprin (dihydrofolate reductase inhibitor[533])	No synergy detected against E. coli[522], S. somaliensis strains[515], therapeutic effect due to sulphamethoxazole alone, clinical use of combination discontinued and converted to single drug[515]	Chequerboard	Both drug target the same single-route folate metabolism pathway, with suphamethoxazole targeting the upstream dihydropteroate synthase[531] and trimethoprin targeting the dounsstream dihydrofolate reductase[533]. Redundent combination if suphamethoxazole effectively inhibits dihydropteroate synthase, trimethoprin inhibition of dihydrofolate reductase serves as a backup when suphamethoxazole becomes less effective
Unclear	Different targets of related pathways	Rifampicin (bacterial DNA-dependent RNA polymerase inhibitor[534])	Fusidic acid (interfered with bacterial protein synthesis by inhibiting the translocation of peptide elongation factor G from the ribosome[535])	Synergistic effect against S. somaliensis strains in Vitro[515]	Chequerboard	Mechanism unclear. There is a report that transcribing activity of DNA-dependent RNA polymerase from Ecoli is inhibited in vitro by addition of preparations of elongation factor Ts purified to homogeneity[536]
Synergistic combination due to	Different targets of related	Erythromycin (binds to bacterial 70S ribosomal complex to inhibit	Penicillin (binds to DD-transpeptidase that links	Combination inhibited 80% of	Chequerboard	Weakening of bacterial cell wall by penicillin[537] enhanced erythromycin

facilitating action	pathways	bacterial protein synthesis[537])	peptidoglycan, which weakens bacterial cell wall[538])	the S. somaliensis strains both synergically and additively[515]		penetration into bacterial cells, thereby enhancing its bioavailability
Potentiative combination by enhanced drug distribution or localization		Ergotamine (5-HT1B/1D receptor agonist[539], agonist of presynaptic dopamine receptors and alpha 2-adrenoceptors, postsynaptic alpha 1 and alpha 2-adrenoceptors, and antagonist of the postsynaptic alpha 1-adrenoceptors[540])	Caffeine (adenosine receptor antagonist[541] that increased dopamine and GABAergic activities[542], cAMP-phosphodiesterase inhibitor[543])	Symptomatic treatment of chronic vascular headache by the combination[516]	Comparison of activity	Caffeine increased water solubility of ergotamine to enhance its absorption[544], producing potentiative effect. Possible synergy may occur at dopamine receptor that require further investigation
Additive combination duo to equivalent action	Different targets of related pathways	Niacin (niacin receptor HM74A agonist that inhibited hepatocyte diacylglycerol acyltransferase and triglyceride synthesis leading to increased intracellular apo B degradation and subsequent decreased secretion of VLDL and LDL cholesterol[545])	Simvastatin (HMG-CoA reductase inhibitor that reduced synthesis of LDL cholesterol and triglycerides and increased HDL-cholesterol [546])	Combination reduced LDL and VLDL, and increased HDL cholesterol[517]	Comparison of activity and protein levels	Niacin reduced secretion of VLDL and LDL cholesterol[545], simvastatin reduced synthesis of LDL cholesterol and triglycerides[546], both drugs equivalently reduced the level of LDL cholesterol
Synergistic combination duo to complementary action	Same target different binding sites	Cisplatin (DNA inter- and intra-strand adduct, preferably bind to the major groove of GG, AG and TACT sites[381] thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis[510])	Cyclophosphamide (metabolite formed DNA adduct at phosphoester[547] and at G N-7 positions[548], thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis[485])	Combination produced response rates of 60% to 80% in small cell lung cancer patients[518])	Comparison of activity	Cisplatin and cyclophosphamide formed DNA adduct at different sites[510,547] possibly at mutually compatable binding conformation because of the small size of the drugs, thereby complementing each other's actions on DNA
Synergistic combination duo to facilitating action	Same target	Methotrexate (dihydrofolate reductase inhibitor[549])	Fluorouracil (anticancer, metabolite inhibited thymidylate synthase that stopped DNA synthesis[550], stabilized and activated P53 by blocking MDM2 feedback inhibition through ribosomal proteins[551])	Synergism in inhibiting viability of L1210 murine tumor cells[519]	Comparison of activity	Apart from methotrexate's anticancer dihydrofolate reductase inhibitory activity[549], methotrexate metabolite formed reversible ternary complexes with fluorouracil on one site of thymidylate synthase to enhance its binding to the enzyme[552], thereby facilitating fluorouracil's anticancer thymidylate synthase inhibitory activity

Synergistic combination duo to complementary action	Same target	Diclofenac (non-selective COX inihibitor[553], COX1 inhibition increased brain KYNA formation to produce analgesic effect[553])	Paracetamol (major ingredient of Tylenol and Panadol, metabolite agonized cannabinoid receptors to produce analgesic effect[489,554], reduced active oxidized form of COX to resting form[555], selective COX2 variant inhibitor[556])	Synergy in treatment of acute pain in humans[520]	Isobolographic analysis	Apart from its analgesic action via cannabinoid receptors[489,554], paracetamol reduced active oxidized form of COX to resting form[555] to complement diclofenac's analgesic action of COX1 inhibition[553]

Literature-described MI profiles appear to provide clues to the possible reasons for 9 of the 10 combinations. The synergistic salmeterol-fluticasone, methotrexate-fluorouracil, and erythromycin-penicillin combinations likely involve facilitating actions[530,537,552], diclofenac-paracetamol synergism may arise from complementary action[553,555]. Amoxicillin-clavulanate synergism possibly stems from potentiative enhancement of drug distribution[525]. We were unable to find information for assessing the reported synergism of rifampicin-fusidic acid combination[515]. The reported non-synergistic suphamethoxazole-trimethoprin combination appears to involve redundant actions in targeting upstream and downstream targets of a single-route pathway, with the downstream drug acting as a second line of defense[531,533]. For the 3 combinations that did not report the types of combination actions, cisplatin-cyclophosphamide combination likely produces synergistically complementary action[510,547], caffeine in the ergotamine/caffeine combination may involve the potentiation of ergotamine's action by enhancing its distribution[544], and niacin-simvastatin combination possibly produces additive effect duo to their equivalent actions[546].

3.3.2.6 ***Pathway analysis***

Pathway analysis is an effective approach for more comprehensive assessment of drug combination effects[557] as well as other drug activities and responses[558,559]. Advances in systems biology and other areas of biomedical and pharmaceutical research have enabled the integration of biomolecular network information, individual MI profiles, omics data, and disease information for drug validation and for understanding of the mechanism of drug actions[560-562]. It is therefore of interest to explore pathway analysis approaches for further studying some of the drug combinations evaluated by MI profiling.

Figure 3-1 shows the related pathways of cisplatin-trastuzumab combination (Appendix Table S1) based on reference and conventional pathways in KEGG database, and literature reports of drug pathways. Cisplatin is a DNA adduct that inhibits DNA polymerization and induces DNA damage to trigger apoptosis[510] (via P53-Bax, P53-Fas, P38-JNK, and p73 paths in Figure 3-1). Trastuzumab is an anti-HER-2/neu antibody that inhibits HER2-mediated proliferation, angiogenesis, survival and migration [563] (via PI3K-AKT-mTOR and RAS-ERK paths). Cisplatin's induction of DNA damage and apoptosis may be attenuated by DNA repair systems in certain cell types[510] (via P53-P21 paths). This counteractive DNA repair action may be reduced by trastuzumab's anti-HER2 activity that suppresses DNA repair pathway [564] and inhibits PI3K-AKT pathway[565] to enhance apoptosis[566]. The corresponding paths (dashed lines in Figure 3-1) involve the inhibition of HER2-PI3K-AKT mediated activation of P21, which reduces P21's activity in facilitating Chk1-P53-P21 and Chk1-P53-Gadd45a-P21 mediated induction of cell cycle arrest important for ATM mediated DNA repair process. Reduction of AKT activity by trastuzumab's inhibition of HER2 also lowers Mdm2's activity in facilitating P53 degradation, which enhances P21 activation to counter-balance the reduced AKT activation of P21. We were unable to identify another counter-balance path, and it is unclear to what extent the Mdm2 mediated counter-balance path affects the overall state of P21 activation.

In addition to the paths of protein-protein, protein-substrate, and protein-nucleic acid interactions, pathway analysis also needs to take into consideration of drug metabolism, transport, drug-drug interaction and complex formation. This can be illustrated by comparative analysis of the anticancer methotrexate-fluorouracil[549,552] and

antibacterial suphamethoxazole-trimethoprin[531,533] combinations (Table 3-8) that target human and bacterial folate metabolism pathways respectively but produce contrasting combination effects. The pathways affected by these two combinations are shown in Figure 3-2 and Figure 3-3 respectively. Although both combinations target upstream and downstream targets in a single route path leading to DNA synthesis (assuming that synthesis of 7,8-dihydropteroate is essential for bacterial growth), only the suphamethoxazole-trimethoprin combination shows the expected redundant effect such that effective inhibition of dihydropteroate synthase by suphamethoxazole renders trimethoprin inhibition of dihydrofolate reductase unnecessary for reducing DNA synethesis[531,533]. The un-expected methotrexate-fluorouracil synergism arises because methotrexate metabolite forms reversible ternary complexes with fluorouracil on one site of thymidylate synthase to enhance its binding to the enzyme [549,552] (dashed line in Figure 3-2), which synergistically facilitates fluorouracil's anticancer thymidylate synthase inhibitory activity.

3.4 Conclusion

Understandings of MI profiles of individual drugs, network crosstalks and regulations, and modes of actions of drug combinations are useful starting points for investigating the effects of drug combinations. For the analyzed cases of synergistic, potentiative, additive, antagonistic and reductive combinations, and likely many others, the literature-described MI profiles of the drugs involved appear to offer useful clues to the mechanism of combination actions from the perspectives of coordinated molecular interactions and network regulations. Clues to other aspects of pharmacodynamic, toxicological, and pharmacokinetic effects may also be obtained from the relevant MI profiles.

Chapter 3: Mechanisms of drug combinations: interaction and network perspectives

Discovery of efficacious drug combinations may be facilitated by targeting key efficacy and toxicity regulating nodes of positive[388,390] and negative regulations[26,28-30], anti-targets and counter-targets[33], compensatory and neutralizing actions[31,32], and transporter and enzyme mediated pharmacokinetic activities[499]. Both the discovery and the analysis of drug combinations can be facilitated by the collective use of different approaches and methods. For instance, signs of MI profiles as well as genes, pathways affected by or responsive to drug-combinations[440] and individual drugs[567-569] may be detected from gene expression or proteomics profiles by using unsupervised hierarchical clustering and supervised machine learning methods[440,567,570,571]. These combined with knowledge of the characteristics and activities of targets[337] and ADME-Tox proteins[355] enable the prediction of responses and markers[567-569], unknown therapeutic actions[570], targets and characteristics[570,572,573], efficacies[574], toxicological effects[570], and resistance profiles[571] of drug combinations and individual drugs

Figure 3-1 Pathways affected by cisplatin-trastuzumab drug combination, figure was generated by Microsoft Visio 2007

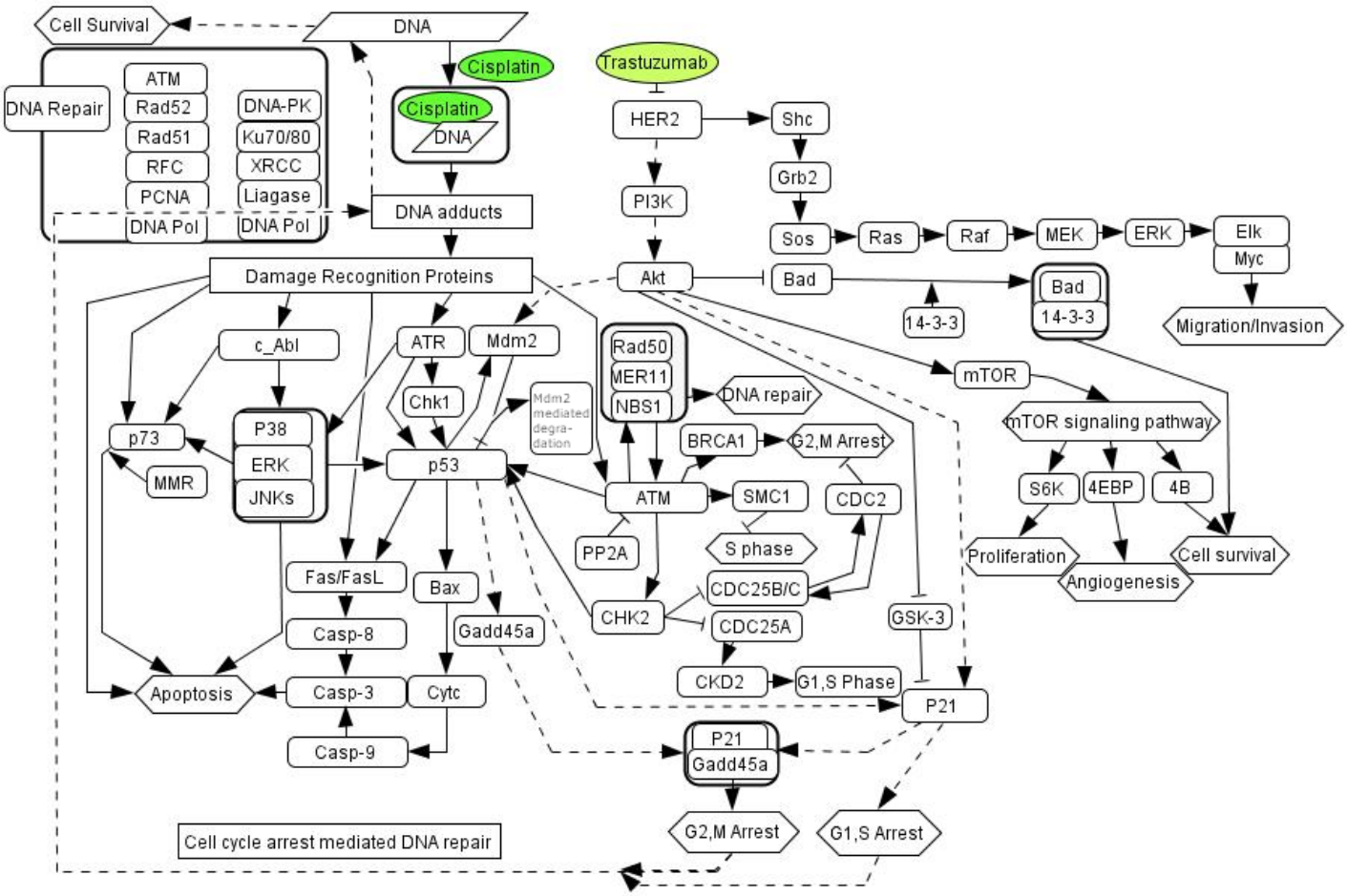


Figure 3-2 Human folate metabolism pathway affected by the combination of methotrexate (MTX) and fluorouracil (5-FU)

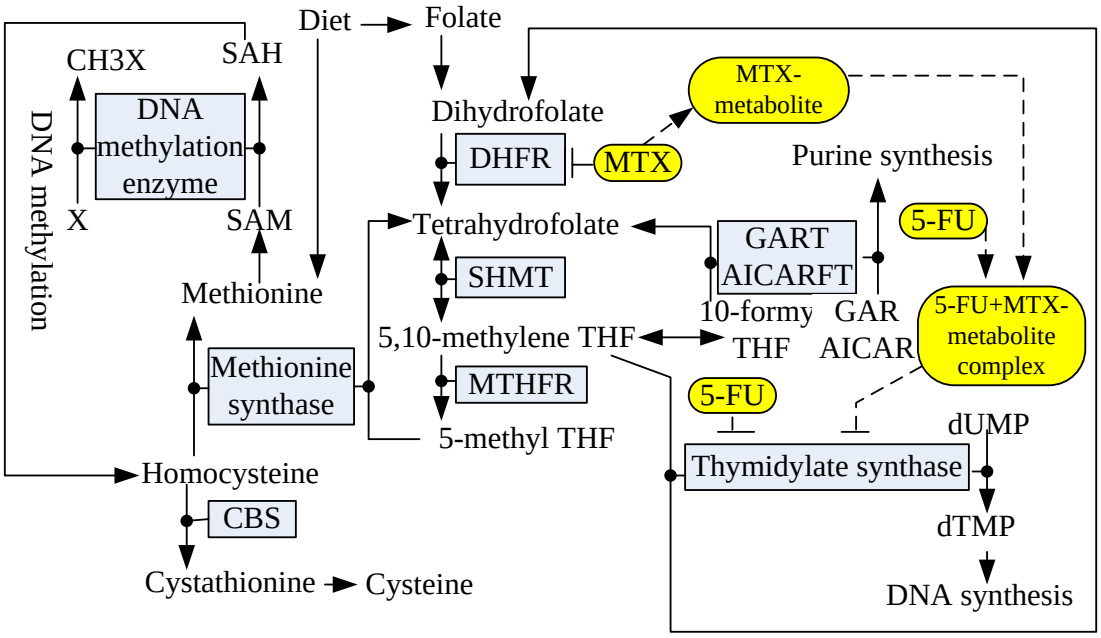
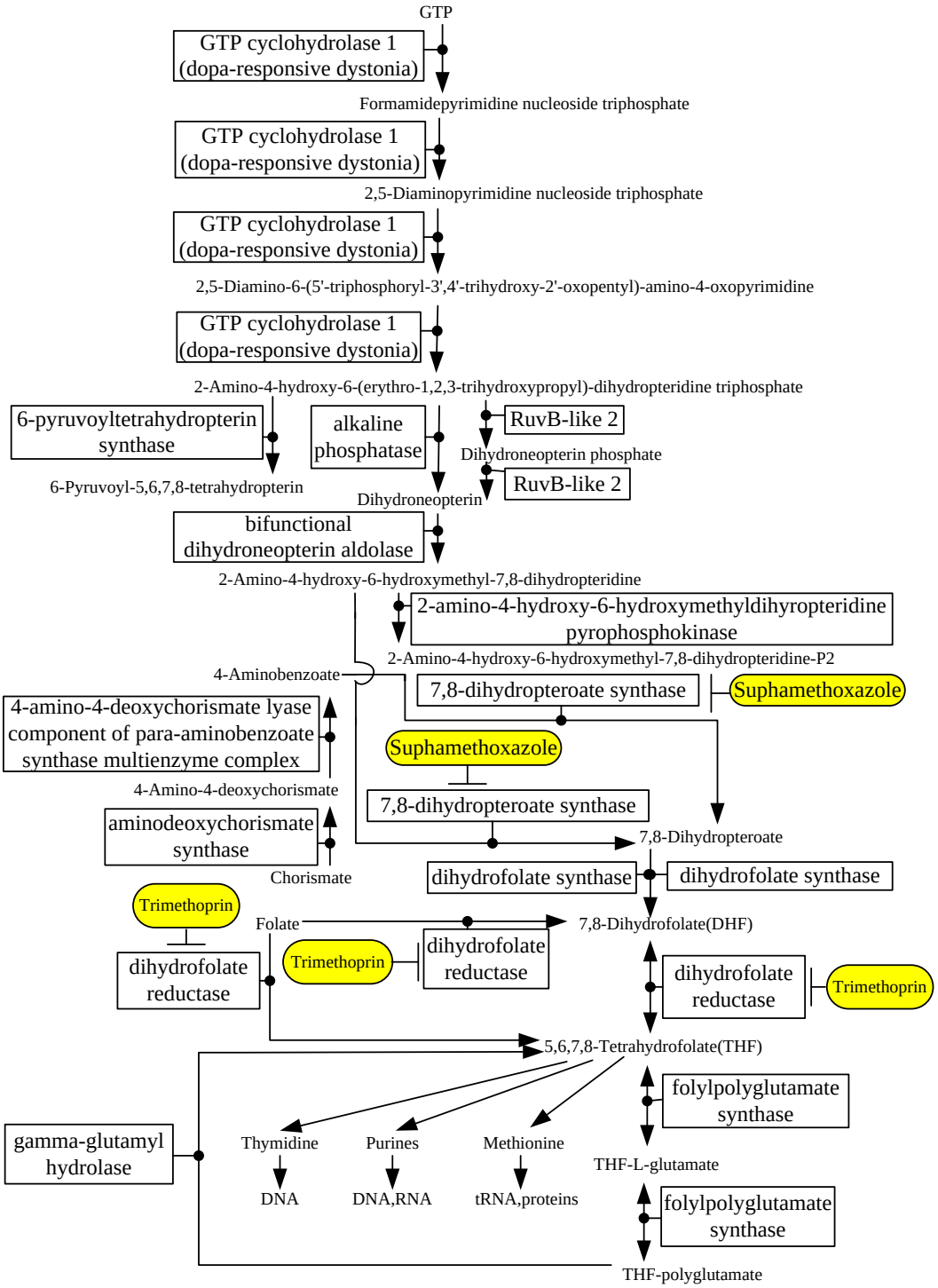


Figure 3-3 Ecoli folate metabolism pathway affected by the suphamethoxazole-trimethoprin drug combination



4 Integrated molecular profiling for predicting drug sensitivity

Receptor tyrosine kinase inhibitors elicit markedly different anticancer clinical response rates in clinical trials[575]. These rates have been linked to mutation and amplification of drug target, activating mutation of RAS, BRAF and PIK3CA, and PTEN loss-of-function[576,577]. The possible correlation with target expression is under investigation[578]. Compensatory and alternative signaling that bypass target inhibition also influence drug response[31,579]. In this work, we retrospectively analyzed mutation, amplification and microarray gene expression profiles of drug target and 8-11 known bypass and downstream drug-resistant genes in 53 NSCLC and 31 breast cancer cell-lines sensitive or resistant to gefitinib, erlotinib, lapatinib, and trastuzumab. The individual profiles correlate with sensitive and resistant cell-lines at levels comparable to the reported correlations with clinical response rates. The collective profiles showed more balanced and improved correlation with sensitive and resistant cell-lines.

4.1 Introduction

Receptor tyrosine kinase inhibitors, such as EGFR inhibitor gefitinib and erlotinib, HER2 monoclonal antibody trastuzumab, and multi-target EGFR and HER2 inhibitor lapatinib, are highly successful anticancer drugs that elicit markedly different anticancer clinical response rates in clinical trials [575,578,580,581]. For instance, the reported clinical response rates of gefitinib and erlotinib are 19.9% and 8.9% for the treatment of NSCLC

[575,578], and those of lapatinib and trastuzumab are 8~24% and 12-34% for the treatment of breast cancer [580,581]. These clinical response rates have been linked to mutation and amplification of the drug target, activating-mutation of downstream signaling genes RAS, BRAF and PIK3CA, and loss-of-function of the downstream signaling regulator PTEN [576,577]. The possible correlation of the response rates with the expression level of the drug target is also under investigation on the basis that the drug target such as EGFR is known to induce carcinogenesis via overexpression, amplification and somatic mutations [578].

Moreover, compensatory, alternative and redundant signaling that bypass target inhibition also influence drug response [31,579]. As summarized in Table 4-1, several such bypass mechanisms against the inhibition of EGFR or HER2 have been reported in the literature. While these mechanisms may appear scattered in individual patients, they collectively are expected to have a significant impact on the overall drug response rates in large patient populations. Therefore, collective consideration of the profiles of the bypass genes that regulate these bypass mechanisms together with usual consideration of the mutational, amplification and expression profiles of the drug targets and the relevant downstream genes may improve the prediction of clinical response to the relevant drugs.

To determine whether the collective profiles show a more balanced and improved correlation with sensitive and resistant cell-lines than the individual profiles, we retrospectively analyzed mutational, amplification and microarray gene expression profiles of drug target and 8-11 known bypass and downstream drug-resistant genes in 53 NSCLC and 31 breast cancer cell-lines sensitive or resistant to gefitinib, erlotinib, lapatinib, and trastuzumab. The clinical relevance of the correlation analysis against cell-line data was

evaluated by comparing the levels of correlation between the individual profiles and the sensitive and resistant cell-lines with the reported level of correlations between these profiles and clinical data to determine if they are consistent with each other.

Gefitinib and erlotinib have been approved for lung and pancreatic cancers, trastuzumab and lapatinib have been approved for breast cancer [576,582]. These drugs were evaluated because of their clinical relevance, knowledge of drug-resistance mechanisms, and availability of drug sensitivity, genetic and expression data for statistically significant number of cell-lines. The analyzed genetic data include drug sensitizing mutations and copy number variations in EGFR, activating mutations in RAS, BRAF, PIK3CA and inactivating mutations in PTEN directly contribute to the resistance against EGFR inhibitors (EGFR-I) or HER2 inhibitors (HER2-I) in significant percentage of patients (>2%) [576,577]. The expression data include microarray gene expression data of EGFR, HER2, PTEN, and several bypass genes directly contribute to the resistance against EGFR-I or HER2-I [31,579].

4.2 Materials and Methods

4.2.1 Data collection and preprocessing

*4.2.1.1 **Drug sensitivity data for NSCLC and breast cell-lines***

We identified from literatures 48, 85, and 83 NSCLC cell-lines with available sensitivity data for gefitinib, erlotinib, and lapatinib, and 24 and 22 breast cancer cell-lines with available sensitivity data for trastuzumab and lapatinib. Overall, 46 NSCLC and 17 breast cancer cell-lines with sensitivity data for one or more drugs were collected. A cell-line was considered to be sensitive to a drug if the drug inhibits it at IC50≤1μM[583], otherwise it

was considered to be resistant to the drug. Sensitivity data of NSCLC cell-lines treated with gefitinib, erlotinib, and lapatinib, and breast cancer cell-lines treated with trastuzumab and lapatinib are summarized in Table 4-2 and Table 4-3 respectively. Table 4-4 listed the literature reported performance of the prediction of EGFR-I sensitive and resistant patients by using mutation-based and amplification-based methods.

*4.2.1.2 **Molecular profiling used for NSCLC and breast cell-lines***

The genetic and microarray gene expression data for 53 NSCLC and 31 breast cancer cell-lines were collected from the published literatures, and COSMIC and GEO databases. We further identified from GEO database the microarray gene expression data for 6 lung and 9 breast cell-lines of healthy people respectively. The relevant data and literature sources for these cell-lines are summarized in Table 4-2, 4-3, and 4-5. These expression data retrieved from GEO database were processed by using R[312] and Bioconductor[311]. Raw data obtained from GEO database was filtering with imagine intensity threshold in the range of 20 to 16,000 units and followed by RMA normalization [310].

Table 4-1 The main therapeutic target, bypass genes, drug-resistant downstream signaling or regulatory genes, and the relevant bypass mechanisms in the treatment of NSCLC and breast cancer

Targeted Cancer: NSCLC **Main Target for the Treatment of Specific Cancer:** EGFR **Drugs Evaluated:** Gefitinib, Erlotinib, and Lapatinib	
Bypass Gene	**Bypass Mechanism**
HER2	EGFR inhibition upregulated HER2 and induced compensatory EGFR-HER2, HER2-HER3, HER2-HER4 heterodimerisation to promote alternative signaling[34-36]
HER3	EGFR inhibition induced compensatory transactivation of HER3 signaling[31]
IGF1R	EGFR inhibition upregulated IGF1R and induced EGFR-IGF1R heterodimerization and activation of IGFR signaling[40,42]
c-MET	EGFR inhibition countered by focal amplification of MET that transactivates HER3 signaling and interacts with EGFR to promote alternative signaling[39,41,46]
PDGFR	EGFR inhibition countered by PDGFR transactivation of HER3 signaling[41]
FGFR	Contributed to EGFR inhibitor resistance via alternative signaling[44]

VEGFR2	Contributed to EGFR inhibitor resistance via alternative signaling[38]
Drug-Resistant Downstream Gene	**Resistance Mechanism**
PTEN	PTEN loss or inactivating mutation contributed to EGFR inhibitor resistance by activation of Akt and EGFR[37,47]
KRAS	KRAS activating mutation mediated EGFR-independent signaling and contributed to EGFR inhibitor resistance[43,45]
PIK3CA	PIK3CA activating mutation mediated EGFR-independent AKT signaling and contributed to EGFR inhibitor resistance[584]
AKT	AKT activating mutation mediated EGFR-independent AKT signaling and could lead to resistance against EGFR inhibitor[585]
Targeted Cancer: Breast Cancer **Main Target for the Treatment of Specific Cancer:** HER2 **Drugs Evaluated:** Trastuzumab, Lapatinib	
Bypass Gene	**Bypass Mechanism**
EGFR	Compensatory crosstalk of HER2-EGFR heterodimerisation promoted alternative ERBB signaling[34,36]
HER3	HER2 inhibition increased HER3 localization[579] and induced compensatory transactivation of HER3 signaling[31] , HER2-HER3, HER2-HER4 heterodimerisation promotes alternative ERBB signaling[34]
IGF1R	Contributed to HER2 inhibitor resistance via IGF1R-HER2 crosstalk and alternative IGF1R signaling[579,580]
c-MET	HER2 inhibition countered by overexpressed MET that interacts with EGFR to promote alternative signaling[46]
ESR1	HER2 inhibition induced compensatory ER signaling via activation of FOXO3a and caveolin-1 mediated ESR1 transcriptional activity[586]
AXL	HER2 inhibition countered by overexpressed AXL that crosstalks with HER3, ER to promote AkT and ER signaling[587]
Drug-Resistant Downstream Gene	**Resistance Mechanism**
PTEN	PTEN loss and inactivating mutation contributed to HER2 inhibitor resistance by activation of Akt[579]
PIK3CA	PIK3CA activating mutation mediated HER2 independent AKT signaling and contributed to HER2 inhibitor resistance[588]

Table 4-2 Clinicopathological features and data source of NSCLC cell-lines used in this study. The available gene expression data, EGFR amplification status, and drug sensitivity data for gefitinib, erlotinib, and lapatinib are included together with the relevant references.

Cell-line	Histological Subtype *	Source *	Gene Expression Sample ID at GEO Database[589]	EGFR Amplification (gene copy number >3)[590]	EGFR Amplification (gene copy number >4) [590]	Mutated Gene/Genes [591,592]	Sensitivity Data		
							Gefitinib[593,594]	Erlotinib[590,594]	Lapatinib[590]
A427	NS	PT	NA			KRAS		R	R
A549	NS	PT	GSM108799			KRAS	R	R	R
Calu1	EC	PE	GSM108801			KRAS	R	R	R
Calu3	AD	PE	GSM108803				S		S
Calu6	APC	PT	GSM108805			KRAS	R	R	R
Colo699	AD	PF	NA	Y				R	R
DV90	AD	PE	NA			KRAS		R	R
EKVX	AD	PT	NA					R	R
H1155	LCC	PT	NA	NA	NA	KRAS,PTEN	R	R	
H1299	LCC	LN	GSM108807			NRAS	R	R	R
H1355	AD	PT	GSM108809			KRAS, BRAF	R	R	R
H1395	AD	PT	GSM108811			BRAF	R	R	R
H1437	AD	PT	GSM108813				R	R	R
H1563	AD	PT	NA			PIK3CA		R	R
H1568	AD	PT	NA	Y	Y			R	R
H157	SQ	PT	GSM108815			KRAS,PTEN	R	R	R
H1648	AD	LN	GSM108817				R	R	S
H1650	AD	PE	GSM108819	Y		EGFR	R	R	R
H1666	AD	PE	GSM108821			BRAF	R	R	S
H1734	AD	PT	NA	Y		KRAS		R	R
H1755	AD	Live	NA			BRAF		R	R
H1770	NE	LN	GSM108825				R	R	
H1781	AD	PE	NA			ERBB2	R	R	R
H1792	AD	PE	GSM171848	Y		KRAS		R	R
H1819	AD	LN	GSM108827	Y			R	R	S
H1838	AD	PT	NA	Y	Y			R	R
H1915	SCC	Brain	NA					R	R
H1944	AD	ST	NA			KRAS		R	R
H1975	AD	PT	GSM108829	Y		EGFR	R	R	R
H1993	AD	LN	GSM108831				R	R	R
H2009	AD	LN	GSM108833			KRAS	R	R	R
H2030	AD	LN	NA			KRAS		R	R
H2052	MT	PE	GSM171854					R	R
H2077	AD	PT	NA					R	R
H2087	AD	LN	GSM108835			BRAF, NRAS	R	R	R
H2110	NS	PE	NA					R	R
H2122	AD	PE	GSM108837			KRAS	R	R	R
H2126	LCC	PE	GSM108839				R	R	R
H2172	NS	PT	NA					R	R
H2228	AD	PT	NA					R	R
H23	AD	PT	GSM171868			KRAS, PTEN		R	R
H2347	AD	PT	GSM108841			NRAS	R	R	R
H2444	NS	PT	NA	Y		KRAS		R	R
H28	MT	PE	GSM171870					R	R
H2882	NS	PT	GSM108843				R	R	R
H2887	NS	PT	GSM108845			KRAS	R	R	R
H3122	AD	PT	GSM171874					R	R
H322	AD	PT	GSM171876	Y			R	R	R

H322M	AD	PT	NA					R	S
H3255	AD	PT	GSM108847	Y	Y	EGFR	S	S	S
H358	AD	PT	GSM108849			KRAS	R	R	R
H441	AD	PT	GSM108851			KRAS	R	R	R
H460	LCC	PE	GSM108853			KRAS, PIK3CA	R	R	R
H520	SQ	PT	NA				R	R	R
H522	AD	PT	NA	Y				R	R
H596	AD	PT	NA	Y		PIK3CA		R	R
H647	ADSQ	PE	NA			KRAS		R	R
H661	LC	LN	GSM171884					R	R
H820	AD	LN	GSM108855	Y		EGFR	R	R	R
HCC1171	NS	PT	GSM108857			KRAS	R	R	R
HCC1195	ADSQ	PT	GSM108859	Y		NRAS	R	R	R
HCC1359	SGC	PT	GSM108861				R	R	R
HCC15	SQ	PT	GSM108863			NRAS	R	R	R
HCC1833	AD	PT	GSM171898					R	R
HCC193	AD	PT	GSM108865	Y			R	R	R
HCC2279	AD	PT	GSM108867	Y	Y	EGFR	S	S	R
HCC2429	NS	PT	GSM171900					R	R
HCC2450	SQ	PT	GSM171902			PIK3CA		R	R
HCC2935	AD	PE	GSM108869			EGFR	S	S	S
HCC364	AD	PT	NA			BRAF		R	R
HCC366	ADSQ	PT	GSM108871				R	R	R
HCC4006	AD	PE	GSM108873	Y	Y	EGFR	S	S	S
HCC44	AD	PT	GSM108875			KRAS	R	R	R
HCC461	AD	PT	GSM108877			KRAS	R	R	R
HCC515	AD	PT	GSM108879			KRAS	R	R	R
HCC78	AD	PE	GSM108881				R	R	R
HCC827	AD	PT	GSM108883	Y	Y	EGFR	S	S	S
HCC95	SQ	PE	GSM108885				R	R	R
HOP62	AD	PT	NA			KRAS		R	R
HOP92	AD	PT	NA	Y				R	R
LCLC103H	LCC	PE	NA					R	R
LCLC97TM	LCC	PT	NA			KRAS		R	R
LouNH91	SQ	PT	NA	Y		EGFR		R	R
PC9	AD	PT	NA	Y		EGFR	S	S	R
SKLU1	AD	PT	NA			KRAS		R	R

* Determined from the ATCC (http://www.atcc.org) and DSMZ (http://www.dsmz.de) websites, and references therein.Abbreviations: AD, lung adenocarcinoma; APC, anaplastic carcinoma; EC, epidermoid carcinoma; LCC, large cell lung cancer; LN, lymph node; MT, mesothelioma; NA: not available; NE, neuroendocrine neoplasm; NS, not specified; NSCLC: non-small cell lung cancer; PE, pleural effusion; PF, pleural fluid; PT, primary tumor; R, resistant; S, sensitive ; SCC, small-cell carcinoma; SGC: spindle and giant cell carcinoma; ST, soft tissue; Y, gene amplified

Table 4-3 Clinicopathological features and data source of breast cancer cell-lines used in this study. The available gene expression data, HER2 amplification status, and drug sensitivity data for trastuzumab and lapatinib are included together with the relevant references.

Sample Name	Histological Subtype* (Subtype Reported in Ref [595,596])	Source*	ER/PR/ HER2 Status*	Gene Expression Sample ID at GEO Database[597]	HER2 Amplification (gene copy number >3) [595]	HER2 Amplification (gene copy number >4) [595]	Mutated Gene/Genes[591,592]	Sensitivity Data	
								Trastuzumab [598-601]	Lapatinib[602-604]
BT20	AC (A)	PT	-/-/-	GSM320590			PIK3CA	R	R
BT474	IDC (L)	PT	+/+/+	GSM320596	Y	Y		S	S
BT549	IDC (B)	PT	-/-/-	GSM320598			PTEN	R	R
CAL51	AC (B)	PE	-/NA/-	GSM320616			PIK3CA		R
CAMA1	AC (L#)	PE	+/-/- #	GSM320599	Y	Y	PTEN	R	R
EFM19	IDC (L)	PE	+/+/-	GSM320618			PIK3CA		R
EFM192A	AC (L)	PE	+/+/+	GSM320619	Y	Y	PIK3CA	R	R
HCC1143	DC (A)	PT	-/-/-	GSM320631				R	
HCC1395	DC (B)	PT	-/-/-	GSM320630				R	
HCC1419	DC (L)	PT	-/-/+	GSM320624	Y	Y		R	
HCC1954	DC (A)	PT	-/-/+	GSM320627	Y	Y	PIK3CA	R	
HCC70	DC (A)	PT	-/-/-	GSM320625				R	
HS578T	C Sar (B)	PT	-/-/-	GSM320601			HRAS		R
JIMT-1	IDC (B#)	PE	NA	GSM320639	Y	Y	PIK3CA	R	R
KPL1	IDC (L#)	PE	NA	GSM320622					R
MCF7	Met AC (L)	PE	+/+/-	GSM320602			PIK3CA	R	R
MDA-MB-157	Med C (B)	PE	-/-/-	GSM421871					R
MDA-MB-175VII	IDC (L)	PE	+/-/-	GSM320603					S
MDA-MB-231	Met AC (B)	PE	-/-/-	GSM320604			BRAF,KRAS	R	R
MDA-MB-361	Met AC (L)	BR	+/+/+	GSM320605	Y	Y	PIK3CA	R	R
MDA-MB-415	AC (L#)	PE	+/-/- #	GSM320606	Y	Y	PTEN	R	
MDA-MB-435s	IDC (B#)	PE	-/-/- #	GSM320607	Y	Y	BRAF	R	R
MDA-MB-436	AC (B)	PE	-/-/-	GSM320608				R	
MDA-MB-453	Met C (L)	PE	-/-/+	GSM320609	Y	Y	PIK3CA,PTEN	S	
MDA-MB-468	Met AC (A)	PE	-/-/-	GSM320610			PTEN	R	R
SK-BR-3	AC (L)	PE	-/-/-	GSM320611	Y	Y		S	S
T47D	IDC (L)	PE	+/+/-	GSM320612			PIK3CA	R	R
UACC812	IDC (L)	PT	+/-/+	GSM320613	Y	Y		S	S
UACC893	IDC (L)	PT	-/-/+	GSM320638	Y	Y	PIK3CA	R	S
ZR-75-1	IDC (L)	AF	+/-/-	GSM320614					R
ZR-75-30	IDC (L)	AF	+/-/+	GSM320633	Y	Y		S	

* Determined from the ATCC (http://www.atcc.org) and DSMZ (http://www.dsmz.de) websites, and references therein, or from this study. # information only obtained from Ref [596]

Abbreviations: A, basal A subtype; AC, adenocarcinoma; AF, ascites fluid; B, basal B subtype; C Sar, carcinoma sarcoma; CWN, chest wall nodule; DC, ductal carcinoma; IDC, invasive ductal carcinoma; Inf, inflammatory carcinoma; ILC, invasive lobular carcinoma; L, luminal subtype; Med C, medullary carcinoma, Met AC, metastatic adenocarcinoma; Met C, metastatic carcinoma; NA, not available; PE, pleural effusion; PT, primary tumor; R, resistant; S, sensitive ; Y, gene amplified

Table 4-4 The literature reported performance of the prediction of EGFR-I sensitive and resistant patients by using mutation-based and amplification-based methods

Study	Study Details	No of EGFR-I Resistant Patients	No of EGFR-I Resistant Patients	No of EGFR-I Resistant Patients	Method for Predicting EGFR-I Sensitivity	Percent of EGFR-I sensitive patients predicted by method	Percent of EGFR-I resistant patients predicted by method
Chan *et al.* 2006	15 studies 3,016 patients 509 with EGFR mutations	280	90	190	EGFR sensitizing mutation	77%	76%
Wu *et al.* 2007	6 studies 506 patients 152 with EGFR mutations	57	22	35	EGFR sensitizing mutation	77%	54%
Murray *et al.* 2008	202 studies 12,244 patients 3,188 with EGFR mutations	1578	498	1080	EGFR sensitizing mutation	80%	75%
Uramoto *et al.* 2007	27 studies 1170 patients 437 with EGFR mutations	1170	384	786	EGFR sensitizing mutation	82%	84%
Linardou *et al.* 2008	17 studies 1008 patients 165 with KRAS mutation	1008	105	903	KRAS activating mutation	95%	23%
Linardou *et al.* 2008	7 studies 756 patients 127 with KRAS mutation	756	81	675	KRAS activating mutation	98%	19%
Uramoto *et al.* 2007	7 studies 663 patients 211 with EGFR amplification	663	108	555	EGFR amplification	61%	74%

Table 4-5 Normal cell-lines (6 from the lung bronchial epithelial tissues and 9 from breast epithelial tissues) obtained from GEO database

Gene Expression Sample ID of Normal Cell-line at GEO Database	Cell-lines	Source of Cell-lines	Reference
GSM427196	NHBE	Normal human bronchial epithelial cells	Ref [605]
GSM427197	NHBE	Normal human bronchial epithelial cells	
GSM427198	BEAS-2B	Immortalized bronchial epithelial cells	
GSM427199	BEAS-2B	Immortalized bronchial epithelial cells	
GSM427200	1799	Immortalized lung epithelial cells	
GSM427201	1799	Immortalized lung epithelial cells	
GSM158659	BPE	Normal human mammary epithelial cells	Ref [606]
GSM158660	BPE	Normal human mammary epithelial cells	
GSM158661	BPE	Normal human mammary epithelial cells	
GSM158662	BPE	Normal human mammary epithelial cells	
GSM158663	BPE	Normal human mammary epithelial cells	
GSM158670	HME	Normal human mammary epithelial cells	
GSM158672	HME	Normal human mammary epithelial cells	
GSM158673	HME	Normal human mammary epithelial cells	
GSM158674	HME	Normal human mammary epithelial cells	

4.2.2 Bypass mechanism of studied tyrosin kinase inhibitors

The clinical efficacy of gefitinib, erlotinib, and lapatinib against NSCLC, and that of trastuzumab and lapatinib against breast cancer are mostly due to their inhibition of the main targets, EGFR and HER2, respectively [576,582]. Resistance to EGFR-I and HER2-I primarily arises from resistant mutations and amplification of the main target, activating mutations of down-stream signaling genes and loss of function of down-stream regulatory genes [31,576,577], and compensatory, alternative and redundant signaling genes frequently up-regulated or amplified in resistant patients [31,39,579]. Efflux-pumps, primarily responsible for the resistance of chemotherapy drugs [607], are not expected to significantly contribute to the resistance of the evaluated drugs because these drugs are either efflux-pump inhibitors [608-610], or monoclonal antibody un-affected by efflux-pumps [611]. Table 3-1 summarizes the literature-reported 11 and 8 bypass-genes and downstream signaling and regulatory genes that directly contribute to EGFR-I and HER2-I resistance respectively, the corresponding bypass and resistance mechanisms, and relevant literatures.

4.2.3 Drug sensitivity evaluation procedure

We retrospectively evaluated the capability of the individual and combinations of the genetic and expression profiles of the main target, downstream signaling and regulatory genes, and bypass genes in Table 3-1 for predicting the sensitivity of the 53 NSCLC cell-lines to gefitinib (6 sensitive, 38 resistant), erlotinib (7 sensitive, 46 resistant), and lapatinib (8 sensitive, 40 resistant), and that of the 31 breast cancer cell-lines to trastuzumab (5 sensitive, 19 resistant) and lapatinib (5 sensitive, 17 resistant). We

evaluated 11 mutation-based, amplification-based, expression-based, and combination methods. Due to inadequate copy number data, the amplification-based methods exclude the profiles of the bypass and downstream genes, which directly contribute to EGFR-I and HER2-I resistance [39], Non-the-less, copy number variation significantly influence gene expression, with 62% of amplified genes showing moderately or highly elevated expression [612]. Thus the effects of amplification of bypass genes are expected to be partially reflected by the expression profiles.

In mutation-based method M1, a NSCLC cell-line is predicted as sensitive to a drug if EGFR contains a mutation sensitizing to the drug [578] and the drug inhibits EGFR at IC50≤500nM [583] (a stricter condition of IC50≤100nM gives the same results in all studied cases), otherwise it is predicted as drug-resistant. In mutation-based method M2, a NSCLC cell-line is predicted as sensitive to a drug if the drug inhibits EGFR at IC50≤500nM [583], and the un-inhibited KRAS has no activating mutation [45]. In mutation-based method M3, a NSCLC or breast cancer cell-line is predicted as sensitive to a drug if:

(1) the drug inhibits EGFR or HER2 at IC50≤500nM [583] and EGFR in NSCLC cell-line has at least one sensitizing mutation [578];

(2) the un-inhibited KRAS, NRAS, BRAF, PIK3CA in NSCLC cell-line [576] or PIK3CA in breast cell-line [578] has no activating mutation;

(3) there is no PTEN loss or PTEN inactivating mutation in NSCLC [576] and breast [578] cell-line.

The mutation profiles of the relevant genes in each cell-line were generated by comparative sequence analysis with respect to the reported sensitizing, activating or inactivating mutations, which are summarized in Appendix Table S8 and Table S9. PTEN loss was assumed to occur if its microarray gene expression level is ≤1/5 of the median level of PTEN in the normal tissue cell-lines [613], based on the comparison of the western-blot staining of a PTEN-deficient cell-line ZR-75-1[614] with that of a PTEN-normal cancer cell-line MCF-7 [614] (variation of this cut-off from 0 to 1/3 of the median level gives the same results in all studied cases).

In amplification-based method A1 and A2, a NSCLC or breast cell-line is predicted as sensitive to a drug if EGFR or HER2 in the respective cell-line is amplified and inhibited by the drug at IC50≤500nM [583]. A gene in a cell-line is considered amplified if its copy number is ≥4 [615] in method A1 and ≥3 [616] in method A2 respectively. Copy numbers of the evaluated genes in the studied cell-lines were from literatures [616,617]. In expression-based method E1, a NSCLC or breast cell-line is predicted as sensitive to a drug if EGFR or HER2 in the respective cell-line is over-expressed [578] and inhibited by the drug at IC50≤500nM [583]. The expression-based method E2 differs from method E1 by an additional condition: all un-inhibited bypass genes in a cell-line are not over-expressed. Bypass genes are frequently up-regulated or amplified in resistant patients [31,39,579], which likely enable the promotion of drug-resistant signaling at significant levels. A gene in cancer cell-lines was assumed as over-expressed if its microarray gene expression level is ≥2-fold higher than the lowest level of the same gene in the corresponding healthy tissue cell-lines [618].

4.3 Results and discussion

4.3.1 Assesment of EGFR-I sensitivity prediction by mutation and amplification profiles

The performance and clinical relevance of the methods M1, M2, A1 and A2 in predicting gefitinib and erlotinib sensitive and resistant NSCLC cell-lines were evaluated. The results are summarized in Table 4-6 together with the reported performance of the mutation-based [45,578,619] and amplification-based [619] methods in predicting EGFR-I sensitive and resistant patients, and are detailed in Appendix Table S10 and Table S11 together with the respective cell-line sensitivity data. The EGFR-I sensitizing mutations in these Tables are applicable to all 3 drugs against NSCLC. The methods M1 and M2 correctly predicted 67% and 100% EGFR-I sensitive and 100% and 47% of EGFR-I resistant cell-lines respectively, which are comparable to the reported 77%~82% and 95%~98% accuracies in predicting EGFR-I sensitive, and 54%~84% and 19%~23% accuracies in predicting EGFR-I resistant patients by EGFR-I sensitizing mutation [578] and KRAS activating mutation [45] methods respectively. The method A1 and A2 correctly predicted 67% and 67% EGFR-I sensitive and 100% and 83% of EGFR-I resistant cell-lines respectively, which are comparable to the reported 61% and 74% accuracy in predicting EGFR-I sensitive and resistant patients by the EGFR amplification method [619]. Thus, the evaluated methods are capable of predicting EGFR-I sensitivity from NSCLC cell-lines at performance levels that reflect the sensitivity of real patients.

Table 4-6 Comparison of the literature reported performance of the prediction of EGFR-I sensitive and resistant patients with that of EGFR-I sensitive and resistant NSCLC cell-lines by using mutation-based and amplification-based methods

Study	**Study Details**	**No of EGFR-I Treated Patients**	**No of EGFR-I Sensitive Patients**	**No of EGFR-I Resistant Patients**	**Method for Predicting EGFR-I Sensitivity**	**Percent of EGFR-I sensitive patients predicted by method**	**Percent of EGFR-I resistant patients predicted by method**
Chan *et al.* 2006	15 studies 3,016 patients 509 with EGFR mutations	280	90	190	EGFR sensitizing mutation	77%	76%
Wu *et al.* 2007	6 studies 506 patients 152 with EGFR mutations	57	22	35	EGFR sensitizing mutation	77%	54%
Murray *et al.* 2008	202 studies 12,244 patients 3,188 with EGFR mutations	1578	498	1080	EGFR sensitizing mutation	80%	75%
Uramoto *et al.* 2007	27 studies 1170 patients 437 with EGFR mutations	1170	384	786	EGFR sensitizing mutation	82%	84%
Linardou *et al.* 2008	17 studies 1008 patients 165 with KRAS mutation	1008	105	903	KRAS activating mutation	95%	23%
Linardou *et al.* 2008	7 studies 756 patients 127 with KRAS mutation	756	81	675	KRAS activating mutation	98%	19%
Uramoto *et al.* 2007	7 studies 663 patients 211 with EGFR amplification	663	108	555	EGFR amplification	61%	74%
Study	**Study Details**	**No of EGFR-I Treated Cell-Lines**	**No of EGFR-I Sensitive Cell-lines**	**No of EGFR-I Resistant Cell-lines**	**Method for Predicting EGFR-I Sensitivity**	**Percent of EGFR-I sensitive cell-lines predicted by method**	**Percent of EGFR-I resistant patients predicted by method**
This work	53 NSCLC cell-lines treated with EGFR-I gefitinib or erlotinib	53	6	47	EGFR sensitizing mutation (M1)	67%	94%
					KRAS activating mutation (M2)	100%	47%
					EGFR amplification (A1)	67%	100%
					EGFR amplification (A2)	67%	83%

4.3.2 Assessment of integrated molecular profiling for predicting TKIs sensitivity

Both the reported studies and our analyses in Table 4-6 indicated that the individual-profile tends to show good performance for sensitive cell-lines at the expense of resistant cell-lines or vice versa. Combined mutation and amplification profiles have shown good correlation with clinical response [620]. It is of interest to evaluate whether more balanced performance can be achieved by using combination-profile methods. We therefore evaluated 5 two-profile methods in both NSCLC and breast cell-lines: M3+A1 (C1), M3+E1 (C2), M3+E2 (C3), A1+E1 (C4), and A1+E2 (C5), and 2 three-profile methods: M3+A1+E1 (C6) and M3+A1+E2 (C7). Overlapping of drug resistant mutation was not identify between drug target and downstream genes in EGFR-I sensitive cell-lines, resulting the similar prediction performance of M1 and M3 in these cell-lines. We therefore used M3 in combined profiles. M2 is KARS drug resistant mutation-based method which was used to compare the reported diagnositic accuracy in clinical trails. Forthermore, gene amplification cutoff in A1 showed better performance than that of A2 in drug resistant cell-lines. Only A1 was included in combined profiles. The results of 5 two-profile and 2 three-profile methods are summarized in Table 4-7, and detailed in Table 4-8 and Table 4-9 which also include the cell-line sensitivity data and the genetic and expression profiles of the main target, bypass genes and downstream signaling and regulatory genes.

Overall, the combination-profile methods showed more balanced and improved predictive performance over the individual-profile methods. Consideration of the contribution of bypass genes substantially improved the predictive performance for resistant cell-lines, except one drug trastuzumab at the expense of reduced performance for sensitive cell-

lines. C4 method showed slightly lower performance than the other combination methods probably due to its limited focus on the amplification and expression profiles of the main target only. The three-profile methods did not show improved performance over the two-profile methods possibly due to two reasons. One is the substantial level of redundancy among drug sensitizing mutation, amplification and expression profiles and among drug resistant activating/inactivating mutation and expression profiles. Another is the high noise levels of microarray gene expression data [140] that negatively affect the performances of the combination methods with expression profiles.

4.3.3 The distribution and coexistence of drug sensitive and resistant profiles

Table 4-10 shows the distribution and coexistence of drug sensitizing mutation, amplification and expression profiles, and drug resistant mutation and expression profiles in the evaluated NSCLC and breast cancer cell-lines. In NSCLC cell-lines, EGFR-I sensitizing mutations are mostly distributed in EGFR-I sensitive cell-lines and substantially coexist with EGFR amplification and over-expression, and the resistance profiles are dominated by RAS activating mutation and HER3 over-expression, which are consistent with literature reports [31,578]. In breast cell-lines, HER2 amplification occurs in majority of the cell-lines and primarily coexists with HER2 over-expression, and with PIK3CA activating mutations and HER3 and IGF1R over-expression that dominate the resistance profiles, which is consistent with recent findings [31,579,621]. Up to 3 of the 7 NSCLC cell-lines with EGFR-I sensitizing mutations and up to 6 of the 11 HER2 amplified breast cancer cell-lines are EGFR-I and HER2-I resistant respectively, primarily due to EGFR resistance mutation and over-expression of bypass genes and downstream genes, suggesting the importance of these genes in drug sensitivity analysis. Drugs

targeted on the EGFR resistance mutation and over-expressed bypass genes can compensate the resistant profiles observed in NSCLC and breast cancer cells. Two NSCLC cell-lines (HCC366, HCC1650) with EGFR sensitizing mutation and one NSCLC cell-line with EGFR amplification (HCC2279) are with null resistance profile but are resistant to gefitinib/erloitinib and lapatinib respectively, probably due to other bypass mechanism not yet included in Table 4-1.

Table 4-7 Percentages of gefitinib, erlotinib, or lapatinib sensitive and resistant NSCLC cell-lines, and percentages of trastuzumab or lapatinib sensitive and resistant breast cancer cell-lines correctly predicted by mutation-based method M1, M2, and M3, amplification-based method A1 and A2, expression-based method E1 and E2, and combination methods C1, C2, C3, C4, C5, C6, and C7. Ps and Pr is percentage for sensitive and resistant cell-lines respectively

Cancer	Main Target of Cancer	Drug (Efficacy Targets)	Number of Cell-lines (Sensitive/ Resistant)	Percentage of Correctly Predicted Sensitive Cell-Lines (P_S) and Resistant Cell-Lines (P_R) by Different Methods														
					Mutation-Based Method			Amplification-Based Method		Expression-Based Method		Combination of Two Methods					Combination of Three Methods	
					M1	M2	M3	A1	A2	E1	E2	C1= M3+A1	C2= M3+E1	C3= M3+E2	C4= A1+ E1	C5= A1+E 2	C6= M3+A1 +E1	C7= M3+A1 +E2
NSCLC	EGFR	Gefitinib (EGFR)	44(6/38)	P_S	67%	100%	67%	67%	67%	50%	50%	83%	50%	50%	83%	83%	83%	83%
				P_R	92%	39%	97%	100%	82%	95%	97%	97%	92%	97%	95%	97%	92%	95%
		Erlotinib (EGFR)	51(5/46)	P_S	80%	100%	80%	80%	80%	60%	60%	100%	100%	100%	100%	100%	100%	100%
				P_R	93%	37%	98%	100%	87%	96%	98%	98%	96%	98%	98%	100%	96%	98%
		Lapatinib (HER2, EGFR)	48(8/40)	P_S	38%	100%	38%	38%	50%	38%	38%	50%	50%	50%	50%	50%	50%	50%
				P_R	90%	43%	95%	98%	83%	95%	98%	95%	90%	93%	93%	95%	90%	93%
Breast Cancer	HER2	Trastuzumab (HER2)	24(5/19)	P_S	-	-	71%	71%	-	100%	71%	57%	71%	57%	100%	71%	71%	57%
				P_R	-	-	58%	74%	-	74%	84%	95%	95%	100%	74%	84%	95%	100%
		Lapatinib (HER2, EGFR)	22(5/17)	P_S	-	-	80%	80%	-	100%	100%	80%	80%	80%	100%	100%	80%	80%
				P_R	-	-	67%	78%	-	72%	89%	100%	100%	100%	72%	89%	100%	100%

Table 4-8 The genetic and expression profiles of the main target, downstream genes and regulator, and bypass genes of 53 NSCLC cell-lines, and the predicted and actual sensitivity of these cell-lines against 3 kinase inhibitors: gefitinib (D1), erlotinib (D2), and lapatinib (D3).

NSCLC Cell lines	**Profile of Main Target (EGFR) Related to Drug Sensitivity**				**Profile of Main Target (EGFR) Related to Drug Resistance**	**Profile of Downstream Signaling Gene or Regulator Directly Contributing to Drug Resistance**				**Profile of Bypass Gene Directly Contributing to Drug Resistance**							**Predicted (Pre) and Actual (Act) Sensitivity to Gefitinib (D1) and Erlotinib (D2)**			**Predicted (Pre) and Actual (Act) Sensitivity to Lapatinib (D3)**	
	over exp	amp (copy no>4)	amp (copy no>3)	s-mut	r-mut	RAS a-mut	BRAF a-mut	PIK3CA a-mut	PTEN loss	HER2 over exp (Not applicable to D3)	HER3 over exp	FGFR1 over exp	IGF1R over exp	VEGFR2 over exp	c-MET over exp	PDGFR over exp	Pre by M1, M2, M3, A1, E1, E2, C1, C2, C3, C4, C5, C6, C7	Act (D1)	Act (D2)	Pre by M1, M2, M3, A1, E1, E2, C1, C2, C3, C4, C5, C6, C7	Act (D3)
Calu3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	S	NA	R,S,R,R,R,R,R,R,R,R,R,R,R	S
H3255	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	S,S,S,S,S,S,S,S,S,S,S,S,S	S	S	S,S,S,S,S,S,S,S,S,S,S,S,S	S
HCC2279	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	S,S,S,S,R,R,S,S,S,S,S,S,S	S	S	S,S,S,S,R,R,S,S,S,S,S,S,S	R
HCC2935	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	S,S,S,R,S,S,S,S,S,S,S,S,S	S	S	S,S,S,R,S,S,S,S,S,S,S,S,S	S
HCC4006	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	S,S,S,S,R,R,S,S,S,S,S,S,S	S	S	S,S,S,S,R,R,S,S,S,S,S,S,S	S
HCC827	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	S,S,S,S,S,S,S,S,S,S,S,S,S	S	S	S,S,S,S,S,S,S,S,S,S,S,S,S	S
A549	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
Calu1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
Calu6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H1299	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H1355	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H1395	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H1437	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H157	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R

H1648	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	S
H1650	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	S,S,S,R,R,R,S,S,R,R,R,S,S	R	R	S,S,S,R,R,R,S,S,R,R,R,S,S	R
H1666	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	S
H1770	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	
H1792	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	R,R,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H1819	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	S
H1975	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	S,S,R,R,R,R,R,R,R,R,R,R,R	R	R	S,S,R,R,R,R,R,R,R,R,R,R,R	R
H1993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H2009	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H2052	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	R,S,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H2087	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H2122	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H2126	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H23	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H2347	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H2882	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H2887	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H3122	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H322	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R	R
H358	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R
H441	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R	R

H460	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R
H661	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	R,S,R,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R
H820	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	S,S,R,R,R,R,R,R,R,R,R,R,R,R	R	R	S,S,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC1171	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC1195	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC1359	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC15	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC1833	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC193	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	R,S,R,R,S,R,R,S,R,S,R,S,R	R	R	R,S,R,R,S,R,R,S,R,S,R,S,R	R
HCC2429	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC2450	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R,R	NA	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC366	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R,S,R,R,S,S,R,R,S,S,S,S,S	R	R	R,S,R,R,S,S,R,R,S,S,S,S,S	R
HCC44	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,S,R,R,R,R,R	R	R	R,R,R,R,R,R,R,S,R,R,R,R,R	R
HCC461	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC515	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,R,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC78	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R
HCC95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R	R	R,S,R,R,R,R,R,R,R,R,R,R,R,R	R

Notes: "1" and "0" indicates the corresponding profile is positive (over-expressed, amplified or mutated) and negative (not over-expressed, amplified or mutated) respectively. "S", "R", "NA", "s-mut", "r-mut", 'a-mut', "amp", "over exp", "pre", and "act" stands for sensitive to drug, resistant to drug, no available drug sensitivity, drug sensitive mutation, drug resistance mutation, activating mutation, amplification, over expression, predicted drug sensitivity, and actual drug sensitivity respectively. The prediction methods M1, M2, M3, A1, E1, E2, C1, C2, C3, C4, C5, C6, and C7 are described in the text.

Table 4-9 The genetic and expression profiles of the main target, downstream signaling genes and regulator, and bypass genes of 31 breast cancer cell-lines, and the predicted and actual sensitivity of these cell-lines aganist 2 kinase inhibitors trastuzumab (D4) and lapatinib (D3).

Breast Cancer Cell line	**Profile of Main Target Related to Drug Sensitivity**			**Profile of Downstream Signaling Gene or Regulator Directly Contributing to Drug Resistance**			**Profile of Bypass Gene Directly Contributing to Drug Resistance**						**Predicted (Pre) and Actual (Act) Sensitivity to Trastuzumab (D4)**		**Predicted (Pre) and Actual (Act) Sensitivity to Lapatinib (D3)**	
	HER2 over exp	HER2 amp (copy no≥4)	HER2 amp (copy no≥3)	PIK3CA mut	PTEN mut	PTEN loss	EGFR over exp (Not applicable to D3)	HER3 over exp	ESR1 over exp	IGF1R over exp	AXL over exp	c-MET over exp	Pre by M3, A1, E1, E2, C1, C2, C3, C4, C5, C6, C7	Act (D4)	Pre by M3, A1, E1, E2, C1, C2, C3, C4, C5, C6, C7	Act (D3)
BT474	1	1	1	0	0	0	0	0	0	0	0	0	S,S,S,S,S,S,S,S,S,S,S	S	S,S,S,S,S,S,S,S,S,S,S	S
HCC1419	1	1	1	0	0	0	0	1	0	0	0	0	S,S,S,R,S,S,R,S,R,S,R	S	S,S,S,R,S,S,R,S,R,S,R	NA
MDA-MB-453	1	1	1	1	0	0	0	0	0	0	0	0	R,S,S,S,R,R,R,S,S,R,R	S	R,S,S,S,R,R,R,S,S,R,R	NA
MDA-MB-175VII	1	0	0	0	0	0	0	0	0	0	0	0	S,R,S,S,R,S,S,S,S,S,S	NA	S,R,S,S,R,S,S,S,S,S,S	S
SK-BR-3	1	1	1	0	0	0	0	0	0	0	0	0	S,S,S,S,S,S,S,S,S,S,S	S	S,S,S,S,S,S,S,S,S,S,S	S
ZR-75-30	1	1	1	0	0	0	0	1	0	0	0	0	S,S,S,R,S,S,R,S,R,S,R	S	S,S,S,R,S,S,R,S,R,S,R	NA
UACC812	1	1	1	0	0	0	0	0	0	0	0	0	S,S,S,S,S,S,S,S,S,S,S	S	S,S,S,S,S,S,S,S,S,S,S	S
UACC893	1	1	1	1	0	0	0	0	0	0	0	0	R,S,S,S,S,R,R,S,S,R,R	R	R,S,S,S,S,R,R,S,S,R,R	S
BT20	0	0	0	1	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R,R	R
BT549	0	0	0	0	1	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R,R	R
CAL51	0	0	0	1	0	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R	NA	R,R,R,R,R,R,R,R,R,R,R	R
CAMA1	0	0	0	0	1	0	0	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R,R	R
EFM19	1	0	0	1	0	0	0	1	0	1	0	0	R,R,S,R,R,R,R,S,R,R,R	NA	R,R,S,R,R,R,R,S,R,R,R	R
EFM192A	1	1	1	1	0	0	0	1	0	0	0	0	R,S,S,R,R,R,R,S,R,R,R	R	R,S,S,R,R,R,R,S,R,R,R	R
HCC1143	0	0	0	0	0	0	0	0	0	0	1	0	S,R,R,R,R,R,R,R,R,R,R	R	S,R,R,R,R,R,R,R,R,R,R	NA
HCC1395	0	0	0	0	0	0	0	0	0	0	0	0	S,R,R,R,R,R,R,R,R,R,R	R	S,R,R,R,R,R,R,R,R,R,R	NA
HCC1954	1	1	1	1	0	0	0	0	0	0	0	0	R,S,S,S,R,R,R,S,R,R,R	R	R,S,S,S,R,R,R,S,R,R,R	NA

HCC70	0	0	0	0	0	0	0	0	0	0	0	0	S,R,R,R,R,R,R,R,R,R	R	S,R,R,R,R,R,R,R,R,R	NA
JIMT-1	1	1	1	1	0	0	0	0	0	0	0	0	R,S,S,S,R,R,R,S,S,R,R	R	R,S,S,S,R,R,R,S,S,R,R	R
Hs578T	0	0	0	0	0	0	0	0	0	0	1	0	S,R,R,R,R,R,R,R,R,R	NA	S,R,R,R,R,R,R,R,R,R	R
KPL1	0	0	0	0	0	0	0	1	0	0	0	0	S,R,R,R,R,R,R,R,R,R	NA	S,R,R,R,R,R,R,R,R,R	R
MCF7	0	0	0	1	0	0	0	1	0	1	0	0	R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R	R
MDA-MB-157	0	0	0	0	0	0	0	0	0	1	0	0	S,R,R,R,R,R,R,R,R,R	NA	S,R,R,R,R,R,R,R,R,R	R
MDA-MB-231	0	0	0	0	0	0	0	0	0	0	1	0	S,R,R,R,R,R,R,R,R,R	R	S,R,R,R,R,R,R,R,R,R	R
MDA-MB-361	1	1	1	1	0	0	0	1	0	1	0	0	R,S,S,R,R,R,S,R,R,R	R	R,S,S,R,R,R,S,R,R,R	R
MDA-MB-415	0	0	0	0	1	0	0	1	0	0	0	0	R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R	NA
MDA-MB-435s	0	0	0	0	0	0	0	0	0	0	0	0	S,R,R,R,R,R,R,R,R,R	R	S,R,R,R,R,R,R,R,R,R	R
MDA-MB-436	0	0	0	0	0	0	0	0	0	0	0	0	S,R,R,R,R,R,R,R,R,R	R	S,R,R,R,R,R,R,R,R,R	NA
MDA-MB-468	0	0	0	0	1	0	1	0	0	0	0	0	R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R	R
T47D	0	0	0	1	0	0	0	1	0	0	0	0	R,R,R,R,R,R,R,R,R,R	R	R,R,R,R,R,R,R,R,R,R	R
ZR-75-1	0	0	0	0	0	0	0	0	0	0	0	0	S,R,R,R,R,R,R,R,R,R	NA	S,R,R,R,R,R,R,R,R,R	R

Notes: "1" and "0" indicates the corresponding profile is positive (over-expressed, amplified or mutated) and negative (not over-expressed, amplified or mutated) respectively. "S", "R", "NA", "s-mut", "r-mut", 'a-mut', "amp", "over exp", "pre", and "act" stands for sensitive to drug, resistant to drug, no available drug sensitivity, drug sensitive mutation, drug resistance mutation, activating mutation, amplification, over expression, predicted drug sensitivity, and actual drug sensitivity respectively. The prediction methods M3, A1, E1, E2, C1, C2, C3, C4, C5, C6, and C7 are described in the text.

Table 4-10 The distribution and coexistence of drug sensitizing mutation, amplification and expression profiles, and the drug resistance mutation and expression profiles in NSCLC and breast cancer cell-lines

Cancer: NSCLC
Main Target for the Treatment of Specific Cancer: EGFR
Drugs Evaluated: gefitinib (D1), erlotinib (D2), and lapatinib (D3)

Drug Sensitizing or Resistance Profile (*index*)	**Number of Cell-Lines with This Profile**	**Number of These Cell-Lines with Another Sensitizing Profile**			**Number of These Cell-Lines with Another Resistance Profile**												**Number of These Cell-Lines Sensitive/ Resistant to Drug**		
		Drug Sensitizing Profile			***Drug Resistance Profile***														
Drug Sensitizing profile		***S1***	***S2***	***S3***	***R1***	***R2***	***R3***	***R4***	***R5***	***R6***	***R7***	***R8***	***R9***	***R10***	***R11***	***R12***	***D1***	***D2***	***D3***
EGFR s-mut (***S1***)	7		3	2	2						1						4/3	4/3	3/4
EGFR amp(copy no≥4) (***S2***)	4	3		2													4/0	4/0	3/1
EGFR over exp (***S3***)	5	2	2									1					3/2	3/1	3/2
Drug Resistance profile																			
EGFR r-mut (***R1***)	2										1						0/2	0/2	0/2
RAS a-mut (***R2***)	22						1				7	2				1	0/20	0/22	0/21
BRAF a-mut (***R3***)	3					1					1						0/3	0/3	1/2
PIK3CA a-mut (***R4***)	2										2						0/1	0/2	0/2
PTEN loss (***R5***)	0																		
HER2 over exp (***R6***)	2										2						1/1	0/1	2/0
HER3 over exp (***R7***)	18				1	7	1	2		2		1				1	1/14	0/17	3/12
MET over exp (***R8***)	5					2					1						0/5	0/5	1/4
PDGFR over exp (***R9***)	4																		
IGF1R over exp (***R10***)	0																		
FGFR1 over exp (***R11***)	0																		
VEGFR2 over exp (***R12***)	1					1					1						0/1	0/1	0/1
Cancer: Breast cancer																			

Main Target for the Treatment of Specific Cancer: HER2 **Drugs Evaluated:** trastuzumab (D1) and lapatinib (D2)														
Drug Sensitizing or Resistance Profile (*index*)	**Number of Cell-Lines with This Profile**	**Number of These Cell-Lines with Another Sensitizing Profile**		**Number of These Cell-Lines with Another Resistance Profile**									**Number of These Cell-Lines Sensitive/ Resistant to Drug**	
		Drug Sensitizing Profile		***Drug Resistance Profile***										
Drug Sensitizing profile		***S1***	***S2***	***R1***	***R2***	***R3***	***R4***	***R5***	***R6***	***R7***	***R8***	***R9***	***D1***	***D2***
HER2 amp (copy no ≥4) (S1)	11	11		5	1			4	1				5/6	4/3
HER2 over exp (S2)	13		11	7	1			6	2				5/8	5/4
Drug Resistance profile														
PIK3CA a-mut (R1)	11							5	3				1/8	1/8
PTEN ina-mut (R2)	4						1	2					0/4	0/3
PTEN loss (R3)	0													
EGFR over exp (R4)	1				1								0/1	0/1
HER3 over exp (R5)	9			5	2				3				1/6	0/6
IGF1R over exp (R6)	4			3				3					0/2	0/4
AXL over exp (R7)	3												0/2	0/2
ESR1 over exp (R8)	0													
MET over exp (R9)	0													

4.4 Summary

In summary, collective analysis of mutation, amplification and expression profiles of target, bypass genes, and drug-resistant downstream signaling and regulatory genes are potentially useful for facilitating drug sensitivity prediction. Drug sensitivity prediction can be greatly potentiated by collective analyzing the profiles of target, bypass genes, and drug-resistant downstream signaling and regulatory genes. In the long run, an accurate patient differentiation, better safety profile, improved response rate and personalized treatment can also be archieved by this system-oriented approach. Development, integration and expanding application of next generation sequencing [622], microarrays [623], and copy number variation [624] detection tools and methods coupled with expanded knowledge of cancer biology and drug resistance bypass mechanisms enable more accurate prediction of drug sensitivity.

5 Collective approach for tumor-specific antigen discovery

An important application of high-throughput technique, such as DNA microarray, is to discover potential disease targets which can be used for therapeutic molecule design and achieve the goal of effective disease treatment and prevention. Our previous study indicates that disease molecular profiles are useful sources of discover new diagnostic and effective therapeutic targets with targeted disease population. Tumor-specific antigens (TSAs), often overexpressed or mutated on particular tumors, but nonetheless present at lower frequencies in normal tissues, can potentially be explored for applications in cancer diagnosis and immunotherapy vaccine. The lack of effective vaccines for many cancers has prompted strong interest in improved TSA search methods. In this chapter, a collective method of analyzing genome-scale TSA was developed, which helps to identify the novel TSA in human cancer genome scale. We collectively analyzed genome-scaled tumor-specific somatic mutations, microarray gene expression data, and T-cell recognition of peptides derived cancer genome. Collectively considering the available profiles of TSAs showed a fairly prediction performance in melanoma and lung cancer. With improved data quality and analysis methods, the collective approach is potentially useful for facilitating genome-scale TSA search.

5.1 Introduction

Among the tumor antigens, some may be tumor specific, namely tumor-specific antigens (TSAs), while others may be also expressed by normal tissues, namely tumor-associated antigens (TAAs). In practice, cancer antigens targeted by active immunotherapies have

more often been tumor associated: overexpressed or mutated on tumors, but nonetheless present at lower frequencies in normal tissues [212]. TSA is therefore have been explored as sources of vaccines for cancer immunotherapy and immunoprevention [217,218,220] and biomarkers for cancer diagnosis.

TSAs elicit cancer immunogenicity by presenting genetically variant and differentially over-expressed epitopes distinguished from those of normal cells [211,625,626]. Tumor-specific mutations in these "self-antigens" provide structural or physicochemical features distinguished from those of normal cells for them to be recognized as"non-self" [626]. Moreover, these antigens are primarily genes prevalently over-expressed in cancer patients, and less expressed in normal persons to override immune tolerance thresholds [627] and weak immunogenicity [211], and to prevent outgrowth of non-immunogenic variants in cancer patients[216]. Many known TSAs are cytotoxic T lymphocyte (CTL) epitopes [216,217,220] that are cleaved from intracellular protein antigens by proteasomes, transported to the endoplasmic reticulum by TAP, assembled with MHC-I, and presented at the surface of cancer cells to be recognized by CTLs [221] .

In recent years, the diagnosis and treatment of cancers have improved, but the long-term survival rate, especially the survival rate for advanced cases, still has not been markedly increased. Therefore, it is very important to search for tumor-specific antigen (TSA) and tumor-associated antigen (TAA) to ensure the early detection, early diagnosis and early treatment of list of various cancers. In spite of extensive efforts, effective vaccine for most cancer types is still lacking partly due to difficulties in overcoming immunosuppressive activities[628,629], selective pressures towards non-epitope variants[630] and other

factors in immune tolerance and evasion [631,632]. There is a need to search for cancer vaccines from more diverse sources[37,216,217,219] that takes into consideration not only tumor-specific mutations and MHC-binding but also the expression profiles of the antigens, processing and transport of the epitopes, and availability of T-cell repertoire in specific tumors [220]

The discovery of potential therapeutic targets from human cancers is greatly facilitated by microarray technology [211,629]. Furthermore, there have been significant progresses in genome-wide profiling of tumor-specific somatic mutations[633-635] and in improving T-cell epitope prediction by collective analysis of proteasomal cleavage, TAP mediated transport and MHC-binding[223,231,247-249,251,261]. These progresses make it possible to conduct genome-scale search of TSAs via collective analysis of tumor-specific mutations, expression profiles, and T-cell recognition of the epitopes that include the cleavage, transport and MHC-binding of the epitopes.

This chapter covers the usage of collective approach in genome-scale search of TSAs from melanoma and lung cancer, which early on was found to have tumor-specific antigens and has been targeted frequently using the protein or peptide approach. Genome-scaled tumor-specific somatic mutations, microarray gene expression data, and in silico T-cell recognition were incorporated into this collective approach. Collectively considering the available profiles of TSAs showed a fairly prediction performance in melanoma and lung cancer.

5.2 Materials and Methods

5.2.1 Collection of genomic, mutation and expression data

The human genome sequence (release 38) was obtained from NCBI database (http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/), which contains 180,000 full-length protein-coding transcripts. Tumor-specific somatic mutation data for melanoma and lung cancer were from comprehensive literature search from sources such as COSMIC database (release 35) and a number of other publications in PubMed (1970−March 2008) using key words "mutation" and "melanoma" or "lung cancer" in the title or abstract.Our search identified 2361 articles (2315 from COMIC database and 46 from PubMed), which report 841 and 340 somatic nonsynonymous mutations in 491 and 338 encoded proteins for melanoma and lung cancer respectively.

The microarray gene expression data for melanoma (GSE4845) and lung cancer (GSE1037) were from GEO database. The melanoma dataset contains the expression profiles of 33,000 genes from 12 melanoma patients and 3 normal persons, and the lung cancer dataset is composed of the expression patterns of 14,211 genes and 15,276 ESTs from 27 lung cancer patients and 19 normal persons.

5.2.2 Collection of tumor-specific antigen

Tumor cells expressing epitopes derived from TSAs and TAAs is recognized and destroyed by Cytotoxic T-lymphocytes (CTLs) derived from T8-lymphocytes (CD8+ T cells) [636] . CTLs recognize antigen on target and APCs[637] as epitopes composed of peptide fragments, 8–12 amino acids long, that are completed to MHC-I molecules [637-639]. Several CTL epitopes restricted by HLA-A2, the most common human

histocompatibility molecule (40–50%), have been previously reported [640]. Definition of HLA-A restricted TSAs is thus an important step in the development of specific CTL-based cancer immunotherapies.

In this study, with the focus of HLA-A restricted TSAs, the collection of human tumor antigens recognized by CD8+ CTLs were carried out by search Cancer Immunity database (http://www.cancerimmunity.org/peptidedatabase/Tcellepitopes.htm) and literatures. The 12 known melanoma TSAs of the 5 MHC-I alleles A1, A2, A3, A24, and A6801, and the 4 known lung cancer TSAs of the 3 MHC-I alleles A1, A2, A6802 were collected in this study for validation purpose. The sequence, binding HLA alleles, and information of host proteins of these known TSAs are given in Tables 5-2 and Tables 5-3 for melanoma and lung cancer respectively.

5.2.3 Computational procedures

5.2.3.1 ***Identify overexpressed candidates from microarray data***

All of the known tumor-specific somatic mutations of melanoma and lung cancer were substituted into the protein products of human genome sequence to generate the human cancer genomes for melanoma and lung cancer. Since the length of identified the 16 known TSAs are in range of 8 to 11 amino acids, only the 8−11mer tumor-specific mutant peptides were extracted from the corresponding encoded protein sequences such that each peptide contains at least one mutated residue uniquely found in melanoma or lung cancer. The expression levels of the corresponding host proteins of these tumor-specific mutant peptides were evaluated.

We calculated the percentage of cancer patients whose selected genes were at least x-fold overexpressed, in comparison with y% of normal people. A range of x (1.5, 2, 2.5, 3) and y (50% to 100%, which means to compare with at least half of normal samples) values were tested. When y value was varied at a fixed value of x, a stable percentage (>40%) of patients carrying the overexpressed genes were observed with the increase in y. We therefore fix the x to 2 and y to 50%.

Therefore, peptides were selected such that their host proteins are expressed at higher levels (2 fold-change) in >40% of patients (5 out of 12 for melanoma, and 11 out of 27 for lung cancer) than those in >50% of normal persons (2 out 3 for melanoma, and 10 out of 19 for lung cancer) in the melanoma and lung cancer datasets respectively.

5.2.3.2 Derivation of structural and physicochemical properties from peptide sequence

As introduced in the chapter of methodology, a feature vector can be constructed for representing the structural and physicochemical properties of a *peptide*. Given the sequence of a peptide, its amino acid composition and the properties of every constituent amino acid are computed and then used to generate this vector. The computed amino acid properties include hydrophobicity, normalized Van der Waals volume, polarity, polarizability, charge, surface tension, secondary structure and solvent accessibility [254].

For each of these properties, amino acids are divided into three or six groups such that those in a particular group are regarded to have approximately the same property. Three descriptors, composition (C), transition (T), and distribution (D), are introduced to describe global composition of each of these properties. Overall, there are 51 elements

representing these three descriptors: 6 for C, 15 for T and 30 for D, with a six grouped property; and 21 elements: 3 for C, 3 for T and 15 for D, with a three grouped property.

Construction of peptide feature vectors can be illustrated by the generation of the amino acid composition descriptors of two hypothetical peptides, sequence I (AEAAAEAEEAAAA) and sequence II (AEAEEEAAEEAEEEAAE). Sequence I contains 9 alanines (n1=9) and 4 glutamic acids (n2=4) and sequence II includes 7 alanines (n1=7) and 10 glutamic acids (n2=10). The composition is C1= (n1×100/ (n1+n2), n2×100/ (n1+n2)) = (69.23, 30.77) for sequence I, and C2= (41.18, 58.82) for sequence II. There are 6 A=>E and E=>A transitions in sequence I, and 9 such transitions in sequence II. The total number of transitions is 12 in sequence I and 16 in sequence II. The percent frequency of transition is thus T1= (6/12) ×100=50.00 for sequence I, and T2= (9/16) ×100=56.25 for sequence II. The first, 25%, 50%, 75% and 100% of As and Cs are located within the 1^{st}, 4^{th}, 10^{th}, 11^{th}, 13^{th} residues and the 2^{nd}, 2^{nd}, 6^{th}, 8^{th}, 9^{th} residues respectively for sequence I, and within the 1^{st}, 3^{rd}, 8^{th}, 15^{th}, 16^{th} residues and the 2^{nd}, 5^{th}, 9^{th}, 13^{th}, 17^{th} residues respectively for sequence II. The distribution is then D1=(1/13×100, 4/13×100, 10/13×100, 11/13×100, 13/13×100, 2/13 ×100, 2/13×100, 6/13×100, 8/13×100, 9/13×100)=(7.69, 30.77, 76.92, 84.62, 96.67, 15.38, 15.38, 46.15, 61.54, 69.23) for sequence I, and D2=(1/17 ×100, 3/17×100, 8/17×100, 15/17×100, 16/17×100, 2/17 ×100, 5/17×100, 9/17×100, 13/17×100, 17/17×100)=(5.88, 17.65, 47.06, 88.23, 94.12, 11.76, 29.41, 52.94, 76.47, 100.00) for sequence II. Overall, the amino acid composition feature vector is **x**1=(C1, T1, D1)=(69.23, 30.77, 50.00, 7.69, 30.77, 76.92, 84.62, 96.67, 15.38, 15.38, 46.15, 61.54, 69.23) for sequence I, and **x**2=(C2, T2, D2)=(41.18, 58.82, 56.25, 5.88, 17.65, 47.06, 88.23, 94.12, 11.76, 29.41, 52.94, 76.47, 100.00) for sequence II. The two vectors **x**1 and **x**2 thus have equal length, which is useful for classification of peptides

of variable lengths by using statistical learning methods.

5.2.3.3 ***Application of computational method for T-cell recognition prediction***

The differentially overexpressed peptides and their neighboring residues in the host proteins were analyzed by NetChop3.0 (Nielsen et al., 2005), a neural networks based prediction tool developed form the experimental validated MHC Class I molecules[641], to determine which of them are cleavable by the proteasome (C-term 3.0 score with cutoff larger than 0.5). The predicted cleavable peptides were then screened by NetCTL1.2 [223] to select those transportable by TAP (TAP transport efficiency score with cutoff lower than 1). The selected proteasome cleavable and TAP transportable peptides were then screened by using MHC-BPS[261], a SVM-based MHC binding prediction software developed by our research group, to determine which of them can bind to MHC-I alleles A1, A2, A3, A24, A6801 for melanoma and MHC-I alleles A1, A2, A0301, A6802 for lung cancer respectively.

Table 5-2 Known melanoma tumor-specific antigens presented by MHC-I. The label “E”, “C”, “T”, “B” indicates the tumor-specific antigen was selected by gene expression, proteasome cleavage, TAP-mediated transport, HLA-binding analysis respectively.

HLA allele	TSA and mutated residue(underline)	Sequence location	Length of TSA	Host protein and gene name	Function of host protein	Reference (PMID)	TSA passed search steps
A1	FLEGNEVGKTY(G)	446–455	11	Hedgehogacyltransferase (MART2)	Hedgehogsignaling	11160356	CTB
A1	ILDTAGREEY(Q)	55–64	10	Neuroblastoma RAS viral (v-ras) oncogene homolog (NRAS)	Signaling	11971032	C
A2	ACDPHSGHFV(R)	23–32	10	Cyclin-dependent kinase 4 (CDK4)	Cellcycle	7652577	CTB
A2	GIVEGLITTV(M)	168–177	10	Glyceraldehyde3-phosphate dehydrogenase(GAPDH)	Energymetabolism	15614045	CTB
A2	SLADEAEVYL(H)	281–290	10	Growth arrest-specific 7 (GAS7)	Cellcycle	15614045	ECTB
A2	LLLDDLLVSI(S)	163–172	10	Peroxiredoxin5(PRDX5)	Oxidativestress	15695408	ECTB
A24	SYLDSGIHF(S)	29–37	9	Catenin 1(CTNNB1)	Cell–celladhesion/Wntsignaling	8642260	ECTB
A3	KILDAVVAQK(E)	668–677	10	Elongation factor Tu GTP binding domain containing 2(EFTUD2)	RNAprocessing	16247014	ECTB
A3	TLDWLLQTPK(G)	179–188	10	Glycoproteinnmb (GPNMB)	Melanosomalprotein	16247014	CT
A3	KINKNPKYK(E)	911–919	9	MyosinclassI(MYO1B)	Cellularmotility	10064075	ECTB
A3	KIFSEVTLK(P)	192–200	9	Sirtuin-2(SIRT2)	Transcriptionalsilencing	16247014	CTB
A6801	EAFIQPITR(S)	322–330	9	Melanoma associated antigen-3(MUM3)	Nucleicacidmetabolism	10820291	ECTB

Table 5-3 Known lung cancer tumor-specific antigens presented by MHC-I. The label “E”, “C”, “T”, “B” indicates the tumor-specific antigen was selected by gene expression, proteasome cleavage, TAP-mediated transport, HLA-binding analysis respectively.

HLA allele	TSA and mutated residue (underline)	Sequence location	Length of TSA	Host protein and gene name	Function of host protein	Reference (PMID)	TSA passed search steps
A1	FLEGNEVGKTY(G)	446–455	11	Hedgehog acyltransferase (MART2)	Hedgehog signaling	11160356	ECTB
A2	FIASNGVKLV(K)	118–127	10	Actinin, alpha 4 (ACTN4)	Adhesion	11358829	ECTB
A2	FLDEFMEG(A)	224–231	8	Malic enzyme 1 (ME1)	Energy metabolism	11325844	EB
A6802	ETVSEQSNV(E)	581–589	9	Eukaryotic translation elongation factor 2 (EEF2)	Protein translation	9823325	ECTB

5.3 Results and Discussion

5.3.1 Performance of collective approach in genome-scaled TSAs identification

The results of the collective approach in predicting melanoma and lung cancer TSAs from the human cancer genomes are given in Tables 5-4 and Tables 5-5 respectively. A total of 36, 250, 31, 22, and 8 putative melanoma TSAs were predicted for HLA A1, A2, A24, A3, and A6801 alleles, and a total of 17, 359, and 14 putative lung cancer TSAs were predicted for HLA A1, A2, and A6802 alleles respectively.

The differentially overexpressed peptides and their neighboring residues in the host proteins were analyzed by NetChop3.0[231] (Nielsen et al., 2005), a neural networks based prediction tool developed form the experimental validated MHC Class I molecules[641], to determine which of them are cleavable by the proteasome (C-term 3.0 score with cutoff larger than 0.5). The predicted cleavable peptides were then screened by NetCTL1.2 [223] to select those transportable by TAP (TAP transport efficiency score with cutoff lower than 1). The selected proteasome cleavable and TAP transportable peptides were then screened by using MHC-BPS[261], a SVM-based MHC binding prediction software developed by our research group, to determine which of them can bind to MHC-I alleles A1, A2, A3, A24, A6801 for melanoma and MHC-I alleles A1, A2, A0301, A6802 for lung cancer respectively.

The predicted TSAs include 50% and 75% of the 12 and 4 known T-cell defined tumor-specific melanoma and lung cancer antigens in Cancer Immunity database. Overall, the yields, hit rates and enrichment factors (with respect to mutation analysis alone) of the collective approach are 50% and 75%, 1.9% and 0.8%, and 29 and 35 for melanoma and

lung cancer respectively. The yields are slightly less than the 70–90% levels of the prediction tools that solely predict T-cell epitopes [238,240,242,244], suggesting that the collective approach is capable of identifying TSAs at accuracy levels not too far away from those of the tools that only predict T-cell epitopes.

Table 5-4 Results of genome-search of melanoma tumor-specific antigens by collective analysis of mutation, expression and T-cell recognition. The number of "√" represents the number of known tumor-specific antigens passed a particular search step.

Known melanoma tumor-specific antigens			Putative tumor-specific antigens in human genome identified by integrated analysis of mutation, expression and T-cell recognition					Known melanoma tumor-specific antigens by collective analysis			
HLA allele	Antigen length	No and host gene name of known melanoma tumor-specific antigens of this length	No of tumor-specific mutated peptides of this length in coding region of human genome	No of peptides in column 4 whose host protein is differentially over-expressed in >40% patients (5 out of 12) with respect to >50% of normal person (2 out of 3)	No of peptides in column 5 predicted to be proteasome cleavable	No of peptides in column 6 predicted to be TAP substrate	No of peptides in column 7 predicted to be HLA binder	No of known melanoma tumor-specific antigens selected	Yield	Hit rate	Enrichment factor (with respect to mutation analysis)
A1	10	1	6,211(√)	985	566	326	15	0	–	–	–
A1	11	1	6,819(√)	1,081	598	356	21	0	–	–	–
A2	10	4	6,211(√√√√)	985(√√)	566(√√)	326(√√)	250(√√)	2	50%	0.8%	12
A24	9	1	5,601(√)	888(√)	503(√)	291(√)	31(√)	1	100%	3.2%	181
A3	9	2	5,601(√√)	888(√)	503(√)	291(√)	9(√)	1	50%	11.1%	311
A3	10	2	6,211(√√)	985(√)	566(√)	326(√)	13(√)	1	50%	7.7%	240
A6801	9	1	5,601(√)	888(√)	503(√)	291(√)	8(√)	1	100%	12.5%	700
Overall	9–11	12	18,631	2,945	1,667	973	316	6	50%	1.9%	29

Table 5-5 Results of genome-search of lung cancer tumor-specific antigens by collective analysis of mutation, expression and T-cell recognition. The number of "√" represents the number of known tumor-specific antigens passed a particular search step

Known lung cancer tumor-specific antigens			Putative tumor-specific antigens in human genome identified by integrated analysis of mutation, expression and T-cell recognition					Known lung cancer tumor- specific antigens by collective analysis			
HLA allele	Antigen length	No and host gene name of known lung cancer tumor-specific antigens of this length	No of tumor-specific mutated peptides of this length in coding region of human genome	No of peptides in column 4 whose host protein is differentially over-expressed in >40% patients (11 out of 27) with respect to >50% of normal person (10 out of 19)	No of peptides in column 5 predicted to be proteasome cleavable	No of peptides in column 6 predicted to be TAP substrate	No of peptides in column 7 predicted to be HLA binder	No of known melanoma tumor-specific antigens selected	Yield	Hit rate	Enrichment factor (with respect to mutation analysis)
A1	11	1	5,366(√)	4,301(√)	1,690(√)	844(√)	17(√)	1	100%	5.9%	316
A2	8	1	3,916(√)	3,137(√)	1,247	657	54	0	–	–	–
A2	10	1	4,884(√)	3,914(√)	1,550(√)	793(√)	305(√)	1	100%	0.3%	16
A6802	9	1	4,401(√)	3,526(√)	1,401(√)	730(√)	14(√)	1	100%	7.2%	314
Overall	8–11	4	18,567	14,878	5,888	3,024	390	3	75%	0.8%	35

Evidences indicative of higher expression levels of some of the host proteins of these malnoma TSAs have been reported. CDK4 [642] and GPNMB [643] has been reported to be amplified in melanoma. Hedgehog signaling has been found to be required for melanoma [644], which suggests the possibility of higher levels of hedgehog acyltransferase in melanoma patients. Therefore, increased sampling sizes as well as the

enhancement of the quality of measurement and analysis of microarray data [645] may enable the further improvement of the TSA prediction performance of the collective approach in genome-scale search studies.

5.4 Conclusion

The collective consideration of somatic mutations, expressions, and T-cell recognition appears to show fairly good capability in predicting TSAs in genome-scale search campaigns. The numbers of predicted putative TSAs from genome-scale search studies are within the range manageable by typical screening campaigns, and the hit rates are enriched to levels that enable effective identification of TSAs. The prediction performance, the yield and hit rates, of the collective approach may be further improved by expanding the sampling size as well as improving the quality of measurement and analysis of gene expression data[645]. With expanding knowledge of the respective peptides and the further development of in silico tools, the performance for predicting proteasomal cleavage, TAP mediated transport, and HLA-binding are expected to be further improved [223,231,261], which helps to increase the capability of the collective approach in facilitating the genome-scale search of TSAs.

6 Lung adenocarcinoma survival marker selection

Micoarrays have been explored for deriving molecular signatures, subsets of genes differentially expressed in patients of different outcomes, for determining disease-outcomes, mechanisms, targets, and treatment-strategies. While exhibiting good predictive performance, derived signatures are unstable due to noises arising from measurement variability and biological differences. Improvements in measurement, annotation and signature-selection methods have been proposed. We explored a new gene signature selection method by incorporating consensus scoring of multiple random sampling and multi-step evaluation of gene-ranking consistency for maximally avoiding erroneous elimination of predictor-genes. The best prediction performance was achieved in cancer biomarkers discovery of a well-studied 62-sample colon-cancer dataset. This chapter provides a case study of applying this gene selection system to survival biomarker selection from a 86-sample lung adenocarcinoma dataset. The derived gene signatures of 10 sampling-sets, composed of 5,000 training-testing sets, are fairly stable with 40%~62% of all predictor-genes shared by all 10 sampling-sets. These shared predictor-genes include 15 cancer-related and 5 cancer-implicated genes. The predictive ability of these survival markers are evaluated by neural network models, SVM models, and unsupervised hierarchical clustering methods. The derived signatures outperform all previously-derived signatures in predicting patient outcomes from an independent dataset, suggesting its usefulness for deriving stable signatures in facilitating biomarker and target discovery.

6.1 Introduction

The fundamental goals of discovery cancer biomarker includes the prediction of cancer stages, the likelihood of disease redeveloping following an apparent resolution of a disease or to predict outcomes such as life expectancy, survivability, progression, and drug sensitivity after initial diagnosis [646]. Molecular risk factors are commonly used prognostically to stratify the subtype of cancer patients and to prescribe the appropriate treatment regimens that match their risk profiles, so that proper treatment regimen can be applied and ultimately extend the survival of the patients[11]. Certain cancer types, such as lung cancer, are prescribed various types of treatments such as chemotherapy and radiation therapy based on the known molecular factors, such as the status of EGFR expression[647,648]; however, there is no assurance that metastases and recurrence will never occur[647,648]. The ability to predict the metastases and invasions behavior of lung cancer still remains one of the greatest clinical challenges in thisfield.

In order for clinician developing proper treatment regime and ultimately extend the survival of the patients, the accurate identification of cancer subtype and prognosis effect is crucial [11]. Extensive studies have been recently conducted to discover cancer subtype and prognostic prediction based on disease and patients molecular expression profiles[4,5,117-120]. The successful rate of current prediction is low due to the complex and very heterogeneous of lung cancer[120]. However, studies on lung adenocarcinoma survival marker discovery and prognosis prediction still provide a platform for subtype discovery and prognostic prediction based on disease and patients molecular details [4,5,117-120]

Lung cancer is the leading cause of cancer-related mortality not only in the United States but also around the world[649]. In 2007, lung cancer accounts for 29% of all cancer deaths (31% in men, 26% in women) in US [650]. The 2 main types of lung cancer are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC); NSCLC accounts for approximately 85% of all cases of lung cancer[651]. The NSCLC can be further classified as squamous cell carcinoma and adenocarcinoma, and large-cell lung cancer[652]. The proportional of squamous cell carcinoma is around 20~25%, and the proportional of adenocarcinoma is 50~60% [651]. Despite advances in early detection and standard treatment, prognosis for both NSCLC and SCLC lung cancers are poor. NSCLC is often diagnosed at an advanced stage and has a poor prognosis. The average survival time of advanced NSCLC is 6 months for untreated patients, and 9 months for patients treated with chemotherapy [651]. Five-year survival rate is 60~70% for patients with stage I disease and zero for patients with stage IV disease [651]. The treatment and prevention of lung cancer are major unmet needs that can probably be improved by a better understanding of the molecular origins and evolution of the disease.

The lung cancer patients can be roughly stratified from the morphological assessment based on conventional sputum cytology and chest radiography. These techniques have not yet demonstrated an impact on decreasing lung-cancer mortality [653]. In one study, only 41% of cases that independent lung pathologists agreed on lung adenocarcinoma subclassification [654]. Recently some specific indicators, including tumor size, poor differentiation and high tumor-proliferative index, have been identified to predict the survival of lung cancer patients [655-658], However, these indicators have only limited power in survival prediction.

It is suggested that common molecular features may be able to predict such outcome discrepancies among patients more reliably. For instance, the efficacy of EGFR antagonists has been shown to depend on expression and mutation status of its target EGFR in the tumor [659]. Also, the beneficial effect of chemotherapies might depend on the expression of certain enzymes such as thymidine synthetase for pemetrexed[660]. Thus, improved classification of NSCLC by using molecular indicators is of considerable clinical interest.

The development of microarray technology holds the potential to find molecular biomarkers of lung cancer subtype and outcome prediction systematically [661-664]. These biomarkers allow new insights in the process of lung carcinogenesis, and they may provide new tools for determination of prognosis and identification of innovative treatments. However, the reproducibility of gene expression signatures to predict high-risk of relapse or recurrence is rarely reported. The molecule marker selection is strongly dependent on the patient samples, causing the significantly different marker signatures in different groups for lung adenocarcinoma prognosis (Table 6-1) [662,663] and diminishing their application potential for prognosis [139]. Moreover, the prognostic power of previous selected survival genes for individual patients was seldom reported in their studies [662,663]. Guo et al. provided the prediction accuracy for their selected survival genes. However, their selected survival genes were only applicable to one dataset, and the predictive power to the independent dataset was very limited [661]. It is therefore highly desirable to identify stable molecular markers that can reliably determination of prognosis and predict specific subgroups of high- and low-risk patients. This would be helpful to select the most appropriate therapy for individual patient

Table 6-1 Statistics of lung adenocarcinoma survival marker signatures from references

Study	Number of selected survival genes in signature	Number of genes selected by other N studies				
		4	3	2	1	0
Lu et al [665]	125	0	0	0	8	116
Chen et al [666]	16	0	0	0	0	16
Xu et al [667]	5	0	0	0	2	3
Beer et al [662]	100	0	0	0	8	92
Guo et al [661]	37	0	0	0	4	34

In this chapter, we developed a new recursive feature selection method based on a model built from SVM in identifying novel survival markers with respect to the interactions among genes. The consensuses scoring of multiple random sampling and the evaluation of gene-ranking consistency have been embedded in the recursive feature selection system to reduce the chances of erroneous elimination of predictor-genes and improving the stability across the different sample groups. The new method has been applied to identify important biomarkers in prediction of the survivability of individual patients with lung adenocarcinoma. A total of 21 genes were selected after repeating smapling of the same experiment for 10 times. . Results show that the prediction models can accurately predict the clinical outcome for individual patients with lung adenocarcinoma by use of independent datasets. The differential expression analysis, function prediction, and literature searches of the identified biomarkers implies that this group of genes plays important roles in lung adenocarcinoma progress and may contain novel therapeutic targets.

6.2 Materials and Methods

6.2.1 Lung adenocarcinoma microarray datasets and data preprocess

Two independent datasets of clinical samples were used for lung adenocarcinoma survival marker gene selection and validation of the effect of our selected genes. The original gene expression profiles of patient samples have been reported in previous publications [662,663].

The dataset for survival marker gene selection contained the gene expression profiles from 86 primary lung adenocarcinomas (Beer's dataset) [443,662], including 67 stage I and 19 stage III tumors, from oligonucleotide arrays seen at the University of Michigan Hospital between May 1994 and July 2000. This gene expression profile, containing 7129 gene expression levels, was obtained before surgery. Sixty two patients survived (survivable patients) whereas 24 patients died at last follow-up (non-survivable patients). The detailed clinical information of samples is listed in Appendix Table S12. For preprocessing, those genes with little variation (less than 2) across all of 86 samples were removed, and 6009 genes were used for survival gene selection [139,331]. Although this gene filtering procedure may potentially result in the loss of some information, it in this manner decreased the possibility of introduce the noise into the machine learning method and clustering algorithm which might be strongly influenced by genes with little or no expression[443,662].

The robustness of our selected signatures in predicting survivability in lung adenocarcinomas was tested using oligonucleotide gene-expression data obtained from a

completely independent lung adenocarcinoma dataset (Bhattacharjee's dataset) [663,668]. To ensure equivalent testing power and comparability of samples, 84 primary lung tumor samples of which at least 40% samples being cancer cells were selected [662]. In these 84 samples, 41 patients were alive at last follow-up (survivable patients), whereas 43 died (non-survivable patients). The detailed sample clinical information is listed in Appendix Table S13.

In order to present a statistically meaningful evaluation, signature selection was conducted based on multiple random sampling on the Beer's dataset [662]. In multiple random sampling, this dataset was randomly divided into a training set containing 43 samples (including 12 poor outcome samples and 31 good outcome samples) and an associated test set containing the other 43 samples (including the other 12 poor outcome samples and 31 good outcome samples). To reduce computational cost, 5,000 training-test sets, each containing a unique combination of samples, were generated. These 5,000 training-test sets were randomly placed into 10 sampling groups; each containing 500 training-test sets. Every sampling group was then used to derive a signature by using the similar way as colon cancer marker discovery. Finally, the 10 different signatures derived from these sampling groups were compared in order to test the level of stability of selected predictor-genes.

Repeated random sampling was used to archieve statistically meaningful evaluation in analyzing Beer's dataset [139]. In multiple random sampling, the dataset was randomly divided into a training set containing 43 samples (including 12 poor outcome samples and 31 good outcome samples) and an associated test set containing the other 43 samples

(including the other 12 poor outcome samples and 31 good outcome samples). To reduce computational cost, 5,000 training-test sets, each containing a unique combination of samples, were generated. These 5,000 training-test sets were randomly placed into 10 sampling groups; each containing 500 training-test sets. Finally, the 20 different signatures derived from these sampling groups were compared in order to test the level of stability of selected predictor-genes.

Each of the 10 sampling groups was used to derive a set of survival marker genes. In the 500 training-testing sets in every sampling group, each training-set was used to select genes by RFE based on SVM system. For all iterations and testing-sets, SVM system employed a set of globally modified parameters which gave the best average class-differentiation accuracy over the 500 testing-sets.

In every sampling group, three gene-ranking consistency evaluation steps were implemented on top of the normal RFE procedures in all sampling groups:

(1) For every training-set, subsets of genes ranked in the bottom 10% (if no gene was selected in current iteration, this percentage was gradually increased to the bottom 40%) with combined score lower than the first top-ranked gene were selected such that collective contribution of these genes less likely outweighed higher-ranked ones;

(2) For every training-set, genes selected from the step (1) were further evaluated to choose those not ranked in the upper 50% in previous iteration so as to ensure that these genes were consistently ranked lower;

(3) A consensus scoring scheme was applied to genes selected from the step (2)such that only those appearing in >90% (if no gene was selected in current iteration, this

percentage was gradually reduced to 60%) of the 500 training-sets were eliminated..

6.2.2 Performance evaluation of survival marker signatures

The predictive capability of survival marker signatures were evaluated by using the SVM and PNN classification system on 500 randomly-generated training-testing sets generated from the Bhattacharjee's dataset [663] and the Beer's dataset [662]. For each training-testing set, the training data was used to construct a classifier model, whereas the testing data was used to evaluate the performance of the model. The predictive performance of selected signatures was evaluated by the accuracies for survival patients (Sensitivity, SE) (Equation 2-17), non-survivable patients (Specificity, SP) (Equation 2-18), and overall accuracies (Q) (Equation 2-16) over the 500 models. Besides the evaluation by using supervised classifiers, unsupervised hierarchical clustering analysis was also applied to evaluate the performance of signatures.

Hierarchical cluster analysis was conducted by using the selected survival on the 86 samples from Beer et al. [662] and the independent dataset from Bhattacharjee *et al.* [663]. Kaplan-Meier survival analysis, often referred as survival analysis, was used in this study together with hierarchical cluster analysis. This analysis is popularly employed in medical research to estimate the percentage of patients living for a certain amount of time after surgery. It allows the estimation of survival over time, even when patients drop out or are studied for different lengths of time. A typical application of Kaplan-Meier analysis involves (1) grouping patients into different categories, and (2) comparing the survival curves from those categories by the log-rank test to assess the statistical significance of the

difference among the survival curves for the categories. The Kaplan-Meier analysis was performed by using XLSTAT software [669].

6.3 Results and discussion

6.3.1 System of the lung adenocarcinoma survival marker selection

10 sets of survival marker genes were obtained. The number or predictor-genes in each sampling group ranged from 34 to 57 (Table 6-2, Appendices Table S14). The stability of selected signatures was evaluated from the percentage of predictor-genes shared across every sampling group. As shown in Table 6-2 and Table 6-3 a total of 21 predictor-genes were presented in all experiments, accounting 40% to 62% genes identified by 10 sampling groups. The identified survival markers shown a moderate stability when comparing to the results from 5 previous studies (Table 6-1), which shows that 5~125 selected predictor-genes in each of the 5 previous studies were seldom presented in the other 4 studies.

PNN and SVM classifiers were used to evaluate classifier accuracy of selected predictor-genes. The classification capability of selected 21 predicator-genes was further evaluated by PNN and SVM classifiers, hierarchical clustering method and Kaplan-Meier survival analyses as shown in Figure 6-1.

Figure 6-1 System for lung adenocarcinoma survival marker derivation and survivability prediction

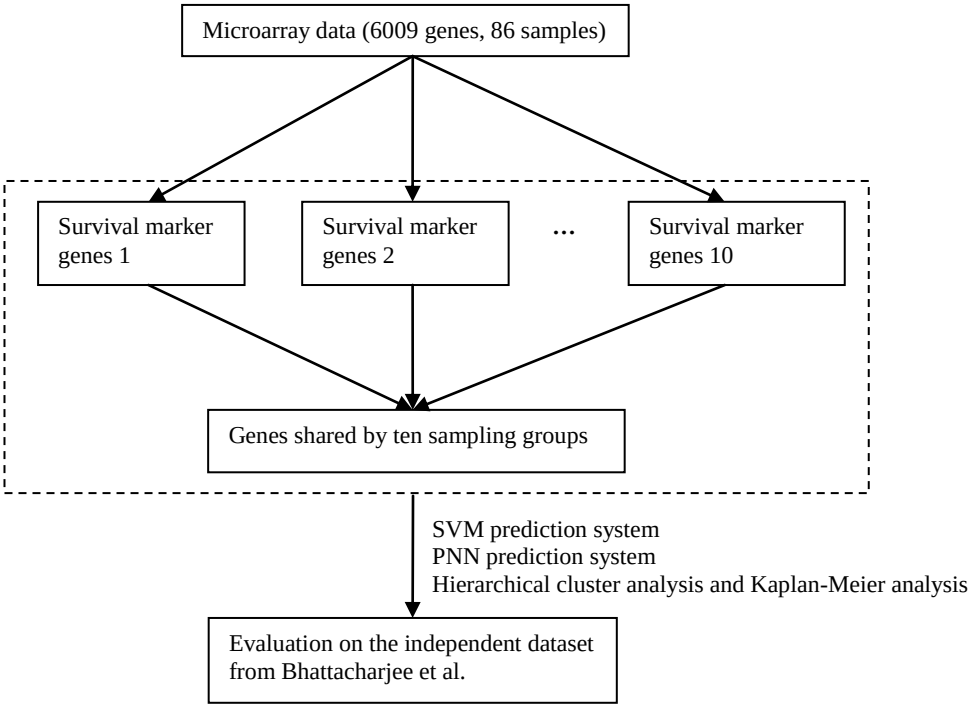


Table 6-2 Statistics of the lung adenocarcinoma survival markers by class-differentiation systems constructed from 10 different sampling-sets, each composed of 500 training-testing sets generated by random sampling.

Signature (method)	Number of selected survival genes in signature	Number of survival-genes also included in N other signatures derived by using different sampling-set									
		9	8	7	6	5	4	3	2	1	0
1	51	21	4	1	7	5	3	3	2	1	4
2	54	21	6	1	6	3	2	5	5	2	3
3	42	21	6	2	4	3	2	2	2	0	0
4	34	21	3	2	1	2	2	1	2	0	0
5	46	21	6	2	7	5	1	2	2	0	0
6	54	21	6	2	8	5	3	2	2	2	3
7	57	21	5	1	7	2	1	3	5	2	10
8	50	21	6	2	6	2	1	4	5	2	1
9	53	21	6	1	5	5	1	4	3	4	3
10	47	21	6	2	5	4	1	2	3	1	2

6.3.2 Consistency analysis of the identified markers

The optimal SVM parameter, σ, for the 10 sample sets were in the range of 41 to 46, and the highest average accuracies across the 10 sampling-sets were 84.1%~88.4% for non-survivable (those died at last follow-up) and 100% for survivable patients (those alive at last follow-up) respectively (Table 6-4). The accuracies for the 5,000 individual testing-sets ranged from 63.6%~100% for non-survivable and 100% for survivable patients respectively. The relatively small variations of optimal SVM parameters and prediction accuracies across the 10 sampling-sets suggest that the performance of the SVM class-differentiation systems constructed by using globally optimized parameters and RFE iteration steps are fairly stable across different sampling combinations.

Table 6-3 Gene information for lung adenocarcinoma survival markers shared by all of 10 signatures.

Gene Name	Gene description	Chromosome Location	Type	Family	Function in metagenesis	Gene Ontology: Function	Gene Ontology: Process	Pathway (from KEGG, Reactome, proteinlounge)	References
VEGF	vascular endothelial growth factor	6p12	Growth Factor	PDGF/VEGF Family of Growth Factors	Angiogenesis, therapeutic target for lung cancer therapy	extracellular matrix binding; growth factor activity; growth factor activity; heparin binding; protein binding; protein homodimerization activity; vascular endothelial growth factor receptor binding	angiogenesis; anti-apoptosis; cell migration; cell proliferation; epithelial cell differentiation; eye photoreceptor cell development; induction of positive chemotaxis; lung development; mesoderm development; multicellular organismal development; nervous system development; nervous system development; positive regulation of epithelial cell proliferation; positive regulation of vascular endothelial growth factor receptor signaling pathway; regulation of progression through cell cycle; response to hypoxia; signal transduction; vasculogenesis	VEGF Pathway; Inhibition of Angiogenesis by TSP1; eNOS Signaling; Relaxin Pathway; Phospholipase-C Pathway; CRHR Pathway; mTOR Pathway; Paxillin Interactions; PAK Pathway; Ras Pathway; Cellular Apoptosis Pathway; Rap1 Pathway; GPCR Pathway; TGF-Beta Pathway; MAPK Family Pathway; P2Y Receptor Signaling; RhoGDI Pathway ; NF-KappaB Family Pathway; FGF Pathway; HIF1Alpha Pathway; Rac1 Pathway; JAK/STAT Pathway; Renin-Angiotensin Pathway; Mitochondrial Apoptosis; NF-KappaB (p50/p65) Pathway; Telomerase Components in Cell Signaling; Rho Family GTPases	[670-675]
BSG	basigin	19p13.3			Tumor marker, angiogenesis, immunoangiostasis	mannose binding; signal transducer activity; sugar binding	cell surface receptor linked signal transduction		[676-680]
CXCL3	chemokine (C-X-C motif) ligand 3	4q21	Cytokine	Intercrine Alpha (Chemokine CXC) Family	Oncogene, immune tolerance gene, angiogenesis, organ-specific metastases	chemokine activity	G-protein coupled receptor protein signaling pathway; chemotaxis; immune response; inflammatory response	Rho Family GTPases	[681-684]
CHRNA2	cholinergic receptor, nicotinic, alpha 2 (neuronal)	8p21	Receptor, Transporter, Neurotransmitter	Ligand-Gated Ionic Channel (TC 1.A.9) Family; autocrine growth factors	therapeutic target for lung cancer therapy	acetylcholine receptor activity; extracellular ligand-gated ion channel activity; ion channel activity; nicotinic acetylcholine-activated cation-selective	ion transport; signal transduction; synaptic transmission		[675,685,686]

						channel activity			
FUT3	fucosyltransferase 3	19p13.3				transferase activity, transferring glycosyl groups	carbohydrate metabolic process; protein amino acid glycosylation		[687-690]
FXYD3	FXYD domain containing ion transport regulator 3	19q13.11-q13.12	ion channel activity, chloride channel activity			chloride channel activity; chloride ion binding; ion channel activity	chloride transport; ion transport		[691,692]
PLD1	phospholipase D1	3q26	Signal Transduction	PLD Family		hydrolase activity; phosphoinositide binding; phospholipase D activity; protein binding	Ras protein signal transduction; cell communication; chemotaxis; lipid catabolic process; metabolic process; phospholipid metabolic process	Ras pathway; Rho Family GTPases; RhoA Pathway ;Rac1 Pathway; Endothelin-1 Signaling Pathway	[693,694]
POLD3	polymerase (DNA-directed), delta 3, accessory subunit	11q14				DNA binding; delta DNA polymerase activity; transferase activity	DNA synthesis during DNA repair; mismatch repair	DNA polymerase; Purine metabolism; Pyrimidine metabolism; Cell Cycle (Mitotic); DNA Repair; DNA Replication; Maintenance of Telomeres	[695]
PRKACB	protein kinase, cAMP-dependent, catalytic, beta	1p36.1	Kinase	Ser/Thr Family of Protein Kinases (cAMP Subfamily)		ATP binding; cAMP-dependent protein kinase activity; magnesium ion binding; nucleotide binding; protein kinase activity; protein serine/threonine kinase activity; transferase activity	G-protein signaling, coupled to cAMP nucleotide second messenger; protein amino acid phosphorylation; signal transduction	Apoptosis; Calcium signaling pathway; Gap junction; GnRH signaling pathway; Hedgehog signaling pathway; Insulin signaling pathway; Long-term potentiation; MAPK signaling pathway; Olfactory transduction; Taste transduction; Wnt signaling pathway; PKA pathway(17333334)	[696]
CXCR7	chemokine (C-X-C motif) receptor 7	2q37.3			Immune tolerance gene, therapeutic target for lung cancer therapy, organ-specific metastases	receptor activity; rhodopsin-like receptor activity	G-protein coupled receptor protein signaling pathway; biological_process; signal transduction		[675,681,684, 697,698]

REG1A	regenerating islet-derived 1 alpha	2p12				sugar binding	positive regulation of cell proliferation		[699,700]
RPS3	ribosomal protein S3	11q13.3-q13.5	Structural Protein	S3P Family of Ribosomal Proteins.	involved in DNA repair pathway and apoptosis pathway, interacted with metastasis suppressor nm23	RNA binding; structural constituent of ribosome	translation	DNA repair pathway and apoptosis pathway	[701,702]
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	7q21.3-q22	Metabolic	SERPIN Family	angiogenesis				[703-706]
SLC2A1	solute carrier family 2 (facilitated glucose transporter)	1p35-p31.3	Transport	Sugar Transporter (Subfamily-Glucose Transporter)	providing energy to rapidly dividing tumor cells,	glucose transporter activity; protein binding; sugar porter activity; transporter activity	carbohydrate transport; glucose transport		[707]
SPRR1B	small proline-rich protein 1B	1q21-q22	structural molecule activity			protein binding, bridging; structural molecule activity	epidermis development; keratinization; keratinocyte differentiation; peptide cross-linking		[708,709]
TUBA4A	tubulin, alpha 4a	2q35	Structural	Tubulin Superfamily	angiogenesis		GTP binding; GTPase activity; nucleotide binding; protein binding; structural molecule activity		[710-712]
VDR	vitamin D (1,25-dihydroxyvitamin D3) receptor	12q13.11	Transcription Factor	Nuclear Hormone Receptor Family (NR1 Subfamily)	Research tumor target	metal ion binding; protein binding; sequence-specific DNA binding; steroid hormone receptor activity; transcription factor activity; vitamin D3 receptor activity;	calcium ion homeostasis; calcium ion transport; intestinal absorption; multicellular organismal development; negative regulation of transcription; organ morphogenesis; regulation of transcription, DNA-dependent; signal transduction; skeletal development; transcription	MAPK	[713]

						zinc ion binding			
ADFP	Adipose differentiation-related protein	9p22.1							
ANXA8	annexin A8	10q11.2				calcium ion binding; calcium-dependent phospholipid binding	blood coagulation		
GALNT4	GalNAc transferase 4	12q21.3-q22		GalNAc-transferase family		calcium ion binding; manganese ion binding; sugar binding; transferase activity, transferring glycosyl groups	carbohydrate metabolic process		
LDHB	lactate dehydrogenase B	12p12.2-p12.1	Metabolic	Ldh Family		L-lactate dehydrogenase activity; oxidoreductase activity	anaerobic glycolysis; tricarboxylic acid cycle intermediate metabolic process		

Table 6-4 Average survivability prediction accuracy of 500 SVM class-differentiation systems on the optimal SVM parameters for lung adenocarcinoma prediction. The sigma is the optimal SVM parameter which gives the best average class-differentiation accuracy over the 500 testing-sets. The accuracies are obtained from 500 testing-sets.

Signature (method)	Optimal SVM parameter	Number of selected survival genes in signature	Non-survivable patients			Survivable patients			Q
			TP	FN	SE	TN	FP	SP	
1	45	51	5275	814	86.6%	14908	3	100%	96.1%
2	44	54	5175	939	84.6%	14886	0	100.%	95.5%
3	43	42	5173	909	85.1%	14918	0	100%	95.7%
4	41	34	5347	802	87.0%	14845	6	100%	96.2%
5	43	46	5340	703	88.4%	14956	1	100%	96.6%
6	45	54	5230	865	85.8%	14905	0	100%	95.9%
7	45	57	5139	972	84.1%	14889	0	100%	95.4%
8	46	50	5201	949	84.6%	14850	0	100%	95.5%
9	43	53	5313	801	86.9%	14886	0	100%	96.2%
10	46	47	5333	757	87.6%	14910	0	100%	96.4%

6.3.3 Disease relevant of identified markers

The 21 predictor-genes shared by all 10 sampling-sets included 16 cancer-related genes (4 anticancer targets, 1 oncogene, 2 angiogenesis genes, 1 organ-specific metastases gene, 1 tumor stability gene, 2 apoptosis genes, 1 cancer growth gene, and 4 cancer-pathway-affiliated genes). In our analysis, anticancer targets were obtained from the latest version of therapeutic target database (http://bidd.nus.edu.sg/group/cjttd/ttd.asp) [337,714], and the cancer-related genes and cancer-pathways were taken from recent publications [715-720].

Table 6-5 List of predictor-genes of lung adnocacinoma cancer dataset shared by all 10 sampling-sets.

Gene Group	Predictor-genes selected by this work only
Therapeutic target	VEGFA, CHRNA2, TUBA4A, VDR
Oncogene	CXCL3
Organ-specific metastases	FUT3
Cellular growth of carcinomas	FXYD3
Apoptosis	RPS3, REG1A
Stability genes	POLD3
Cancer pathway affiliated gene	PRKACB, PLD1, SERPINE1, SLC2A1

6.3.4 The predictive ability of identified markers

The predictive capabilities of our selected and 10 previously-derived signatures were evaluated using the SVM and PNN classification system on 500 randomly-generated training-testing sets that were generated from the Bhattacharjee's dataset [663] and the Beer's dataset [662]. For each training-test set, the training data was used to construct a classifier model, whereas the test data was used to evaluate the performance of the model. The predictive performance of selected signatures was evaluated by the overall accuracies (Q) of the 500 models.

Table 6-6 gives the prediction accuracies from the SVM prediction system constructed by independent samples from Bhattacharjee's dataset [663] using our selected gene signatures and 9 other previous-derived signatures. The accuracies for non-survival patients, survival patients and all patients of the our selected 10 gene signatures over the 500 test sets were in the range of 77.8% to 81.2%, 74.3% to 80.1% and 77.6% to 80.2% correspondingly,

while the standard deviations of the accuracy of all patients were ranged from 4.7% to 4.9%. The accuracies for non-survival patients, survival patients and overall accuracies for all patients of the 21 survival genes shared by all of the 10 signatures over the 500 test sets were 78.9%, 76.8% and 77.9% respectively, while the standard deviation of the accuracy of all patients was 4.8%. In contrast, the accuracies for non-survivable patients, survival patients and all patients of the 9 previous-derived signatures were in the range of 70.1%~77.2%, 56.4% to 78.0% and 66.0% to 77.6% respectively, and the standard deviations of the accuracy of all patients were in the range of 5.5%~6.9%. These results suggest that the performance of our signatures is more stable than those of other signatures by using independent database and by applying the SVM models.

Table 6-7 illustrates the predictive performance of the 500 testing sets by using the PNN classification system and the 500 randomly generated training-testing dataset generated from the independent Bhattacharjee's dataset [663] using our selected genes. The accuracies for non-survivable patients, survival patients and all patients of our selected 10 signatures over the 500 test sets were, respectively, in the range of 69.3% to 80.2%, 64.5% to 78.0% and 69.1% to 76.6%, while the standard deviations of the accuracy of all patients were in the range of 4.2% to 4.9%. The accuracies for non-survivable patients, survivable patients and all patients of the 21 survival genes shared by all the 10 signatures over the 500 test sets were 75.2%, 62.6% and 69.2% respectively, while the standard deviation of the overall accuracy was 4.4%. The accuracies for non-survivable patients, survival patients and patients of the 9 previous-derived signatures were in the range of 53.5%~95.6%, 45.7% to 76.8% and 64.0% to 71.7% respectively, while the standard deviations of accuracy of all patients were in the range of 4.7%~6.1%. These results recommend that the survival genes we selected have a slightly better consistency and more

stable predictive performance than those of the signatures selected by other studies with the PNN classification system.

Table 6-6 Average survivability prediction accuracy of the 500 SVM class-differentiation systems constructed by 84 samples from independent Bhattacharjee's lung adenocarcinoma dataset [663] using each of the signatures derived from this study and 9 previous studies. The accuracies were obtained from the 500 testing-sets.

Signature (method)	Number of selected survival genes in signature	Non-survivable patients			Survivable patients			Q	STDEV
		TP	FN	SE	TN	FP	SP		
1*	51	8495	2369	78.2%	7864	2272	77.6%	77.9%	4.8%
2*	54	8602	2262	79.2%	7783	2353	76.8%	78.0%	4.7%
3*	42	8745	2119	80.5%	8014	2122	79.1%	79.8%	4.8%
4*	34	8452	2412	77.8%	7837	2299	77.3%	77.6%	4.8%
5*	46	8723	2141	80.3%	8117	2019	80.1%	80.2%	4.9%
6*	54	8600	2264	79.2%	7731	2405	76.3%	77.8%	4.7%
7*	57	8802	2062	81.0%	7807	2329	77.0%	79.1%	4.8%
8*	50	8414	2450	77.4%	7533	2603	74.3%	75.9%	4.8%
9*	53	8655	2209	79.7%	7992	2144	78.8%	79.3%	4.7%
10*	47	8823	2041	81.2%	7899	2237	77.9%	79.6%	4.8%
Genes selected by all sampling sets*	21	8571	2293	78.9%	7788	2348	76.8%	77.9%	4.8%
Beer et al [662]	100	8287	2577	76.3%	7540	2596	74.4%	75.4%	6.2%
Beer et al [662]	50	7616	3248	70.1%	7407	2729	73.1%	71.5%	6.3%
Chen et al [666]	16	7755	3109	71.4%	7255	2881	71.6%	71.5%	6.6%
Chen et al [666]	5	7684	3180	70.7%	6820	3316	67.3%	69.1%	6.4%
Guo et al [661]	37	8088	2776	74.4%	7443	2693	73.4%	74.0%	6.4%
Guo et al [661]	8	8386	2478	77.2%	7904	2232	78.0%	77.6%	6.6%
Lu et al [665]	125	8348	2516	76.8%	7588	2548	74.9%	75.9%	5.8%
Lu et al [665]	64	8237	2627	75.8%	7612	2524	75.1%	75.5%	5.5%
Xu et al [667]	5	8141	2723	74.9%	5720	4416	56.4%	66.0%	6.9%

* Data from this study

Table 6-7 Average survivability prediction accuracies of the 500 PNN class-differentiation systems constructed by 84 samples from independent Bhattacharjee's lung adenocarcinoma dataset [663] using each of the signatures derived from this study and 9 previous works.

Signature (method)	Number of selected survival genes in signature	Non-survivable patients			Survivable patients			Q	STDEV
		TP	FN	SP	TN	FP	SE		
1*	51	7769	3156	71.1%	7270	2805	72.2%	71.6%	4.5%
2*	54	7837	3088	71.7%	7478	2597	74.2%	72.9%	4.9%
3*	42	8762	2163	80.2%	7333	2742	72.8%	76.6%	4.6%
4*	34	8656	2269	79.2%	6810	3265	67.6%	73.6%	4.3%
5*	46	7995	2930	73.2%	7863	2212	78.0%	75.5%	4.6%
6*	54	8019	2906	73.4%	6502	3573	64.5%	69.1%	4.5%
7*	57	8177	2748	74.8%	7518	2557	74.6%	74.7%	4.4%
8*	50	8000	2925	73.2%	7514	2561	74.6%	73.9%	4.2%
9*	53	7575	3350	69.3%	7140	2935	70.9%	70.1%	4.6%
10*	47	8379	2546	76.7%	7413	2662	73.6%	75.2%	4.7%
Genes selected by all sampling sets*	21	8217	2708	75.2%	6305	3770	62.6%	69.2%	4.4%
Beer et al [662]	100	7537	3388	69.0%	7515	2560	74.6%	71.7%	5.5%
Chen et al [666]	5	10446	479	95.6%	4600	5475	45.7%	71.6%	4.7%
Guo et al [661]	8	7752	3173	71.0%	7189	2886	71.4%	71.1%	5.2%
Guo et al [661]	37	7537	3388	69.0%	7284	2791	72.3%	70.6%	5.5%
Xu et al [667]	5	7884	3041	72.2%	6844	3231	67.9%	70.1%	5.6%
Beer et al [662]	50	9220	1705	84.4%	5310	4765	52.7%	69.2%	4.9%
Chen et al [666]	16	6780	4145	62.1%	7734	2341	76.8%	69.1%	5.7%
Lu et al [665]	125	6874	4051	62.9%	7591	2484	75.3%	68.9%	6.1%
Lu et al [665]	64	5845	5080	53.5%	7591	2484	75.3%	64.0%	6.1%

* Data from this study

The predictive accuracies of the 500 SVM survivability prediction systems from the original Beer' dataset [662] are shown in Table 6-8. These 500 training sets and 500 test sets were different from those used for survivability gene signatures selection. The accuracies for non-survivable patients, survival patients and all patients of the 10 survival gene signatures over the 500 test sets were in the range of 94.2% to 96.1%, 99.8 to 100% and 98.3% to 98.9% respectively, and the standard deviations of accuracy of all patients

were in the range of 3.2~3.7%. The accuracies for non-survival patients, survival patients and all patients of the 21 survival genes shared by all the 10 signatures over the 500 test sets were 90.5%, 99.5% and 96.9% respectively, and the standard deviation of the accuracy of all patients was 4.0%. The performances of our selected genes were both higher and more stable than those of the other 9 studies, in which the accuracies for non-survivable patients, survival patients and all patients were in the range of 52.5% to 66.6%, 81.8% to 96.8% and 75.6% to 88.3% respectively, and the standard deviations of accuracy of all patients were in the range of 5.8% to 8.0%. Overall, the survival predictor-genes selected from this study showed a better performance and relative low standard deviations than those previously identified ones when evaluated by the similar SVM classification system.

The predictive accuracies of the 500 PNN classification systems for survivability prediction from the original Beer' dataset [662] are shown in Table 6-9. The accuracies for non-survivable patients, survival patients and all patients of the 10 survival gene signatures over the 500 test sets were in the range of 79.6% to 89.8%, 95.9% to 98.9% and 93.4% to 95.5% respectively, and the standard deviations (STDEV) were in the range of 4.3% to 5.2%. The accuracies for non-survivable patients, survival patients and all patients of the 21 survival genes shared by all the 10 signatures over the 500 test sets were 75.1%, 96.2% and 90.2% respectively, and the standard deviations of the overall accuracy was 5.7%. In contrast, the accuracies for non-survivable patients, survival patients and all patients of the 9 gene signatures from other studies over the 500 test sets were in the range of 57.2% to 76.1%, 73.5% to 89.7% and 72.1% to 80.6% respectively, and the standard deviation were in the range of 7.5% to 11.0%. This comparison indicated that the

performance of our selected gene signatures is better and stabler than those of other studies using the PNN classification methods for survivability prediction.

Table 6-8 Average survivability prediction accuracy of 500 SVM class-differentiation systems constructed by 86 samples from Beer's lung adenocarcinoma dataset [662].

Signature (method)	Number of selected survival genes in signature	Non-survival patients			Survivable patients			Q	STDEV
		TP	FN	SE	TN	FP	SP		
1*	51	5589	342	94.2%	15047	22	99.9%	98.3%	3.4%
2*	54	5671	260	95.6%	15043	26	99.8%	98.6%	3.2%
3*	42	5622	309	94.8%	15061	8	99.9%	98.5%	3.5%
4*	34	5630	301	94.9%	15037	32	99.8%	98.4%	3.3%
5*	46	5679	252	95.8%	15039	30	99.8%	98.7%	3.5%
6*	54	5664	267	95.5%	15054	15	99.9%	98.7%	3.7%
7*	57	5678	253	95.7%	15059	10	99.9%	98.7%	3.4%
8*	50	5694	237	96.0%	15069	0	100%	98.9%	3.3%
9*	53	5702	229	96.1%	15047	22	99.9%	98.8%	3.3%
10*	47	5686	245	95.9%	15052	17	99.9%	98.8%	3.3%
Genes selected by all sampling sets *	21	5369	562	90.5%	14987	82	99.5%	96.9%	4.0%
Beer et al [662]	100	3951	1980	66.6%	14589	480	96.8%	88.3%	5.8%
Beer et al [662]	50	3302	2629	55.7%	14134	935	93.8%	83.0%	6.7%
Lu et al [665]	64	3526	2405	59.5%	13658	1411	90.6%	81.8%	6.4%
Lu et al [665]	125	3467	2464	58.5%	13570	1499	90.0%	81.1%	6.2%
Guo et al [661]	37	2760	3171	46.5%	13974	1095	92.7%	79.7%	7.0%
Chen et al [666]	16	2925	3006	49.3%	13702	1367	90.9%	79.2%	7.0%
Xu et al [667]	5	3696	2235	62.3%	12432	2637	82.5%	76.8%	7.5%
Chen et al [666]	5	3577	2354	60.3%	12325	2744	81.8%	75.7%	8.0%
Guo et al [661]	8	3113	2818	52.5%	12760	2309	84.6%	75.6%	7.3%

* Data from this study

Table 6-9 Average survivability prediction accuracies of the 500 PNN class-differentiation systems constructed by 86 samples from Beer's lung adenocarcinoma dataset [662].

Signature (Method)	No. of selected predictor genes in signature	Non-survivable patients			Survivable patients			Q	STDEV
		TP	FN	SE	TN	FP	QN		
1*	51	5069	862	85.5%	14635	434	97.1%	93.8%	4.8%
2*	54	5062	869	85.3%	14726	343	97.7%	94.2%	4.6%
3*	42	4939	992	83.3%	14715	354	97.7%	93.6%	4.7%
4*	34	4719	1212	79.6%	14904	165	98.9%	93.4%	5.2%
5*	46	5210	721	87.8%	14798	271	98.2%	95.3%	4.5%
6*	54	5326	605	89.8%	14730	339	97.8%	95.5%	4.3%
7*	57	5214	717	87.9%	14533	536	96.4%	94.0%	4.9%
8*	50	5089	842	85.8%	14707	362	97.6%	94.3%	4.5%
9*	53	5319	612	89.7%	14450	619	95.9%	94.1%	4.4%
10*	47	5100	831	86.0%	14571	498	96.7%	93.7%	4.8%
Genes selected by all sampling sets*	21	4454	1477	75.1%	14495	574	96.2%	90.2%	5.7%
Beer et al [662]	50	3393	2538	57.2%	13523	1546	89.7%	80.6%	7.5%
Beer et al [662]	100	4183	1748	70.5%	12648	2421	83.9%	80.1%	9.0%
Lu et al [665]	64	4515	1416	76.1%	11700	3369	77.6%	77.2%	10.0%
Xu et al [667]	5	4205	1726	70.9%	11960	3109	79.4%	77.0%	7.5%
Chen et al [666]	5	3601	2330	60.7%	11985	3084	79.5%	74.2%	7.9%
Guo et al [661]	8	3743	2188	63.1%	11768	3301	78.1%	73.9%	8.2%
Chen et al [666]	16	3569	2362	60.2%	11936	3133	79.2%	73.8%	7.8%
Lu et al [665]	125	4310	1621	72.7%	11078	3991	73.5%	73.3%	12.5%
Guo et al [661]	37	3903	2028	65.8%	11232	3837	74.5%	72.1%	11.0%
* Data from this study									

6.3.5 Patient survival analysis using survival markers

Hierarchical cluster analysis can cluster the samples according to their expression profiles across the gene we selected. The comparison of the survival curves from these clusters can be used to assess the statistical significance of the survivability difference among the clusters.

By using 21 identified markers, hierarchical cluster analysis grouped 86 lung adenocarcinoma patients in the Beer's dataset [662] into three clusters (Figure 6-2). Kaplan-Meier survival analysis showed that the survival time after therapy was significantly different in the three patient clusters ($P<0.0001$, log-rank test, Figure 6-3). Cluster 1 was the poor prognosis group. The average survival time of patients in this cluster was 50.6 months. In this cluster, the numbers of survivable patients (SP) and non-survivable patients (NSP) were 12 and 14 respectively (Table 6-10). The survival percentage, which defined by SP/(SP+NSP), were 46%. Cluster 2 was the good prognosis groups with average survival time of 82.2 months. The SP, NSP and survival percentage were 26, 1 and 96% respectively. Cluster 3 was the moderate prognosis group with average survival time of 74.8 months. The SP, NSP and survival percentage were 22, 9 and 72% respectively. By using the similar way, Guo et al [661] clustered these samples [662] into three clusters by using 37 genes and the survival percentages were 69%, 72% and 75% for poor, moderate and good prognosis clusters, respectively (Table 6-10). The survival percentage for three clusters generated by 100 genes in Beer et al [662] are 43%, 57% and 88% for poor, moderate and good prognosis clusters, respectively (Table 6-10). These results indicated that the 21 genes selected by using our method can be classified into better clinically meaningful groups for further prognosis than the genes selected by other group.

Hierarchical clustering of the 21 genes on the independent validation dataset - Bhattacharjee's dataset [663] showed the similar results (Figure 6-4). Three clusters had significant difference by using Kaplan-Meier analysis with $P<0.001$ from log-rank test (Figure 6-5). The average survival time for cluster 1, which was poor prognosis group, was 35.7 months. The average survival time for cluster 2, which was moderate prognosis

group, was 32.0 months. The average survival time for cluster 3, which was good prognosis group, was 78.3 months. The survival percentages of the three clusters were 30%, 43 % and 73% for poor, moderate and good prognosis clusters, as shown in Table 6-11. By using the similar strategies, Guo et al [661] clustered the sample into three clusters. However, the survivability percentages among the clusters were 45%, 46% and 51% for three clusters by using the Kaplan-Meier analysis, showing little statistically different among the clusters (Table 6-10). The survival percentage of three clusters formed by 21 genes we selected were more spread out than those formed by the genes selected by other researchers, further suggesting that 21 genes we selected have robust behavior for prognosis prediction.

Table 6-10 Comparison of the survival rate in clusters with other groups, by using different signatures and Beer's microarray dataset [662].

Study	Gene number in signatures	Poor prognosis cluster			Moderate prognosis cluster			Good prognosis cluster		
		SP[1]	NSP[2]	Survival rate[3]	SP	NSP	Survival rate	SP	NSP	Survival rate
This study	21	12	14	46%	22	9	72%	26	1	96%
Guo's group [661]	37	25	11	69%	15	6	71%	20	7	74%
Beer's group [662] [4]	100	25	19	43%	23	19	57%	37	5	88%

[1] SP: the number of survivable patients
[2] NSP: the number of non-survivable patients
[3] Survival rate= SP/(SP+NSP)
[4] The cluster analysis was done on 128 lung cancer samples

Figure 6-2 Hierarchical clustering analysis of the 21 lung adenocarcinoma survival markers from Beer's microarray dataset [662]. The tumor samples were aggregated into three clusters. Substantially elevated (red) and decreased (green) expression of the genes is observed in individual tumors.

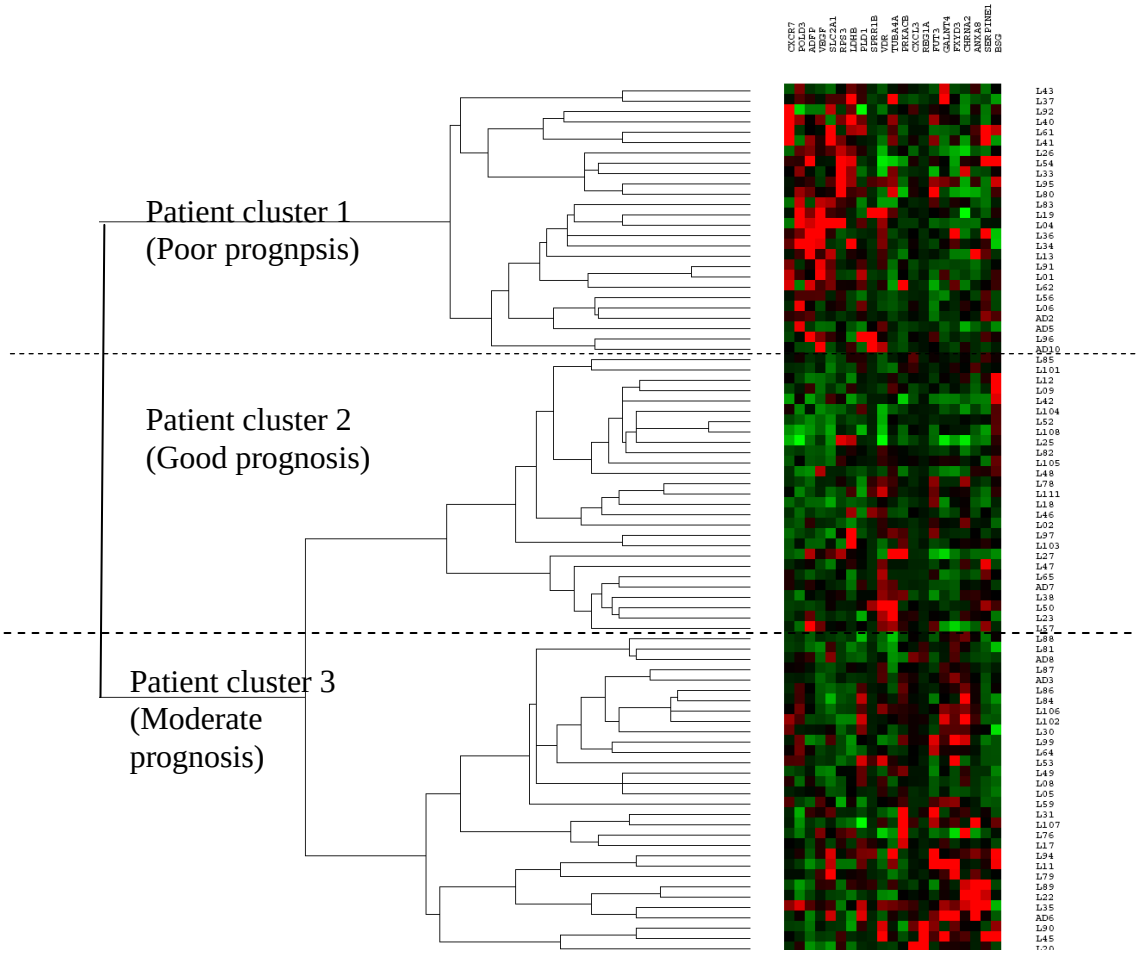
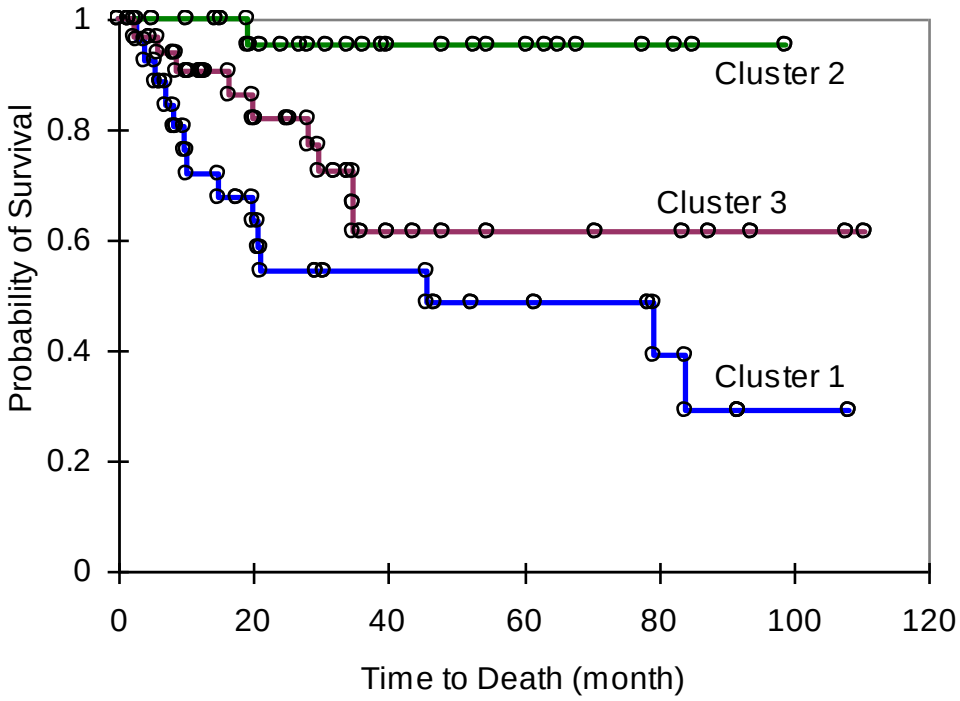


Figure 6-3 Kaplan-Meier survival analysis of the three clusters of patients from Figure 6-2. Average survival time of patients in cluster 1 is 50.6 months; average survival time of patients in cluster 2 is 82.2 months; average survival time of patients in cluster 3 is 74.8 months (P<0.0001, log-rank test).



Study	Gene number in signatures	Poor prognosis cluster			Moderate prognosis cluster			Good prognosis cluster		
		SP[1]	NSP[2]	Survival rate[3]	SP	NSP	Survival rate	SP	NSP	Survival rate
This study	21	12	14	46%	22	9	72%	26	1	96%
Guo's group [661]	37	25	11	69%	15	6	71%	20	7	74%
Beer's group [662] [4]	100	25	19	43%	23	19	57%	37	5	88%

Figure 6-4 Hierarchical clustering analysis of the 21 lung adenocarcinoma markers from Bhattacharjee's microarray dataset [663]. The tumor samples were aggregated into three clusters. This 21-gene signature are shared by 10 survival genes sets of lung adenocarcinoma derived by using datasets from Beer et al [662] and by using multiple random sampling method.

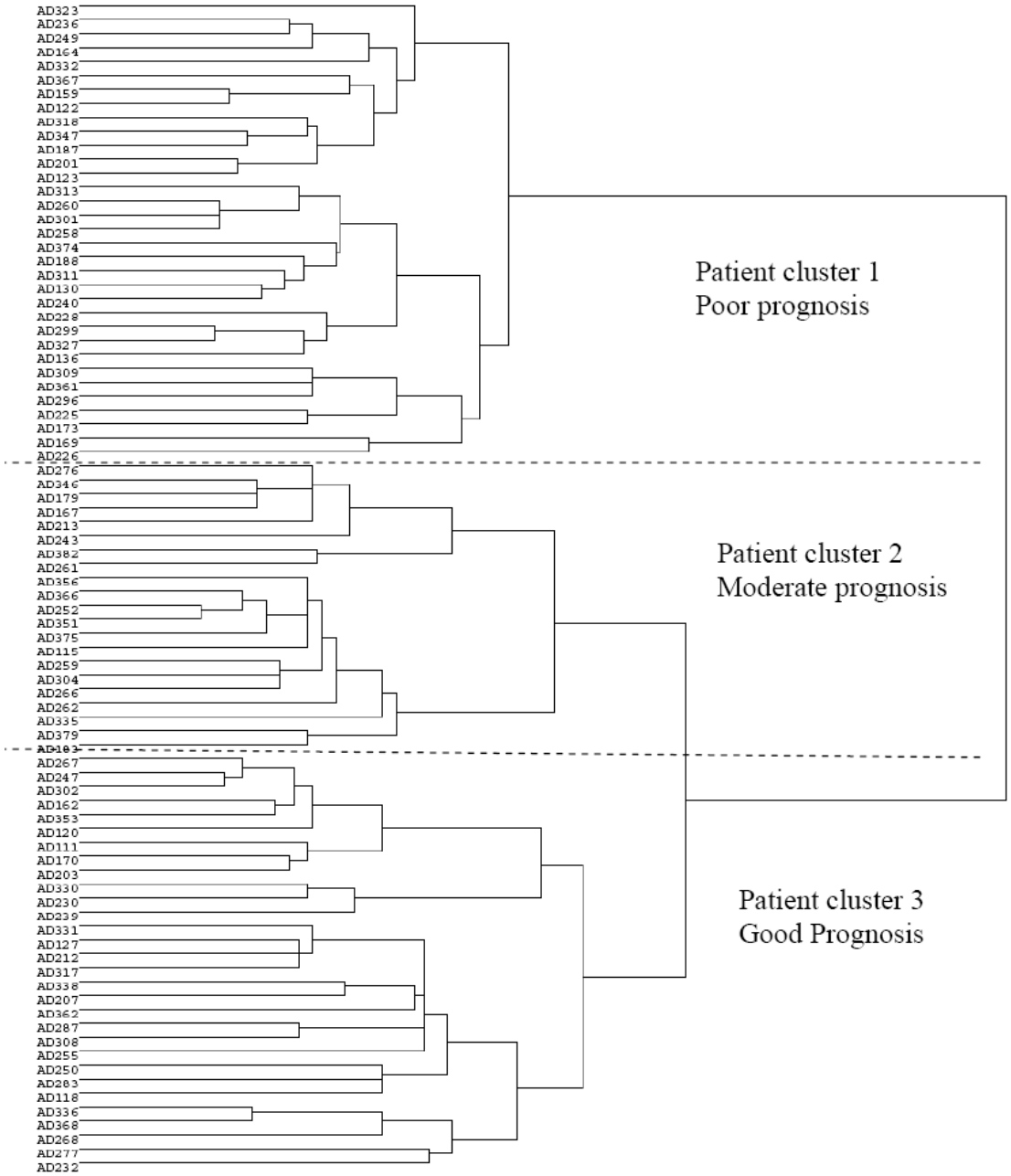
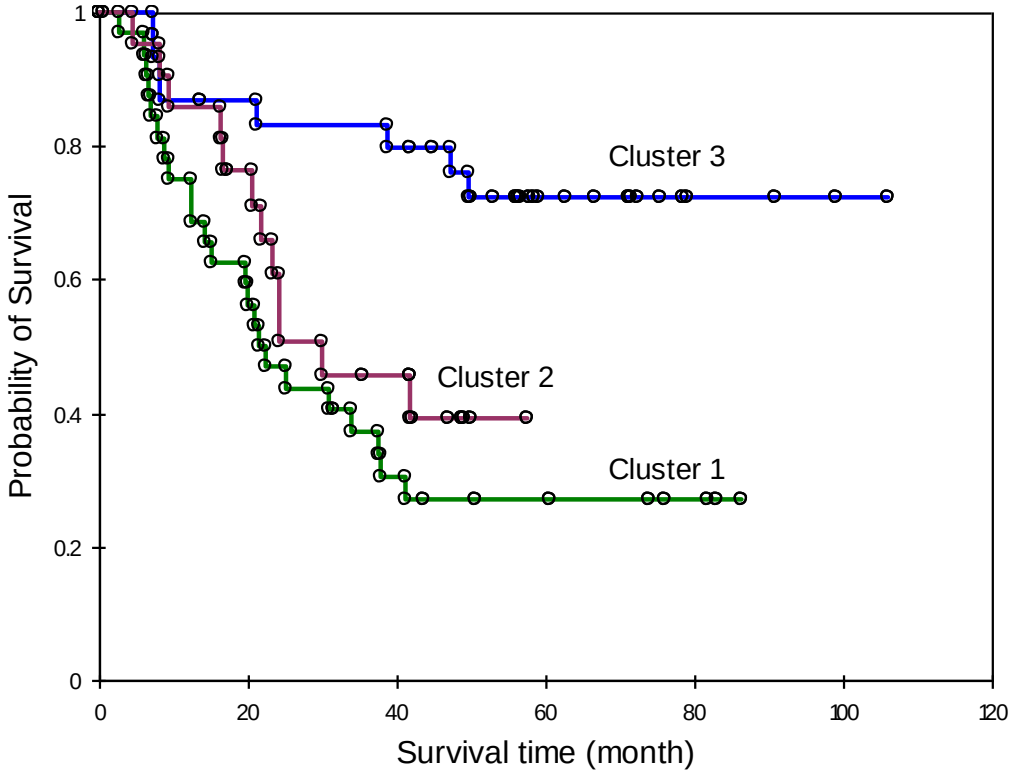


Figure 6-5 Kaplan-Meier survival analysis of the three clusters of patients from Figure 6-4. Average survival time of patients in cluster 1 is 35.7 months; average survival time of patients in cluster 2 is 32.0 months; average survival time of patients in cluster 3 is 78.3 months (P<0.001, log-rank test).



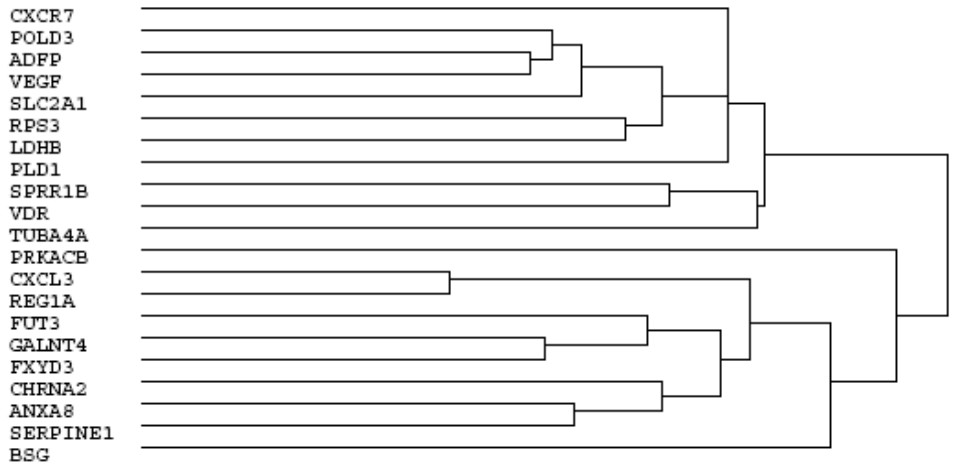
Study	Gene number in signatures	Poor prognosis cluster			Moderate prognosis cluster			Good prognosis cluster		
		SP	NSP	Survival rate	SP	NSP	Survival rate	SP	NSP	Survival rate
This study	21	10	23	30%	9	12	43%	22	8	73%
Guo's group [661]	37	9	11	45%	11	13	46%	20	19	51%

6.3.6 Hierarchical clustering analysis of the survival markers

In the hierarchical analysis for 86 lung adenocarcinoma patients in the Beer's dataset [662], 21 survival genes were formed into two clusters (Figure 6-6). Genes in gene cluster 1 are correlated with a poor prognosis of lung adenocarcinoma, whereas genes in cluster 2 are correlated with a moderate prognosis of lung adenocarcinoma. Figure 6-6 shows that

higher expression of the genes in cluster 1 is associated with poor prognosis in patients in lung adenocarcinoma, and higher expression of the genes in cluster 2 is associated with moderate prognosis in patients in lung adenocarcinoma. On the other hand, the lower expression of all these 21 genes in both cluster 1 and 2 is associated with good prognosis. The 11 poor-prognosis genes in cluster 1 are CXCR7, POLD3, ADFP, VEGF, SLC2A1, RPS3, LDHB, PLD1, SPRR1B, VDR, and TUBA4A, of which four genes, VEGF, CXCR7, TUBA4A and VDR, are therapeutic tumor targets. The 10 moderate-prognosis genes in cluster 2 consist of PRKACB, CXCL3, REG1A, FUT3, GALNT4, FXYD3, CHRNA2, ANXA8, SERPINE1 and BSG. CHRNA2 is a molecular target for lung cancer therapy. The target information was obtained from the latest version of therapeutic target database [337,714],

Figure 6-6 Hierarchical clustering analysis of the 21 lung adenocarcinoma survival markers from Beer's microarray dataset [662]



Most of the selected genes were experimentally proved that high expression of these genes was related to adverse survivability of patients. High level of serum VEGF (vascular endothelial growth factor) in the NSCLC may directly predict worse survival [721,722], and acts as a crucial parameter in lung cancer, especially associated with NSCLC

[721,722]. The expression of VDR (Vitamin D receptor) was observed in lung adenocarcinoma [723]. Increased SLC2A1 (solute carrier family 2) expression in tumors was identified as an adverse prognostic factor and a predictive prognosis marker [724]. Elevated PLD1 (phospholipase D1) activity could promote tumor progress and show high invasive potential [694,725-727]. Up-regulated expression of FXYD3 (FXYD domain containing ion transport regulator 3) in cancer indicated that FXYD3 might contribute to the proliferative activity of malignancy [691]. In vivo experiments demonstrated that BSG (basigin; CD147) overexpression stimulated tumor angiogenesis and growth [678]. Higher expression of FUT3 (fucosyltransferase 3) was often observed in high grade and poor prognosis tumors [728]. The expression level of SERPINE1 (plasminogen activator inhibitor-1) in tissue was significantly and positively correlated with tumor severity and tumor size [729], and high level of SERPINE1 could indicate an aggressive phonotype of carcinomas [730,731], serving as an indicator of poor prognosis in adenocarcinomas of the lung [732]. REG1A (regenerating islet-derived 1 alpha) expression was reported to be closely related to the carcinoma invasiveness of neoplasm [733], and to be an independent predictor of overall cancer patient survival as well [734]. The over-expression of SPRR1B (small proline-rich protein 1B) for prolonged periods might disrupt normal progression of mitosis [708]. Therefore, the expression of most of our selected survival marker genes has been validated as either directly or closely related to cancer metastasis and prognosis in the literatures indicated in Table 6-3.

6.4 Summary

In this chapter, the comprehensive gene selection system was further evaluated on the selection of survival marker for lung adenocarcinoma. By way of multiple random sampling, 21 genes were selected by all of ten sets of lung adenocarcinoma survival

marker signatures, in which 34 to 53 genes were selected. These 21 markers were then used to develop PNN and SVM prediction models to predict prognosis for lung adenocarcinoma patients from different datasets. The survivability analysis by hierarchical clustering analysis and Kaplan-Meier survival analysis further suggested that the derived signatures from our system could provide better performance when comparing with other signatures. Most of the selected genes have been experimentally proved that high expression of the genes is relevant to adverse survivability of patients. 12 markers, including 5 known targets and 7 novel targets, were successfully predicted as therapeutic targets by using a therapeutic target prediction system.

7 Conclusion and Future Work

This last chapter summarizes the major findings and contributions (section 7.1) of this work to the progress of using integrated molecular profiling and machine learning approaches for therapeutic mechanism, response, target, and biomarker discovery. Limitations of the present work (section 7.2) and possible areas for future studies (section 7.3) are also discussed.

7.1 Major findings and contributioins

Drug discovery efforts are nowadays to search for therapeutic regimens that comprise more than one active ingredient and drugs that are composed of a single chemical entity but combat multiple targets. In chapters 3, a focus study in the early drug discovery process on identifying and optimizing the activity of combinations of molecules was carried out and the newly identified drug combination actions can result in the identification of more effective drug regimens in clinical setting. It is the first time that we comprehensivly observed the coordinated interactions and network regulations from a systems perspective to understand the mechanism and mode of actions of successful drug combinations. The identified modes of actions of drug combinations reveal seven important categories of multicomponents therapeutics of current successful drug-combinations and multi-targeting agents, which will be a starting point to guide a rational combination screening by using different mode of actions.

In chapters 4 and 5, the application of integrated molecular profiling, including mutational, amplification and microarray gene expression profiles, suggests a useful approach for efficiously exploring of drug efficacy issues and developing of novel therapeutic target.

First of all, the use of integrated molecular profilingprovide a reasonable explanation for the variations observed in clinical TKI drug responses and prediction as its application in the future. The identified co-altered genes may serve as potential targets for new drug development and choice of combination therapy. Secondly, the exploration of the potential epitopes help better understanding of the antigen recognition mechanism and dramatically reduce the workload for experimental identification of antigenic scources. The methodology developed in this work could be further extended to the studies of other fields of immunology in disease treatment..

Finally, a robust computational system for gene signature derivation from microarray data was developed. A popular and accurate machine learning method, support vector machines, was applied to classify the samples. Recursive feature selection incorporating with multiple random sampling method and gene consistency evaluation strategies was used in gene selection procedure. This system has been successfully applied to selection colon cancer markers and lung adenocarcinoma survivability markers. A total of 21 lung adenocarcinoma survivability biomarkers were identified and shared by all of the 10 sampling-sets. The results from the lung adenocarcinoma survival gene selection suggest that, our system is able to derive stable and good predictive marker signatures. The use of consensus scoring for multiple random sampling and evaluation of gene-ranking consistency seem to have impressive capability in avoiding erroneous elimination of predictor-genes due to such noise as measurement variability and biological differences. This approach can be further implemented in biomarker selection for other highthroughput biological data.

7.2 Limitations

Some obvious limitations of using microarray data are the main obstacles hindering the identification of the real TSAs and biomarkers in this study In tumor antigen discovery, six known melanoma TSAs were misidentified which mainly because of failing of expression analysis. Due to the high cost of microarray experiments, the sample size is much smaller than what is required for a satisfactory diagnosis and prognosis of a certain disease such as cancer. The currently available platforms for microarray data are different. Increasing sample size via synchronizing the different platforms remains a challenged task. Although we introduced a multiple random sampling strategies in cancer biomarker identification from microarray data, which has shown improved consistency and stability while feature selection from 5000 different microarray dataset combinations. The combination is far from the complex of heterogenic cancer patients [Ref]. Therefore, large size of representative samples may improve the accuracy of our system. Further improvement in experimental design, measurement quality, annotation accuracy and coverage, and signature-selection will enable the derivation of more accurate signatures for facilitating biomarker and target discovery.

In this study, the application integrated molecular profiling is currently limited to the RTKs, in which the signaling pathways are rather well established. Sufficient information could also be collected for their possible assistant genes. However, it may be much more difficult to do so for less characterized pathways involving novel genes. Moreover, downstream signaling genes and proteins themselves could be not only actively mutated but also highly activated due to the posttranslational modifications and translocation activiation. This way makes the analysis on the profiles of gene expression and genetic variation is extremely complicated and susceptible to errors. Therefore, more factors

should be considered in the cilincial application despite the importance of the drug effeicacy evaluation strategy. On the other heand, there is still lack of information about clinical trials for many drugs, such as those targeting IGFR-1 and MET. Once their results are released, the information could be added to further validate the usefulness of method.

7.3 Directions for future research

Both therapeutic selectivity and rational combination therapeutics are major challenge in drug discovery. This is especially true for the treatment of cancer, metabolic, or inflammation disorders, which must rely on targets that are present in both healthy and diseased tissues, and which are thwarted by the compensatory mechanisms available to complex biological systems. This work aims to translate the fundamental insight of disease causing-genes for systemic drug efficacy studies. Development strategies to improve selectivity of targeted and multicomponent therapy will be the focus of my research in the near future. To address these challenges, I will mainly focus on the development of efficient computational methods for agnostic screening of compound combination and multi-target agents from a library of chemical and biological agents that perturb a diverse set of molecular targets.

In practical study of compound combination, a comparatively small number of compounds will provide a very large number of combinations; a collection of 1,000 compounds yields more than 500,000 pairwise combinations, and many more higher-order combinations. Moreover, variations in molar ratio and timing of compound addition can be relevant and increase the size of the search space. Therefore, efficient methods, such as improving the strategy of dose-response matrix design, are needed to facilitate the screening or simulation the possible and rational combination pairs. Moreover, experimental strategies, including antibodies, negative dominant controls, antisense oligonucleotides, ribozymes,

small-interfering RNAs, and mouse reverse genetics such as knockout phenotypes, are need to be established for validation of developed combinations and various ratios of the component drugs.

Currently multi-target kinase inhibitors are among the most successful clinical anticancer drugs (e.g. sunitinib against PDGFR and VEGFR, dasatinib against Abl and Src, sorafenib against Braf and VEGFR, and lapatinib against EGFR and HER2) and have been actively pursued in current drug discovery efforts[735,736]. Methods for efficient search of multi-target agents are highly desired. It is known that virtual screening (VS) methods have been widely explored for facilitating lead discovery against individual targets[737-739]. In particular, molecular docking[740], pharmacophore[741], QSAR[742], machine learning[743], and combination methods[744] have been extensively used for VS of single-target kinase inhibitors, but few multi-target VS studies have been reported[745,746]. Thus, it is interesting to develop strategies by using VS method from the know multi-target kinase inhibitors and facilicate the highthrouput screening of novel agents targeting the larger kinase set[746].

.

BIBLIOGRAPHY

1 Sandberg, A.A. and Chen, Z. (1994) Cancer cytogenetics and molecular genetics: detection and therapeutic strategy. *In Vivo* 8 (5), 807-818

2 Chen, Z. and Sandberg, A.A. (2002) Molecular cytogenetic aspects of hematological malignancies: clinical implications. *Am J Med Genet* 115 (3), 130-141

3 Mrozek, K. et al. (2004) Cytogenetics in acute leukemia. *Blood Rev* 18 (2), 115-136

4 Golub, T.R. et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286 (5439), 531-537

5 Ramaswamy, S. et al. (2001) Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 98 (26), 15149-15154

6 Robinson, S. and Kessling, A. (1992) Diabetes secondary to genetic disorders. *Baillieres Clin Endocrinol Metab* 6 (4), 867-898

7 Clee, S.M. et al. (2006) Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat Genet* 38 (6), 688-693

8 Li, D. (2006) Positional cloning: single-gene cardiovascular disorders. *Methods Mol Med* 128, 125-136

9 Gulcher, J. and Stefansson, K. (2006) Positional cloning: complex cardiovascular traits. *Methods Mol Med* 128, 137-152

10 Hotta, K. (2005) [Genetic testing and gene-based testing for obesity]. *Nippon Rinsho* 63 Suppl 12, 280-284

11 Zhang, W. et al. (2006) A method for predicting disease subtypes in presence of misclassification among training samples using gene expression: application to human breast cancer. *Bioinformatics* 22 (3), 317-325

12 Sachidanandam, R. et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409 (6822), 928-933

13 Larsson, T.P. et al. (2005) Comparison of the current RefSeq, Ensembl and EST databases for counting genes and gene discovery. *FEBS Lett* 579 (3), 690-698

14 Knowles, J. and Gromo, G. (2003) A guide to drug discovery: Target selection in drug discovery. *Nat Rev Drug Discov* 2 (1), 63-69

15 Collins, I. and Workman, P. (2006) New approaches to molecular cancer therapeutics. *Nat Chem Biol* 2 (12), 689-700

16 Workman, P. (2005) Genomics and the second golden era of cancer drug development. *Mol Biosyst* 1 (1), 17-26

17 Shimoji, T. et al. (2006) [Gene expression profiling for prediction of response to chemotherapy]. *Gan To Kagaku Ryoho* 33 (1), 1-5

18 Workman, P. and Kaye, S.B. (2002) Translating basic cancer research into new cancer therapeutics. *Trends Mol Med* 8 (4 Suppl), S1-9

19 Sahin, O. and Wiemann, S. (2009) Functional genomics and proteomics approaches to study the ERBB network in cancer. *FEBS Lett* 583 (11), 1766-1771

20 Uberall, I. et al. (2008) The status and role of ErbB receptors in human cancer. *Exp Mol Pathol* 84 (2), 79-89

21 Miller, T.W. et al. (2009) Loss of Phosphatase and Tensin homologue deleted on chromosome 10 engages ErbB3 and insulin-like growth factor-I receptor signaling to promote antiestrogen resistance in breast cancer. *Cancer Res* 69 (10), 4192-4201

22 Normanno, N. et al. (2009) Target-based therapies in breast cancer: current status and future perspectives. *Endocr Relat Cancer* 16 (3), 675-702

23 Park, B.H. and Davidson, N.E. (2007) PI3 kinase activation and response to Trastuzumab Therapy: what's neu with herceptin resistance? *Cancer Cell* 12 (4), 297-299

24 Ocampo, M.T. et al. (2002) Targeted deletion of mNth1 reveals a novel DNA repair enzyme activity. *Mol Cell Biol* 22 (17), 6111-6121

25 Papp, B. (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429, 661-664

26 Smalley, K.S. et al. (2006) Multiple signaling pathways must be targeted to overcome drug resistance in cell lines derived from melanoma metastases. *Mol Cancer Ther* 5 (5), 1136-1144

27 Pilpel, Y. et al. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 29 (2), 153-159

28 Peng, X.H. et al. (2006) Cross-talk between epidermal growth factor receptor and hypoxia-inducible factor-1alpha signal pathways increases resistance to apoptosis by up-regulating survivin gene expression. *J Biol Chem* 281 (36), 25903-25914

29 Muller, R. (2004) Crosstalk of oncogenic and prostanoid signaling pathways. *J Cancer Res Clin Oncol* 130 (8), 429-444

30 Massarweh, S. and Schiff, R. (2006) Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. *Endocr Relat Cancer* 13 Suppl 1, S15-24

31 Sergina, N.V. et al. (2007) Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* 445 (7126), 437-441

32 Kassouf, W. et al. (2005) Uncoupling between epidermal growth factor receptor and downstream signals defines resistance to the antiproliferative effect of Gefitinib in bladder cancer cells. *Cancer Res* 65 (22), 10524-10535

33 Christopher M. et al. (2006) Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nature Reviews Cancer* 6, 227-239

34 Citri, A. and Yarden, Y. (2006) EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol* 7 (7), 505-516

35 Engelman, J.A. and Cantley, L.C. (2006) The role of the ErbB family members in non-small cell lung cancers sensitive to epidermal growth factor receptor kinase inhibitors. *Clin Cancer Res* 12 (14 Pt 2), 4372s-4376s

36 Reid, A. et al. (2007) Dual inhibition of ErbB1 (EGFR/HER1) and ErbB2 (HER2/neu). *Eur J Cancer* 43 (3), 481-489

37 Mellinghoff, I.K. et al. (2007) PTEN-mediated resistance to epidermal growth factor receptor kinase inhibitors. *Clin Cancer Res* 13 (2 Pt 1), 378-381

38 Perveen, R. et al. (2007) A heterozygous c-Maf transactivation domain mutation causes congenital cataract and enhances target gene activation. *Hum Mol Genet* 16 (9), 1030-1038

39 Engelman, J.A. et al. (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* 316 (5827), 1039-1043

40 Morgillo, F. et al. (2007) Implication of the insulin-like growth factor-IR pathway in the resistance of non-small cell lung cancer cells to treatment with gefitinib. *Clin Cancer Res* 13 (9), 2795-2803

41 Sawyers, C.L. (2007) Cancer: mixing cocktails. *Nature* 449 (7165), 993-996

42 Guix, M. et al. (2008) Acquired resistance to EGFR tyrosine kinase inhibitors in cancer cells is mediated by loss of IGF-binding proteins. *J Clin Invest* 118 (7),

2609-2619

43 Raponi, M. et al. (2008) KRAS mutations predict response to EGFR inhibitors. *Curr Opin Pharmacol* 8 (4), 413-418

44 Thomson, S. et al. (2008) Kinase switching in mesenchymal-like non-small cell lung cancer lines contributes to EGFR inhibitor resistance through pathway redundancy. *Clin Exp Metastasis* 25 (8), 843-854

45 Linardou, H. et al. (2008) Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *Lancet Oncol* 9 (10), 962-972

46 Agarwal, S. et al. (2009) Association of constitutively activated hepatocyte growth factor receptor (Met) with resistance to a dual EGFR/Her2 inhibitor in non-small-cell lung cancer cells. *Br J Cancer* 100 (6), 941-949

47 Sos, M.L. et al. (2009) PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR. *Cancer Res* 69 (8), 3256-3261

48 Blume-Jensen, P. and Hunter, T. (2001) Oncogenic kinase signalling. *Nature* 411 (6835), 355-365

49 Brent, R. (2000) Genomic biology. *Cell* 100 (1), 169-183

50 Kanehisa, M. et al. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res* 30 (1), 42-46

51 Kitano, H. (2002) Systems biology: a brief overview. *Science* 295 (5560), 1662-1664

52 Zhang, P. (1999) The cell cycle and development: redundant roles of cell cycle regulators. *Curr Opin Cell Biol* 11 (6), 655-662

53 Shaheen, R.M. et al. (2001) Tyrosine kinase inhibition of multiple angiogenic growth factor receptors improves survival in mice bearing colon cancer liver metastases by inhibition of endothelial cell survival mechanisms. *Cancer Res* 61 (4), 1464-1468

54 Zimmermann, G.R. et al. (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov Today* 12 (1-2), 34-42

55 Chou, T.C. and Talalay, P. (1984) Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv Enzyme Regul* 22, 27-55

56 Chou TC, T.P. (1983) Analysis of combined drug effects: a new look at a very old problem. *Trends Pharmacol Sci* 4, 450-454

57 Loewe, S. (1953) The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* 3 (6), 285-290

58 Borisy, A.A. et al. (2003) Systematic discovery of multicomponent therapeutics. *Proc Natl Acad Sci U S A* 100 (13), 7977-7982

59 Zon, L.I. and Peterson, R.T. (2005) In vivo drug discovery in the zebrafish. *Nat Rev Drug Discov* 4 (1), 35-44

60 Keith, C.T. et al. (2005) Multicomponent therapeutics for networked systems. *Nat Rev Drug Discov* 4 (1), 71-78

61 Csermely, P. et al. (2005) The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci* 26 (4), 178-182

62 Kitano, H. (2007) A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* 6 (3), 202-210

63 Kamb, A. et al. (2007) Why is cancer drug discovery so difficult? *Nat Rev Drug Discov* 6 (2), 115-120

64 http://www.centerwatch.com/patient/drugs/druglist.html.
65 Wishart, D.S. et al. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 34 (Database issue), D668-672
66 Saltz, L. et al. (2006) Panitumumab. *Nat Rev Drug Discov* 5 (12), 987-988
67 Yang, X.D. et al. (1999) Eradication of established tumors by a fully human monoclonal antibody to the epidermal growth factor receptor without concomitant chemotherapy. *Cancer Res* 59 (6), 1236-1243
68 Carter, P. et al. (1992) Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc Natl Acad Sci U S A* 89 (10), 4285-4289
69 Goldenberg, M.M. (1999) Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. *Clin Ther* 21 (2), 309-318
70 Rosenfeld, P.J. et al. (2006) Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med* 355 (14), 1419-1431
71 Carter, P.J. (2006) Potent antibody therapeutics by design. *Nat Rev Immunol* 6 (5), 343-357
72 Reichert, J.M. et al. (2005) Monoclonal antibody successes in the clinic. *Nat Biotechnol* 23 (9), 1073-1078
73 Boiesen, P. et al. (2000) Histologic grading in breast cancer--reproducibility between seven pathologic departments. South Sweden Breast Cancer Group. *Acta Oncol* 39 (1), 41-45
74 Dietel, M. and Sers, C. (2006) Personalized medicine and development of targeted therapies: The upcoming challenge for diagnostic molecular pathology. A review. *Virchows Arch* 448 (6), 744-755
75 Orsenigo, C. (2008) Gene Selection and Cancer Microarray Data Classification Via Mixed-Integer Optimization *Lecture Notes in Computer Sciences* 4973, 141-152
76 Isaacs, C. et al. (2001) New prognostic factors for breast cancer recurrence. *Semin Oncol* 28 (1), 53-67
77 Ross, D.T. et al. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24 (3), 227-235
78 Yeang, C.H. et al. (2001) Molecular classification of multiple tumor types. *Bioinformatics* 17 Suppl 1, S316-322
79 Ooi, C.H. and Tan, P. (2003) Genetic algorithms applied to multi-class prediction for the analysis of gene expression data. *Bioinformatics* 19 (1), 37-44
80 Peng, S. et al. (2003) Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. *FEBS Lett* 555 (2), 358-362
81 Massion, P.P. and Carbone, D.P. (2003) The molecular basis of lung cancer: molecular abnormalities and therapeutic implications. *Respir Res* 4, 12
82 Mishra, A.V., M. . (2010) Cancer Biomarkers: Are We Ready for the Prime Time? *Cancers* 2, 190-208
83 Mahmoud H. Hamdan, D.M.D., Nico M. Nibbering. (2007) *Cancer biomarkers: analytical techniques for discovery*, WILEY
84 Sarker, D. and Workman, P. (2007) Pharmacodynamic biomarkers for molecular cancer therapeutics. *Adv Cancer Res* 96, 213-268
85 Bubley, G.J. et al. (1999) Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. *J Clin Oncol* 17 (11), 3461-3467
86 Rustin, G.J. (2003) Use of CA-125 to assess response to new agents in ovarian

cancer trials. *J Clin Oncol* 21 (10 Suppl), 187s-193s

87 Rustin, G.J. et al. (2004) Use of CA-125 in clinical trial evaluation of new therapeutic drugs for ovarian cancer. *Clin Cancer Res* 10 (11), 3919-3926

88 Slamon, D.J. et al. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235 (4785), 177-182

89 Kelloff, G.J. and Sigman, C.C. (2005) New science-based endpoints to accelerate oncology drug development. *Eur J Cancer* 41 (4), 491-501

90 Rossi, A. et al. (2009) Biological prognostic and predictive factors in lung cancer. *Oncology* 77 Suppl 1, 90-96

91 Giallourakis, C. et al. (2005) Disease gene discovery through integrative genomics. *Annu Rev Genomics Hum Genet* 6, 381-406

92 Kim, H.L. and Steinberg, G.D. (2000) New insights and candidate genes and their implications for care of patients with hereditary prostate cancer. *Curr Urol Rep* 1 (1), 9-14

93 Malkin, D. et al. (1990) Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250 (4985), 1233-1238

94 Cui, J.F. et al. (2006) Identification of metastasis candidate proteins among HCC cell lines by comparative proteome and biological function analysis of S100A4 in metastasis in vitro. *Proteomics* 6 (22), 5953-5961

95 Pharoah, P.D. et al. (2007) Association between Common Variation in 120 Candidate Genes and Breast Cancer Risk. *PLoS Genet* 3 (3), e42

96 Smith, A.K. and Meyers, D.A. (2005) Family studies and positional cloning of genes for asthma and related phenotypes. *Immunol Allergy Clin North Am* 25 (4), 641-654

97 Schrijvers, D. (2009) ESMO Handbook of Cancer Diagnosis and Treatment Evaluation.

98 Hutchens TW, Y.T. (1993) New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 7, 576–580

99 Tang, N. et al. (2004) Current developments in SELDI affinity technology. *Mass Spectrom Rev* 23 (1), 34-44

100 Link, A.J. et al. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17 (7), 676-682

101 Wolters, D.A. et al. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* 73 (23), 5683-5690

102 Wang, H. and Hanash, S. (2003) Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 787 (1), 11-18

103 O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250 (10), 4007-4021

104 Aebersold, R. and Goodlett, D.R. (2001) Mass spectrometry in proteomics. *Chem Rev* 101 (2), 269-295

105 Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422 (6928), 198-207

106 Klose, J. and Kobalz, U. (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 16 (6), 1034-1059

107 Espina, V. et al. (2003) Protein microarrays: molecular profiling technologies for clinical specimens. *Proteomics* 3 (11), 2091-2100

108 MacBeath, G. (2002) Protein microarrays and proteomics. *Nat Genet* 32 Suppl, 526-532

109 Petricoin, E.F. and Liotta, L.A. (2003) Clinical applications of proteomics. *J Nutr* 133 (7 Suppl), 2476S-2484S

110 Caprioli, J. et al. (1996) Temporal corneal phacoemulsification combined with superior trabeculectomy: a controlled study. *Trans Am Ophthalmol Soc* 94, 451-463; discussion 463-458

111 Chaurand, P. et al. (1999) Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. *Anal Chem* 71 (23), 5263-5270

112 Stoeckli, M. et al. (2001) Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat Med* 7 (4), 493-496

113 Datta, Y.H. et al. (1994) Sensitive detection of occult breast cancer by the reverse-transcriptase polymerase chain reaction. *J Clin Oncol* 12 (3), 475-482

114 Krismann, M. et al. (1995) Low specificity of cytokeratin 19 reverse transcriptase-polymerase chain reaction analyses for detection of hematogenous lung cancer dissemination. *J Clin Oncol* 13 (11), 2769-2775

115 Velculescu, V.E. et al. (1995) Serial analysis of gene expression. *Science* 270 (5235), 484-487

116 Tusher, V.G. et al. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98 (9), 5116-5121

117 Alizadeh, A.A. et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403 (6769), 503-511

118 Khan, J. et al. (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7 (6), 673-679

119 Ross, M.E. et al. (2003) Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* 102 (8), 2951-2959

120 Yeoh, E.J. et al. (2002) Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1 (2), 133-143

121 Tillinghast, G.W. Microarrays in the clinic. *Nat Biotechnol* 28 (8), 810-812

122 van de Vijver, M.J. et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347 (25), 1999-2009

123 van 't Veer, L.J. et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415 (6871), 530-536

124 Lind, G.E. et al. (2006) ADAMTS1, CRABP1, and NR3C1 identified as epigenetically deregulated genes in colorectal tumorigenesis. *Cell Oncol* 28 (5-6), 259-272

125 Dhanasekaran, S.M. et al. (2001) Delineation of prognostic biomarkers in prostate cancer. *Nature* 412 (6849), 822-826

126 Shi, H. et al. (2007) Discovery of novel epigenetic markers in non-Hodgkin's lymphoma. *Carcinogenesis* 28 (1), 60-70

127 Mrozek, K. et al. (2007) Influence of new molecular prognostic markers in patients with karyotypically normal acute myeloid leukemia: recent advances. *Curr Opin Hematol* 14 (2), 106-114

128 Babu, M.M. (2004) *An Introduction to Microarray Data Analysis*, Horizon Bioscience

129 Leung, Y.F. and Cavalieri, D. (2003) Fundamentals of cDNA microarray data analysis. *Trends Genet* 19 (11), 649-659

130 Pinkel, D. et al. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20 (2), 207-211

131 Hughes, T.R. et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19 (4), 342-347

132 Lockhart, D.J. et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14 (13), 1675-1680

133 Dalma-Weiszhausz, D.D. et al. (2006) The affymetrix GeneChip platform: an overview. *Methods Enzymol* 410, 3-28

134 Demeter, J. et al. (2007) The Stanford Microarray Database: implementation of new analysis tools and open source release of software. *Nucleic Acids Res* 35 (Database issue), D766-770

135 Barrett, T. et al. (2007) NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* 35 (Database issue), D760-765

136 Parkinson, H. et al. (2007) ArrayExpress--a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* 35 (Database issue), D747-750

137 http://www.mged.org/Workgroups/MIAME/miame.html.

138 Tang, Y. et al. (2007) Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis. *IEEE/ACM Trans Comput Biol Bioinform* 4 (3), 365-381

139 Michiels, S. et al. (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365 (9458), 488-492

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

141 Harrison, R. and DeLisi, C. (2002) Condition specific transcription factor binding site characterization in Saccharomyces cerevisiae. *Bioinformatics* 18 (10), 1289-1296

142 Aach, J. et al. (2000) Systematic management and analysis of yeast gene expression data. *Genome Res* 10 (4), 431-445

143 Su, A.I. et al. (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* 99 (7), 4465-4470

144 Smith, C.M. et al. (2007) The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res* 35 (Database issue), D618-623

145 Wiederkehr, C. et al. (2004) GermOnline, a cross-species community knowledgebase on germ cell differentiation. *Nucleic Acids Res* 32 (Database issue), D560-567

146 Haverty, P.M. et al. (2002) HugeIndex: a database with visualization tools for high-density oligonucleotide array data from normal human tissues. *Nucleic Acids Res* 30 (1), 214-217

147 Argraves, G.L. et al. (2003) The MUSC DNA Microarray Database. *Bioinformatics* 19 (18), 2473-2474

148 Bono, H. et al. (2002) READ: RIKEN Expression Array Database. *Nucleic Acids Res* 30 (1), 211-213

149 Yazaki J, K.N., Ishikawa M, Endo D, Kojima K, MicroArray Center, Kikuchi S. (2002) The Rice Expression Database (RED): gateway to rice functional genomics. *Trends in Plant Science* 12, 563-564

150 Manduchi, E., Pizarro, A., Stoeckert, C. (2001) RAD (RNA Abundance Database): an infrastructure for array data analysis. *Proc. SPIE* 4266, 68-78

151 Dwight, S.S. et al. (2002) Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res* 30 (1), 69-72

152 Cheung, K.H. et al. (2002) YMD: a microarray database for large-scale gene expression analysis. *Proc AMIA Symp*, 140-144

153 Lelandais, G. et al. (2004) yMGV: a cross-species expression data mining tool. *Nucleic Acids Res* 32 (Database issue), D323-325

154 Schoch C, D.M., Kern W, Kohlmann A, Schnittger S, Haferlach T. (2004) "Deep insight" into microarray technology. *Atlas Genet Cytogenet Oncol Haematol*

155 DeRisi, J.L. et al. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278 (5338), 680-686

156 Tavazoie, S. et al. (1999) Systematic determination of genetic network architecture. *Nat Genet* 22 (3), 281-285

157 Jansen, R. et al. (2002) Relating whole-genome expression data with protein-protein interactions. *Genome Res* 12 (1), 37-46

158 Ramirez-Benitez Mdel, C. et al. (2001) VIR.II: a new interface with the antibody sequences in the Kabat database. *Biosystems* 61 (2-3), 125-131

159 Alon, U. et al. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 96 (12), 6745-6750

160 Eisen, M.B. et al. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95 (25), 14863-14868

161 van der Pouw Kraan, T.C. et al. (2003) Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair. *Genes Immun* 4 (3), 187-196

162 Sherlock, G. (2000) Analysis of large-scale gene expression data. *Curr Opin Immunol* 12 (2), 201-205

163 Vapnik, V. (1998) Statistical Learning Theory.

164 Bishop, C. (1995) neural networks for pattern recognition.

165 Qiu, P. et al. (2005) Ensemble dependence model for classification and prediction of cancer and normal gene expression data. *Bioinformatics* 21 (14), 3114-3121

166 Li, F. and Yang, Y. (2005) Analysis of recursive gene selection approaches from microarray data. *Bioinformatics* 21 (19), 3741-3747

167 Pochet, N. et al. (2004) Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction. *Bioinformatics* 20 (17), 3185-3195

168 Isabelle Guyon, J.W., Stephen Barnhill, Vladimir Vapnik. (2002) Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46 (1-3), 389-422

169 Furey, T.S. et al. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16 (10), 906-914

170 Brown, M.P. et al. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* 97 (1), 262-267

171 Bellman., R.E. (1961) Adaptive Control Processes.

172 Koeppen, M. (2000) The Curse of Dimensionality. *5th Online World Conference on Soft Computing in Industrial Applications (WSC5)*

173 Inza, I. et al. (2004) Filter versus wrapper gene selection approaches in DNA microarray domains. *Artif Intell Med* 31 (2), 91-103

174 Model, F. et al. (2001) Feature selection for DNA methylation based cancer

classification. *Bioinformatics* 17 Suppl 1, S157-164

175 Robnik-Šikonja, M. and Kononenko, I. (2003) Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning* 53 (1-2), 23-69

176 Ding, C. and Peng, H. (2005) Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol* 3 (2), 185-205

177 Ben-Bassat, M. (1982) Pattern recognition and reduction of dimensionality. *Handbook of statistics II*, p. 773—791.

178 Cheng, J. and Greiner, R. (1999) Comparing Bayesian Network Classifiers. *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, 101-110

179 Aris V, R.M. (2002) A method to improve detection of disease using selectively expressed genes in microarray data. *Methods of Microarray Data Analysis. Proceedings of the First Conference on Critical Assessment of Microarray Data Analysis, CAMDA'00.* p. 69—80.

180 Beibel, M. (2000) Selection of informative genes in gene expression based diagnosis: a nonparametric approach. *Lecture Notes in Computer Sciences. Proceedings of the First International Symposium in Medical Data Analysis, ISMDA'00* 1933, p. 300-307

181 Ding, C. (2002) Analysis of gene expression profiles: class discovery and leaf ordering. *Proceedings of the Sixth International Conference on Research in Computational Molecular Biology*, p. 127-136

182 Baker, S.G. and Kramer, B.S. (2006) Identifying genes that contribute most to good classification in microarrays. *BMC Bioinformatics* 7, 407

183 Kohavi, R. and John, G.H. (97) Wrappers for feature subset selection. *Artificial Intelligence* Special issue on relevance (1-2), 273 - 324

184 Xiong, M. et al. (2001) Biomarker identification by feature wrappers. *Genome Res* 11 (11), 1878-1887

185 Kohavi, R. and John, G.H. (1997) Wrappers for feature subset selection. *Artificial Intelligence* Special issue on relevance (1-2), 273 - 324

186 Talvinen, K. et al. (2006) Biochemical and clinical approaches in evaluating the prognosis of colon cancer. *Anticancer Res* 26 (6C), 4745-4751

187 Ancona, N. et al. (2006) On the statistical assessment of classifiers using DNA microarray data. *BMC Bioinformatics* 7, 387

188 Zhang, X.W. et al. (2005) Molecular diagnosis of human cancer type by gene expression profiles and independent component analysis. *Eur J Hum Genet* 13 (12), 1303-1311

189 Li, W. and Yang, Y. (2002) How Many Genes Are Needed for a Discriminant Microarray Data Analysis ? *Methods of Microarray Data Analysis. Proceedings of the First Conference on Critical Assessment of Microarray Data Analysis, CAMDA'00.*, 137-150

190 Grate, L.R. (2005) Many accurate small-discriminatory feature subsets exist in microarray transcript data: biomarker discovery. *BMC Bioinformatics* 6, 97

191 Slonim, D.K. et al. (2000) Class prediction and discovery using gene expression data. *Proceedings of the 4th Annual International Conference on Computational Molecular Biology (RECOMB)*

192 Ahmed, A.A. and Brenton, J.D. (2005) Microarrays and breast cancer clinical studies: forgetting what we have not yet learnt. *Breast Cancer Res* 7 (3), 96-99

193 Brenton, J.D. et al. (2005) Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol* 23 (29), 7350-7360

194 Bullinger, L. and Valk, P.J. (2005) Gene expression profiling in acute myeloid leukemia. *J Clin Oncol* 23 (26), 6296-6305

195 Bullinger, L. et al. (2004) Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia. *N Engl J Med* 350 (16), 1605-1616

196 Valk, P.J.M. et al. (2004) Prognostically Useful Gene-Expression Profiles in Acute Myeloid Leukemia. *N Engl J Med* 350 (16), 1617-1628

197 Ntzani, E.E. and Ioannidis, J.P. (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 362 (9394), 1439-1444

198 Zhou, X. and Mao, K.Z. (2005) LS Bound based gene selection for DNA microarray data. *Bioinformatics* 21 (8), 1559-1564

199 Bo, T. and Jonassen, I. (2002) New feature subset selection procedures for classification of expression profiles. *Genome Biol* 3 (4), RESEARCH0017

200 Huang, T.M. and Kecman, V. (2005) Gene extraction for cancer diagnosis by support vector machines--an improvement. *Artif Intell Med* 35 (1-2), 185-194

201 Liu, X. et al. (2005) An entropy-based gene selection method for cancer classification using microarray data. *BMC Bioinformatics* 6 (1), 76

202 Draghici, S. et al. (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* 22 (2), 101-109

203 Ioannidis, J.P. (2005) Microarrays and molecular research: noise discovery? *Lancet* 365 (9458), 454-455

204 Gardner, S.N. and Fernandes, M. (2005) Prediction of cancer outcome with microarrays. *Lancet* 365 (9472), 1685

205 Biganzoli, E. et al. (2005) Prediction of cancer outcome with microarrays. *Lancet* 365 (9472), 1683; author reply 1684-1685

206 Winegarden, N. (2003) Microarrays in cancer: moving from hype to clinical reality. *Lancet* 362 (9394), 1428

207 Disis, M.L. et al. (1997) High-titer HER-2/neu protein-specific antibody can be detected in patients with early-stage breast cancer. *J Clin Oncol* 15 (11), 3363-3367

208 Jager, D. et al. (2002) Identification of tumor-restricted antigens NY-BR-1, SCP-1, and a new cancer/testis-like antigen NW-BR-3 by serological screening of a testicular library with breast cancer serum. *Cancer Immun* 2, 5

209 Scanlan, M.J. et al. (2001) Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression. *Cancer Immun* 1, 4

210 Altuvia, Y. and Margalit, H. (2004) A structure-based approach for prediction of MHC-binding peptides. *Methods* 34 (4), 454-459

211 Adam, J.K. et al. (2003) Immune responses in cancer. *Pharmacol Ther* 99 (1), 113-132

212 Goldman, B. and DeFrancesco, L. (2009) The cancer vaccine roller coaster. *Nat Biotechnol* 27 (2), 129-139

213 De Leo, A.B. (2005) p53-based immunotherapy of cancer. Approaches ro reversing unresponsiveness to T lymphocytes and preventing tumor escape. *Adv Otorhinolaryngol* 62, 134-150

214 Mapara, M.Y. and Sykes, M. (2004) Tolerance and cancer: mechanisms of tumor evasion and strategies for breaking tolerance. *J Clin Oncol* 22 (6), 1136-1151

215 Wu, J. and Lanier, L.L. (2003) Natural killer cells and cancer. *Adv Cancer Res* 90, 127-156

216 Eisenbach, L. et al. (2000) Antitumor vaccination using peptide based vaccines. *Immunol Lett* 74 (1), 27-34

217 Lollini, P.L. et al. (2006) Vaccines for tumour prevention. *Nat Rev Cancer* 6 (3), 204-216

218 Mittendorf, E.A. et al. (2007) Breast cancer vaccines: promise for the future or pipe dream? *Cancer* 110 (8), 1677-1686

219 Schlom, J. et al. (2007) Cancer vaccines: moving beyond current paradigms. *Clin Cancer Res* 13 (13), 3776-3782

220 Purcell, A.W. et al. (2007) More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 6 (5), 404-414

221 Lichtman, A.K.A.A.H. (2005) *Cellular and Molecular Immunology, Updated Edition (Book + Student Consult + Evolve* W.B. Saunders Company

222 Shoshan, S.H. and Admon, A. (2004) MHC-bound antigens and proteomics for novel target discovery. *Pharmacogenomics* 5 (7), 845-859

223 Larsen, M.V. et al. (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 35 (8), 2295-2303

224 Tanahashi, N. et al. (1993) Molecular structure of 20S and 26S proteasomes. *Enzyme Protein* 47 (4-6), 241-251

225 Eggers, M. et al. (1995) The cleavage preference of the proteasome governs the yield of antigenic peptides. *J Exp Med* 182 (6), 1865-1870

226 Craiu, A. et al. (1997) Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc Natl Acad Sci U S A* 94 (20), 10850-10855

227 Serwold, T. et al. (2001) ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nat Immunol* 2 (7), 644-651

228 Serwold, T. et al. (2002) ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 419 (6906), 480-483

229 York, I.A. et al. (2002) The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat Immunol* 3 (12), 1177-1184

230 Saric, T. et al. (2002) An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat Immunol* 3 (12), 1169-1176

231 Nielsen, M. et al. (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57 (1-2), 33-41

232 Peters, B. et al. (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol* 171 (4), 1741-1749

233 Nussbaum, A.K. et al. (2003) Using the World Wide Web for predicting CTL epitopes. *Curr Opin Immunol* 15 (1), 69-74

234 Houghton, A.N. et al. (2001) Immunity against cancer: lessons learned from melanoma. *Curr Opin Immunol* 13 (2), 134-140

235 Parmiani, G. et al. (2007) Unique human tumor antigens: immunobiology and use in clinical trials. *J Immunol* 178 (4), 1975-1979

236 Van den Eynde, B.J. and van der Bruggen, P. (1997) T cell defined tumor antigens. *Curr Opin Immunol* 9 (5), 684-693

237 Van Der Bruggen, P. et al. (2002) Tumor-specific shared antigenic peptides recognized by human T cells. *Immunol Rev* 188, 51-64

238 Honeyman, M.C. et al. (1998) Neural network-based prediction of candidate T-cell epitopes. *Nat Biotechnol* 16 (10), 966-969

239 Altuvia, Y. et al. (1997) A structure-based algorithm to predict potential binding

peptides to MHC molecules with hydrophobic binding pockets. *Hum Immunol* 58 (1), 1-11

240 Mallios, R.R. (2001) Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics* 17 (10), 942-948

241 Reche, P.A. et al. (2004) Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 56 (6), 405-419

242 Doytchinova, I.A. et al. (2004) Coupling in silico and in vitro analysis of peptide-MHC binding: a bioinformatic approach enabling prediction of superbinding peptides and anchorless epitopes. *J Immunol* 172 (12), 7495-7502

243 Donnes, P. and Elofsson, A. (2002) Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics* 3, 25

244 Nielsen, M. et al. (2004) Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics* 20 (9), 1388-1397

245 Sung, M.H. and Simon, R. (2004) Candidate epitope identification using peptide property models: application to cancer immunotherapy. *Methods* 34 (4), 460-467

246 Bhasin, M. and Raghava, G.P. (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* 22 (23-24), 3195-3204

247 Petrovsky, N. and Brusic, V. (2004) Virtual models of the HLA class I antigen processing pathway. *Methods* 34 (4), 429-435

248 Donnes, P. and Kohlbacher, O. (2005) Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci* 14 (8), 2132-2140

249 Tenzer, S. et al. (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* 62 (9), 1025-1037

250 Zhao, Y. et al. (2003) Application of support vector machines for T-cell epitopes prediction. *Bioinformatics* 19 (15), 1978-1984

251 Bian, H. and Hammer, J. (2004) Discovery of promiscuous HLA-II-restricted T cell epitopes with TEPITOPE. *Methods* 34 (4), 468-475

252 Rudolph, M.G. et al. (2006) How TCRs Bind MHCs, Peptides, and Coreceptors. *Annu Rev Immunol*

253 Burges, C.J.C. (1998) A tutorial on Support Vector Machine for pattern recognition. *Data Mining and Knowledge Discovery* 2, 121-167

254 Cai, C.Z. et al. (2003) SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res* 31 (13), 3692-3697

255 Han, L.Y. et al. (2004) Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. *Rna* 10 (3), 355-368

256 Han, L.Y. et al. (2005) Prediction of functional class of novel viral proteins by a statistical learning method irrespective of sequence similarity. *Virology* 331 (1), 136-143

257 Matsumura, M. et al. (1992) Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science* 257 (5072), 927-934

258 Zhang, C. et al. (1998) Structural principles that govern the peptide-binding motifs of class I MHC molecules. *J Mol Biol* 281 (5), 929-947

259 McFarland, B.J. and Beeson, C. (2002) Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med Res Rev* 22 (2), 168-203

260 Cui, J. et al. (2007) Prediction of MHC-binding peptides of flexible lengths from sequence-derived structural and physicochemical properties. *Mol Immunol* 44 (5), 866-877

261 Cui, L. et al. (2007) Cytotoxic effect of curcumin on malaria parasite Plasmodium falciparum: inhibition of histone acetylation and generation of reactive oxygen species. *Antimicrob Agents Chemother* 51 (2), 488-494

262 Peters, B. et al. (2005) The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 3 (3), e91

263 Mitchell, T.M. (1997) *Machine Learning*, McGraw-Hill Science/Engineering/Math

264 Pierre Baldi and Brunak, S. (2001) *Bioinformatics: The Machine Learning Approach, Second Edition*, The MIT Press

265 Dietterich, T.G. (2003) Machine Learning. In *Nature Encyclopedia of Cognitive Science* (Vol. 2), pp. 971-981, Nature Publishing Group

266 Vapnik, V. (1979) *Estimation of dependences based on empirical data [in Russian]. [English tanslation: Spring Verlag, New York, 1982]*

267 Vapnik, V. (1995) *The nature of statistical learning theory*, Springer

268 Souheil Ben-Yacoub, Y.A., and Eddy Mayoraz. (1999) Fusion of Face and Speech Data for Person Identity Verification. *IEEE transactions on neural networks* 10, 1065-1074

269 Karlsen, R.E.G., David J.; Gerhart, Grant R. (2000) Target classification via support vector machines. *Optical Engineering* 39 (3), 704-711

270 Shin, C.S.K., K.I. Park, M.H. Kim, H.J. . (2001) Support vector machine-based text detection in digital video. *Pattern recognition* 34, 527-529

271 Yuan, Z. et al. (2002) Prediction of protein solvent accessibility using support vector machines. *Proteins* 48 (3), 566-570

272 Ding, C.H. and Dubchak, I. (2001) Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* 17 (4), 349-358

273 Hua, S. and Sun, Z. (2001) A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. *J Mol Biol* 308 (2), 397-407

274 Bock, J.R. and Gough, D.A. (2001) Predicting protein--protein interactions from primary structure. *Bioinformatics* 17 (5), 455-460

275 Vapnik, V.N. (1995) *The nature of statistical learning theory*, Springer

276 Burges, C.J.C. (1998) A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2 (2), 127-167

277 Pochet, N. et al. (2004) Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction. *Bioinformatics* 20, 3185-3195

278 Li, F. and Yang, Y. (2005) Analysis of recursive gene selection approaches from microarray data. *Bioinformatics* 21, 3741-3747

279 Burges, C. (1998) A tutorial on Support Vector Machine for pattern recognition. *Data Min. Knowl. Disc.* 2, 121-167

280 Karchin, R. et al. (2002) Classifying G-protein coupled receptors with support vector machines. *Bioinformatics* 18 (1), 147-159

281 Cai, Y.D. and Lin, S.L. (2003) Support vector machines for predicting rRNA-, RNA-, and DNA-binding proteins from amino acid sequence. *Biochim Biophys Acta* 1648 (1-2), 127-133

282 Keerthi, S.S. and Lin, C.J. (2003) Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput* 15 (7), 1667-1689

283 Lin, H.-T., C.-J. Lin. (2003) *A study on sigmoid kernels for SVM and the trainingof non-PSD kernels by SMO-type methods. Technical report, Department of Computer Science, National Taiwan University.*

284 Chang, C.C. and Lin, C.J. (2001) LIBSVM : a library for support vector machines.

285 Burges, C.J.C. (1998) A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* 2, 121-167

286 Wang, D. and Larder, B. (2003) Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks. *J Infect Dis* 188 (5), 653-660

287 Draghici, S. and Potter, R.B. (2003) Predicting HIV drug resistance with neural networks. *Bioinformatics* 19 (1), 98-107

288 Specht, D.F. (1990) Probabilistic neural networks. *Neural Networks* 3 (1), 109-118

289 Parzen, E. (1962) On estimation of a probability density function and mode. *Ann. Math. Stat.* 33, 1065-1076

290 Cacoullos, T. (1966) Estimation of a multivariate density. *Ann. I. Stat. Math.* 18, 179-189

291 Kumar, M.S.G.K.V. (2000) A Comparison of Document Clustering Techniques. *KDD Workshop on Text Mining, University of Minnesota*

292 Olson, C.F. (1993) Parallel Algorithms for Hierarchical Clustering. *Technical Report UCB//CSD-94-786, University of California at Berkeley*

293 Jain A.K., M.M.N., and Flynn P.J. (1999) Data Clustering: A Review. *ACM Computing Surveys* 31 (3), 264-323

294 Pang-Ning Tan, M.S., Vipin Kumar. (2005) Introduction to Data Mining. *Addison Wesley*

295 Rong En Fan, P.H.C., Chih Jen Lin. (2005) Working set selection using second order information for training SVM. *Journal of Machine Learning Research* 6, 1889-1918

296 Richard O. Duda, P.E.H., David G. Stork. (2000) Pattern Classification, 2nd Edition. *Wiley*

297 Hawkins, D.M. (2004) The problem of overfitting. *J Chem Inf Comput Sci* 44 (1), 1-12

298 Trevor Hastie, R.T., Jerome H. Friedman. (2001) The elements of statistical learning: data mining, inference, and prediction. *Springer*

299 Matthews, B. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 405 (2), 442-451

300 Gunnarsson, R.K. and Lanke, J. (2002) The predictive value of microbiologic diagnostic tests if asymptomatic carriers are present. *Stat Med* 21 (12), 1773-1785

301 Li, H. et al. (2007) Machine learning approaches for predicting compounds that interact with therapeutic and ADMET related proteins. *J Pharm Sci* 96 (11), 2838-2860

302 Zhao, Y. et al. (2005) An adaptive method for cDNA microarray normalization. *BMC Bioinformatics* 6, 28

303 Yang, Y.H. et al. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30 (4), e15

304 WS, C. (1974) Robust locally weighted regression and smoothing scatterplots. *Jounral of the American Statistical Association* 74, 829-836

305 Kepler, T.B. et al. (2002) Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biol* 3 (7), RESEARCH0037

306 Wang, Y. et al. (2002) Iterative normalization of cDNA microarray data. *IEEE*

Trans Inf Technol Biomed 6 (1), 29-37

307 Workman, C. et al. (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 3 (9), research0048

308 Edwards, D. (2003) Non-linear normalization and background correction in one-channel cDNA microarray studies. *Bioinformatics* 19 (7), 825-833

309 Chen, Y.J. et al. (2003) Normalization methods for analysis of microarray gene-expression data. *J Biopharm Stat* 13 (1), 57-74

310 Irizarry, R.A. et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4 (2), 249-264

311 Irizarry, R.A. et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31 (4), e15

312 Ihaka R, G.R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5 (3), 299-314

313 Chu, W. et al. (2005) Biomarker discovery in microarray gene expression data with Gaussian processes. *Bioinformatics* 21 (16), 3385-3393

314 Michael E. Wall, A.R., Luis M. Rocha. (2002) *Microarray analysis techniques:Singular value decomposition and principal component analysis*, Kluwer Academic Press

315 Saccone, R.A., Rauniyar, R. K. and Patti M.-E. (2002) Sources of Experimental Variability In Expression Data Derived From High-Density Oligonucleotide Microarrays: Practical Experience From An Academic Core Laboratory. *2nd Annual UMass Bioinformatics Conference, UMass Lowell*

316 Pomeroy, S.L. et al. (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415 (6870), 436-442

317 Stegmaier, K. et al. (2004) Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. *Nat Genet* 36 (3), 257-263

318 Schuchhardt, J. et al. (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Res* 28 (10), E47

319 Tu, Y. et al. (2002) Quantitative noise analysis for gene expression microarray experiments. *Proc Natl Acad Sci U S A* 99 (22), 14031-14036

320 Bo, T.H. et al. (2004) LSimpute: accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Res* 32 (3), e34

321 de Brevern, A.G. et al. (2004) Influence of microarrays experiments missing values on the stability of gene groups by hierarchical clustering. *BMC Bioinformatics* 5, 114

322 Hu, J. et al. (2006) Integrative missing value estimation for microarray data. *BMC Bioinformatics* 7, 449

323 Troyanskaya, O. et al. (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17 (6), 520-525

324 Kim, H. et al. (2005) Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics* 21 (2), 187-198

325 Oba, S. et al. (2003) A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics* 19 (16), 2088-2096

326 Scholz, M. et al. (2005) Non-linear PCA: a missing data approach. *Bioinformatics* 21 (20), 3887-3895

327 Bair, E. and Tibshirani, R. (2004) Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2 (4), E108

328 Scheel, I. et al. (2005) The influence of missing value imputation on detection of differentially expressed genes from microarray data. *Bioinformatics* 21 (23), 4272-

4279
329 http://helix-web.stanford.edu/pubs/impute/.
330 Hanson, S.J. and Halchenko, Y.O. (2008) Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. *Neural Comput* 20 (2), 486-503
331 Guyon, I. et al. (2002) Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46 (1-3), 389-422
332 Sima, C. et al. (2005) Superior feature-set ranking for small samples using bolstered error estimation. *Bioinformatics* 21 (7), 1046-1054
333 Fu, W.J. et al. (2005) Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics* 21 (9), 1979-1986
334 Henderson, R.A. et al. (1992) HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science* 255 (5049), 1264-1266
335 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287 (5460), 1960-1964
336 Peter I., C.S. (2007) Drugs,their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery* 5 (5460), 821-834
337 Zheng, C.J. et al. (2006) Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev* 58 (2), 259-279
338 Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 3 (8), 673-683
339 Nelson, H.S. (2001) Advair: combination treatment with fluticasone propionate/salmeterol in the treatment of asthma. *J Allergy Clin Immunol* 107 (2), 398-416
340 Gupta, E.K. and Ito, M.K. (2002) Lovastatin and extended-release niacin combination product: the first drug combination for the management of hyperlipidemia. *Heart Dis* 4 (2), 124-137
341 Larder, B.A. et al. (1995) Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science* 269 (5224), 696-699
342 Dancey, J.E. and Chen, H.X. (2006) Strategies for optimizing combinations of molecularly targeted anticancer agents. *Nat Rev Drug Discov* 5 (8), 649-659
343 Silver, L.L. (2007) Multi-targeting by monotherapeutic antibacterials. *Nat Rev Drug Discov* 6 (1), 41-55
344 Shoemaker, R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 6 (10), 813-823
345 Graham, B.A. et al. (2000) Synergistic interactions between two alpha(2)-adrenoceptor agonists, dexmedetomidine and ST-91, in two substrains of Sprague-Dawley rats. *Pain* 85 (1-2), 135-143
346 Kisliuk, R.L. (2000) Synergistic interactions among antifolates. *Pharmacol Ther* 85 (3), 183-190
347 Rand, K.H. and Houck, H. (2004) Daptomycin synergy with rifampicin and ampicillin against vancomycin-resistant enterococci. *J Antimicrob Chemother* 53 (3), 530-532
348 Dryselius, R. et al. (2005) Antimicrobial synergy between mRNA- and protein-level inhibitors. *Journal of Antimicrobial Chemotherapy* 56 (1), 97-103
349 Azrak, R.G. et al. (2006) The mechanism of methylselenocysteine and docetaxel synergistic activity in prostate cancer cells. *Mol Cancer Ther* 5 (10), 2540-2548
350 Bell, A. (2005) Antimalarial drug synergism and antagonism: mechanistic and clinical significance. *FEMS Microbiol Lett* 253 (2), 171-184

351 Robertson, J.G. (2005) Mechanistic basis of enzyme-targeted drugs. *Biochemistry* 44 (15), 5561-5571

352 Zybarth, G. and Kley, N. (2006) Investigating the molecular basis of drug action and response: chemocentric genomics and proteomics. *Curr Drug Targets* 7 (4), 387-395

353 Yao, L.X. et al. (2006) Internet resources related to drug action and human response: a review. *Appl Bioinformatics* 5 (3), 131-139

354 Liu, T. et al. (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35 (Database issue), D198-201

355 Ji, Z.L. et al. (2003) Internet resources for proteins associated with drug therapeutic effects, adverse reactions and ADME. *Drug Discov Today* 8 (12), 526-529

356 Chen, Y.Z. and Zhi, D.G. (2001) Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* 43 (2), 217-226

357 Paul, N. et al. (2004) Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* 54 (4), 671-680

358 Cleves, A.E. and Jain, A.N. (2006) Robust ligand-based modeling of the biological targets of known drugs. *J Med Chem* 49 (10), 2921-2938

359 Armour, C.D. and Lum, P.Y. (2005) From drug to protein: using yeast genetics for high-throughput target discovery. *Curr Opin Chem Biol* 9 (1), 20-24

360 Nettles, J.H. et al. (2006) Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. *J Med Chem* 49 (23), 6802-6810

361 Han, L.Y. et al. (2007) Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. *Drug Discov Today* 12 (7-8), 304-313

362 Chen, X. et al. (2007) Does drug-target have a likeness? *Methods Inf Med* 46 (3), 360-366

363 Greco, W.R. et al. (1995) The search for synergy: a critical review from a response surface perspective. *Pharmacol Rev* 47 (2), 331-385

364 Chou, T.C. (2006) Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol Rev* 58 (3), 621-681

365 Dolara, P. et al. (1992) Sister-chromatid exchanges in human lymphocytes induced by dimethoate, omethoate, deltamethrin, benomyl and their mixture. *Mutat Res* 283 (2), 113-118

366 Johnson, M.D. et al. (2004) Combination antifungal therapy. *Antimicrob Agents Chemother* 48 (3), 693-715

367 Peterson, J.J. and Novick, S.J. (2007) Nonlinear blending: a useful general concept for the assessment of combination drug synergy. *J Recept Signal Transduct Res* 27 (2-3), 125-146

368 Berenbaum, M.C. (1989) What is synergy? *Pharmacol Rev* 41 (2), 93-141

369 Chou, T.C. (2006) Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacological Reviews* 58 (3), 621-681

370 Tallarida, R.J. (2007) Interactions between drugs and occupied receptors. *Pharmacology & Therapeutics* 113 (1), 197-209

371 Jonker, D.M. et al. (2005) Towards a mechanism-based analysis of

pharmacodynamic drug-drug interactions in vivo. *Pharmacology & Therapeutics* 106 (1), 1-18

372 Peters, G.J. et al. (2000) Basis for effective combination cancer chemotherapy with antimetabolites. *Pharmacology & Therapeutics* 87 (2-3), 227-253

373 Barrera, N.P. et al. (2005) Principles: mechanisms and modeling of synergism in cellular responses. *Trends in Pharmacological Sciences* 26 (10), 526-532

374 Falagas, M.E. et al. (2008) Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 22 (2), 338-342

375 Wheeler, D.L. et al. (2004) Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* 32 (Database issue), D35-40

376 Kawakami, H. et al. (2007) Inhibition of heat shock protein-90 modulates multiple functions required for survival of human T-cell leukemia virus type I-infected T-cell lines and adult T-cell leukemia cells. *Int J Cancer* 120 (8), 1811-1820

377 Arthur J. Atkinson, D.R.A., Charles E. Daniels, FASHP, Robert L. Dedrick, and Sanford P. Markey. (2007) Principles of Clinical Pharmacology (Second Edition). *Elsevier*

378 Lin, X. et al. (1999) The role of DNA mismatch repair in cisplatin mutagenicity. *Journal of Inorganic Biochemistry* 77 (1-2), 89-93

379 Rhee, I. et al. (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416 (6880), 552-556

380 van Waardenburg, R.C. et al. (2004) Platinated DNA adducts enhance poisoning of DNA topoisomerase I by camptothecin. *Journal of Biological Chemistry* 279 (52), 54502-54509

381 Grimaldi, K.A. et al. (1994) DNA damage by anti-cancer agents resolved at the nucleotide level of a single copy gene: evidence for a novel binding site for cisplatin in cells. *Nucleic Acids Res* 22 (12), 2311-2317

382 Bassett, E. et al. (2003) Efficiency of extension of mismatched primer termini across from cisplatin and oxaliplatin adducts by human DNA polymerases beta and eta in vitro. *Biochemistry* 42 (48), 14197-14206

383 Koster, D.A. et al. (2007) Antitumour drugs impede DNA uncoiling by topoisomerase I. *Nature* 448 (7150), 213-217

384 D'Incalci, M. et al. (2003) The combination of yondelis and cisplatin is synergistic against human tumor xenografts. *Eur J Cancer* 39 (13), 1920-1926

385 Marco, E. and Gago, F. (2005) DNA structural similarity in the 2:1 complexes of the antitumor drugs trabectedin (Yondelis) and chromomycin A3 with an oligonucleotide sequence containing two adjacent TGG binding sites on opposing strands. *Mol Pharmacol* 68 (6), 1559-1567

386 Dziegielewska, B. et al. (2004) SV40 DNA replication inhibition by the monofunctional DNA alkylator Et743. *Biochemistry* 43 (44), 14228-14237

387 Dai, Z. et al. (2006) 5-Aza-2'-deoxycytidine and depsipeptide synergistically induce expression of BIK (BCL2-interacting killer). *Biochem Biophys Res Commun* 351 (2), 455-461

388 Georgakis, G.V. et al. (2006) The HSP90 inhibitor 17-AAG synergizes with doxorubicin and U0126 in anaplastic large cell lymphoma irrespective of ALK expression. *Exp Hematol* 34 (12), 1670-1679

389 Soja, P.J. et al. (2001) Spontaneous spike activity of spinoreticular tract neurons during sleep and wakefulness. *Sleep* 24 (1), 18-25

390 Staud, R. (2002) Evidence of involvement of central neural mechanisms in

generating fibromyalgia pain. *Curr Rheumatol Rep* 4 (4), 299-305

391 Tham, S.M. et al. (2005) Synergistic and additive interactions of the cannabinoid agonist CP55,940 with mu opioid receptor and alpha2-adrenoceptor agonists in acute pain models in mice. *British Journal of Pharmacology* 144 (6), 875-884

392 Malonga, H. et al. (2005) Taxol anticancer activity and DNA binding. *Mini Rev Med Chem* 5 (3), 307-311

393 Sintchak, M.D. et al. (1996) Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic acid. *Cell* 85 (6), 921-930

394 Marcus, A.I. et al. (2005) The synergistic combination of the farnesyl transferase inhibitor lonafarnib and paclitaxel enhances tubulin acetylation and requires a functional tubulin deacetylase. *Cancer Res* 65 (9), 3883-3893

395 Piperno, G. et al. (1987) Microtubules containing acetylated alpha-tubulin in mammalian cells in culture. *J Cell Biol* 104 (2), 289-302

396 Lai, G.H. et al. (2003) Celecoxib acts in a cyclooxygenase-2-independent manner and in synergy with emodin to suppress rat cholangiocarcinoma growth in vitro through a mechanism involving enhanced Akt inactivation and increased activation of caspases-9 and -3. *Molecular cancer therapeutics* 2 (3), 265-271

397 Alloza, I. et al. (2006) Celecoxib inhibits interleukin-12 alphabeta and beta2 folding and secretion by a novel COX2-independent mechanism involving chaperones of the endoplasmic reticulum. *Mol Pharmacol* 69 (5), 1579-1587

398 Jayasuriya, H. et al. (1992) Emodin, a protein tyrosine kinase inhibitor from Polygonum cuspidatum. *J Nat Prod* 55 (5), 696-698

399 Olsen, B.B. et al. (2007) Emodin negatively affects the phosphoinositide 3-kinase/AKT signalling pathway: a study on its mechanism of action. *Int J Biochem Cell Biol* 39 (1), 227-237

400 Cottagnoud, P. et al. (2003) Vancomycin acts synergistically with gentamicin against penicillin-resistant pneumococci by increasing the intracellular penetration of gentamicin. *Antimicrobial Agents and Chemotherapy* 47 (1), 144-147

401 Yoshizawa, S. et al. (1998) Structural origins of gentamicin antibiotic action. *Embo J* 17 (22), 6437-6448

402 Cegelski, L. et al. (2002) Rotational-echo double resonance characterization of the effects of vancomycin on cell wall synthesis in Staphylococcus aureus. *Biochemistry* 41 (43), 13053-13058

403 Watanakunakorn, C. (1984) Mode of action and in-vitro activity of vancomycin. *J Antimicrob Chemother* 14 Suppl D, 7-18

404 Goddard, J. et al. (2004) Endothelin A receptor antagonism and angiotensin-converting enzyme inhibition are synergistic via an endothelin B receptor-mediated and nitric oxide-dependent mechanism. *Journal of the American Society of Nephrology* 15 (10), 2601-2610

405 Verhaar, M.C. et al. (1998) Endothelin-A receptor antagonist-mediated vasodilatation is attenuated by inhibition of nitric oxide synthesis and by endothelin-B receptor blockade. *Circulation* 97 (8), 752-756

406 Moridaira, K. et al. (2003) ACE inhibition increases expression of the ETB receptor in kidneys of mice with unilateral obstruction. *Am J Physiol Renal Physiol* 284 (1), F209-217

407 Pollock, D.M. et al. (1995) Endothelin receptors and calcium signaling. *Faseb J* 9 (12), 1196-1204

408 Chaney, S.G. et al. (2004) Protein interactions with platinum-DNA adducts: from

structure to function. *J Inorg Biochem* 98 (10), 1551-1559

409 Faivre, S. et al. (2003) DNA strand breaks and apoptosis induced by oxaliplatin in cancer cells. *Biochem Pharmacol* 66 (2), 225-237

410 Koizumi, F. et al. (2004) Synergistic interaction between the EGFR tyrosine kinase inhibitor gefitinib ("Iressa") and the DNA topoisomerase I inhibitor CPT-11 (irinotecan) in human colorectal cancer cells. *Int J Cancer* 108 (3), 464-472

411 Tanaka, R. et al. (2005) Synergistic interaction between oxaliplatin and SN-38 in human gastric cancer cell lines in vitro. *Oncol Rep* 14 (3), 683-688

412 Kobayashi, S. et al. (1993) Singly-linked catenation and knotting of cisplatin-DNA adduct by DNA topoisomerase I. *Nucleic Acids Symp Ser* (29), 137-138

413 Zhao, W.H. et al. (2001) Mechanism of synergy between epigallocatechin gallate and beta-lactams against methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother* 45 (6), 1737-1742

414 Bickle, M. et al. (1998) Cell wall integrity modulates RHO1 activity via the exchange factor ROM2. *Embo J* 17 (8), 2235-2245

415 Abal, M. et al. (2003) Taxanes: microtubule and centrosome targets, and cell cycle dependent mechanisms of action. *Curr Cancer Drug Targets* 3 (3), 193-203

416 Ganansia-Leymarie, V. et al. (2003) Signal transduction pathways of taxanes-induced apoptosis. *Curr Med Chem Anticancer Agents* 3 (4), 291-306

417 Park, S.J. et al. (2004) Taxol induces caspase-10-dependent apoptosis. *J Biol Chem* 279 (49), 51057-51067

418 Okano, J. et al. (2007) The growth inhibition of liver cancer cells by paclitaxel and the involvement of extracellular signal-regulated kinase and apoptosis. *Oncol Rep* 17 (5), 1195-1200

419 Zhang, W. et al. (1999) ERK pathway mediates the activation of Cdk2 in IGF-1-induced proliferation of human osteosarcoma MG-63 cells. *J Bone Miner Res* 14 (4), 528-535

420 Bacus, S.S. et al. (2001) Taxol-induced apoptosis depends on MAP kinase pathways (ERK and p38) and is independent of p53. *Oncogene* 20 (2), 147-155

421 Pennati, M. et al. (2005) Potentiation of paclitaxel-induced apoptosis by the novel cyclin-dependent kinase inhibitor NU6140: a possible role for survivin down-regulation. *Mol Cancer Ther* 4 (9), 1328-1337

422 Lee, E.J. et al. (2007) The epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 (Iressa) suppresses proliferation and invasion of human oral squamous carcinoma cells via p53 independent and MMP, uPAR dependent mechanism. *Ann N Y Acad Sci* 1095, 113-128

423 Fanucchi, M. and Khuri, F.R. (2006) Taxanes in the treatment of non-small cell lung cancer. *Treat Respir Med* 5 (3), 181-191

424 Takabatake, D. et al. (2007) Tumor inhibitory effect of gefitinib (ZD1839, Iressa) and taxane combination therapy in EGFR-overexpressing breast cancer cell lines (MCF7/ADR, MDA-MB-231). *Int J Cancer* 120 (1), 181-188

425 Funakoshi, M. et al. (2001) A MEK inhibitor, PD98059 enhances IL-1-induced NF-kappaB activation by the enhanced and sustained degradation of IkappaBalpha. *Biochem Biophys Res Commun* 283 (1), 248-254

426 Roberts, P.J. and Der, C.J. (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* 26 (22), 3291-3310

427 De Clercq, E. (1993) HIV-1-specific RT inhibitors: highly selective inhibitors of human immunodeficiency virus type 1 that are specifically targeted at the viral reverse transcriptase. *Med Res Rev* 13 (3), 229-258

428 Fattorusso, C. et al. (2005) Specific targeting highly conserved residues in the HIV-1 reverse transcriptase primer grip region. Design, synthesis, and biological evaluation of novel, potent, and broad spectrum NNRTIs with antiviral activity. *J Med Chem* 48 (23), 7153-7165

429 Cruchaga, C. et al. (2005) Inhibition of phosphorolysis catalyzed by HIV-1 reverse transcriptase is responsible for the synergy found in combinations of 3'-azido-3'-deoxythymidine with nonnucleoside inhibitors. *Biochemistry* 44 (9), 3535-3546

430 Rigourd, M. et al. (2002) Primer unblocking and rescue of DNA synthesis by azidothymidine (AZT)-resistant HIV-1 reverse transcriptase: comparison between initiation and elongation of reverse transcription and between (-) and (+) strand DNA synthesis. *J Biol Chem* 277 (21), 18611-18618

431 Drew, R.H. and Gallis, H.A. (1992) Azithromycin--spectrum of activity, pharmacokinetics, and clinical applications. *Pharmacotherapy* 12 (3), 161-173

432 Wynd, M.A. and Paladino, J.A. (1996) Cefepime: a fourth-generation parenteral cephalosporin. *Ann Pharmacother* 30 (12), 1414-1424

433 Fernandez-Cuenca, F. et al. (2003) In vitro activity of azithromycin in combination with amikacin, ceftazidime, ciprofloxacin or imipenem against clinical isolates of Acinobacter baumannii. *Chemotherapy* 49 (1-2), 24-26

434 Gajate, C. and Mollinedo, F. (2005) Cytoskeleton-mediated death receptor and ligand concentration in lipid rafts forms apoptosis-promoting clusters in cancer chemotherapy. *J Biol Chem* 280 (12), 11641-11647

435 Cuadrado, A. et al. (2004) JNK activation is critical for Aplidin-induced apoptosis. *Oncogene* 23 (27), 4673-4680

436 Biscardi, M. et al. (2005) VEGF inhibition and cytotoxic effect of aplidin in leukemia cell lines and cells from acute myeloid leukemia. *Ann Oncol* 16 (10), 1667-1674

437 Pelicano, H. et al. (2006) Targeting Hsp90 by 17-AAG in leukemia cells: mechanisms for synergistic and antagonistic drug combinations with arsenic trioxide and Ara-C. *leukemia* 20 (4), 610-619

438 Abdel-Aziz, W. et al. (2000) Ara-C affects formation of cancer cell DNA synthesome replication intermediates. *Cancer Chemother Pharmacol* 45 (4), 312-319

439 de Vries, J.F. et al. (2006) The mechanisms of Ara-C-induced apoptosis of resting B-chronic lymphocytic leukemia cells. *Haematologica* 91 (7), 912-919

440 Humeniuk, R. et al. (2007) Aplidin synergizes with cytosine arabinoside: functional relevance of mitochondria in Aplidin-induced cytotoxicity. *Leukemia*

441 Hajra, K.M. and Liu, J.R. (2004) Apoptosome dysfunction in human cancer. *Apoptosis* 9 (6), 691-704

442 Haggarty, S.J. et al. (2003) Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. *Proc Natl Acad Sci U S A* 100 (8), 4389-4394

443 Di Francesco, A.M. et al. (2007) The novel atypical retinoid ST1926 is active in ATRA resistant neuroblastoma cells acting by a different mechanism. *Biochem Pharmacol* 73 (5), 643-655

444 Zanchi, C. et al. (2005) Modulation of survival signaling pathways and persistence of the genotoxic stress as a basis for the synergistic interaction between the atypical retinoid ST1926 and the epidermal growth factor receptor inhibitor ZD1839. *Cancer Res* 65 (6), 2364-2372

445 Zwang, Y. and Yarden, Y. (2006) p38 MAP kinase mediates stress-induced

internalization of EGFR: implications for cancer chemotherapy. *Embo J* 25 (18), 4195-4206

446 Reffelmann, T. and Kloner, R.A. (2006) Cardiovascular effects of phosphodiesterase 5 inhibitors. *Curr Pharm Des* 12 (27), 3485-3494

447 Walch, L. et al. (1999) Prostanoid receptors involved in the relaxation of human pulmonary vessels. *Br J Pharmacol* 126 (4), 859-866

448 Parkinson, P.A. et al. (2000) Phospholipase C activation by prostacyclin receptor agonist in cerebral microvascular smooth muscle cells. *Proc Soc Exp Biol Med* 223 (1), 53-58

449 Ashrafpour, H. et al. (2004) Vasodilator effect and mechanism of action of vascular endothelial growth factor in skin vasculature. *Am J Physiol Heart Circ Physiol* 286 (3), H946-954

450 Della Bella, S. et al. (2001) Novel mode of action of iloprost: in vitro down-regulation of endothelial cell adhesion molecules. *Prostaglandins* 65 (2-3), 73-83

451 Ghofrani, H.A. et al. (2002) Combination therapy with oral sildenafil and inhaled iloprost for severe pulmonary hypertension. *Ann Intern Med* 136 (7), 515-522

452 Mullershausen, F. et al. (2006) Desensitization of NO/cGMP signaling in smooth muscle: blood vessels versus airways. *Mol Pharmacol* 69 (6), 1969-1974

453 Yamaki, F. et al. (2001) MaxiK channel-mediated relaxation of guinea-pig aorta following stimulation of IP receptor with beraprost via cyclic AMP-dependent and -independent mechanisms. *Naunyn Schmiedebergs Arch Pharmacol* 364 (6), 538-550

454 Nelson, L.E. et al. (2003) The alpha2-adrenoceptor agonist dexmedetomidine converges on an endogenous sleep-promoting pathway to exert its sedative effects. *Anesthesiology* 98 (2), 428-436

455 Davis, M. et al. (1989) Spinal vs. supraspinal sites of action of the alpha 2-adrenergic agonists clonidine and ST-91 on the acoustic startle reflex. *Pharmacol Biochem Behav* 33 (1), 233-240

456 Philipp, M. et al. (2002) Physiological significance of alpha(2)-adrenergic receptor subtype diversity: one receptor is not enough. *Am J Physiol Regul Integr Comp Physiol* 283 (2), R287-295

457 Gan, L. et al. (2003) The immunosuppressive agent mizoribine monophosphate forms a transition state analogue complex with inosine monophosphate dehydrogenase. *Biochemistry* 42 (4), 857-863

458 Shimmura, H. et al. (2006) Combination effect of mycophenolate mofetil with mizoribine on cell proliferation assays and in a mouse heart transplantation model. *Transplantation* 82 (2), 175-179

459 Jordan, M.A. (2002) Mechanism of action of antitumor drugs that interact with microtubules and tubulin. *Curr Med Chem Anticancer Agents* 2 (1), 1-17

460 Madiraju, C. et al. (2005) Tubulin assembly, taxoid site binding, and cellular effects of the microtubule-stabilizing agent dictyostatin. *Biochemistry* 44 (45), 15053-15063

461 Honore, S. et al. (2004) Synergistic suppression of microtubule dynamics by discodermolide and paclitaxel in non-small cell lung carcinoma cells. *Cancer Res* 64 (14), 4957-4964

462 Black, D.M. (2003) The development of combination drugs for atherosclerosis. *Curr Atheroscler Rep* 5 (1), 29-32

463 Mondimore, F.M. et al. (2003) Drug combinations for mania. *J Clin Psychiatry* 64 Suppl 5, 25-31

464 Curatolo, M. and Sveticic, G. (2002) Drug combinations in pain treatment: a review of the published evidence and a method for finding the optimal combination. *Best Pract Res Clin Anaesthesiol* 16 (4), 507-519

465 Wilmes, A. et al. (2007) Peloruside A synergizes with other microtubule stabilizing agents in cultured cancer cell lines. *Mol Pharm* 4 (2), 269-280

466 Guignard, B. et al. (2005) Beta-lactams against methicillin-resistant Staphylococcus aureus. *Curr Opin Pharmacol* 5 (5), 479-489

467 Braga, P.C. et al. (2002) Daptomycin morphostructural damage in Bacillus cereus visualized by atomic force microscopy. *J Chemother* 14 (4), 336-341

468 Paul, T.R. et al. (1995) Localization of penicillin-binding proteins to the splitting system of Staphylococcus aureus septa by using a mercury-penicillin V derivative. *J Bacteriol* 177 (13), 3631-3640

469 Nishikawa, K. (1998) Angiotensin AT1 receptor antagonism and protection against cardiovascular end-organ damage. *J Hum Hypertens* 12 (5), 301-309

470 Rokoss, M.J. and Teo, K.K. (2005) Ramipril in the treatment of vascular diseases. *Expert Opin Pharmacother* 6 (11), 1911-1919

471 Carlsson, L. and Abrahamsson, T. (1989) Ramiprilat attenuates the local release of noradrenaline in the ischemic myocardium. *Eur J Pharmacol* 166 (2), 157-164

472 Raasch, W. et al. (2004) Combined blockade of AT1-receptors and ACE synergistically potentiates antihypertensive effects in SHR. *J Hypertens* 22 (3), 611-618

473 Touma, S.E. et al. (2005) Retinoic acid and the histone deacetylase inhibitor trichostatin a inhibit the proliferation of human renal cell carcinoma in a xenograft tumor model. *Clinical Cancer Research* 11 (9), 3558-3566

474 Meco, D. et al. (2003) Effective combination of ET-743 and doxorubicin in sarcoma: preclinical studies. *Cancer Chemotherapy and Pharmacology* 52 (2), 131-138

475 Kellogg, G.E. et al. (1998) Identification and hydropathic characterization of structural features affecting sequence specificity for doxorubicin intercalation into DNA double-stranded polynucleotides. *Nucleic Acids Res* 26 (20), 4721-4732

476 Zewail-Foote, M. et al. (2001) The inefficiency of incisions of ecteinascidin 743-DNA adducts by the UvrABC nuclease and the unique structural feature of the DNA adducts can be used to explain the repair-dependent toxicities of this antitumor agent. *Chem Biol* 8 (11), 1033-1049

477 Baruah, H. et al. (2004) Platinum-intercalator conjugates: from DNA-targeted cisplatin derivatives to adenine binding complexes as potential modulators of gene regulation. *Curr Top Med Chem* 4 (15), 1537-1549

478 Nickels, T.J. et al. (2006) Effect of theophylline and aminophylline on transmitter release at the mammalian neuromuscular junction is not mediated by cAMP. *Clin Exp Pharmacol Physiol* 33 (5-6), 465-470

479 Barrington, W.W. et al. (1989) Demonstration of distinct agonist and antagonist conformations of the A1 adenosine receptor. *J Biol Chem* 264 (22), 13157-13164

480 Yao, Q. et al. (2007) Synergism between etoposide and 17-AAG in leukemia cells: critical roles for Hsp90, FLT3, topoisomerase II, Chk1, and Rad51. *Clin Cancer Res* 13 (5), 1591-1600

481 Alves, D.P. et al. (2004) Additive antinociceptive effect of the combination of diazoxide, an activator of ATP-sensitive K+ channels, and sodium nitroprusside and dibutyryl-cGMP. *Eur J Pharmacol* 489 (1-2), 59-65

482 Russ, U. et al. (2003) Binding and effect of K ATP channel openers in the absence

of Mg2+. *Br J Pharmacol* 139 (2), 368-380

483 Soares, A.C. and Duarte, I.D. (2001) Dibutyryl-cyclic GMP induces peripheral antinociception via activation of ATP-sensitive K(+) channels in the rat PGE2-induced hyperalgesic paw. *Br J Pharmacol* 134 (1), 127-131

484 Deka, D.K. and Brading, A.F. (2004) Nitric oxide activates glibenclamide-sensitive K+ channels in urinary bladder myocytes through a c-GMP-dependent mechanism. *Eur J Pharmacol* 492 (1), 13-19

485 Alves, D.S. et al. (2004) Membrane-related effects underlying the biological activity of the anthraquinones emodin and barbaloin. *Biochem Pharmacol* 68 (3), 549-561

486 Campagna-Slater, V. and Weaver, D.F. (2007) Anaesthetic binding sites for etomidate and propofol on a GABA(A) receptor model. *Neurosci Lett* 418 (1), 28-33

487 Nishikawa, K. and Harrison, N.L. (2003) The actions of sevoflurane and desflurane on the gamma-aminobutyric acid receptor type A: effects of TM2 mutations in the alpha and beta subunits. *Anesthesiology* 99 (3), 678-684

488 Harris, R.S. et al. (2006) Interaction of propofol and sevoflurane on loss of consciousness and movement to skin incision during general anesthesia. *Anesthesiology* 104 (6), 1170-1175

489 Sigel, E. (2002) Mapping of the benzodiazepine recognition site on GABA(A) receptors. *Curr Top Med Chem* 2 (8), 833-839

490 Ono, S. et al. (2005) Mechanisms of resistance to imipenem and ampicillin in Enterococcus faecalis. *Antimicrob Agents Chemother* 49 (7), 2954-2958

491 Fuda, C. et al. (2004) The basis for resistance to beta-lactam antibiotics by penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus. *J Biol Chem* 279 (39), 40802-40806

492 Krishna, S. et al. (2006) Re-evaluation of how artemisinins work in light of emerging evidence of in vitro resistance. *Trends Mol Med* 12 (5), 200-205

493 Nandakumar, D.N. et al. (2006) Curcumin-artemisinin combination therapy for malaria. *Antimicrob Agents Chemother* 50 (5), 1859-1860

494 Furuya, R. et al. (2006) In vitro synergistic effects of double combinations of beta-lactams and azithromycin against clinical isolates of Neisseria gonorrhoeae. *J Infect Chemother* 12 (4), 172-176

495 Huang, W. et al. (2002) Ion channel behavior of amphotericin B in sterol-free and cholesterol- or ergosterol-containing supported phosphatidylcholine bilayer model membranes investigated by electrochemistry and spectroscopy. *Biophys J* 83 (6), 3245-3255

496 Walsh, T.J. et al. (2000) New targets and delivery systems for antifungal therapy. *Med Mycol* 38 Suppl 1, 335-347

497 Meletiadis, J. et al. (2006) Triazole-polyene antagonism in experimental invasive pulmonary aspergillosis: in vitro and in vivo correlation. *J Infect Dis* 194 (7), 1008-1018

498 Carrillo-Munoz, A.J. et al. (2006) Antifungal agents: mode of action in yeast cells. *Rev Esp Quimioter* 19 (2), 130-139

499 Thanou, M. et al. (2001) Oral drug absorption enhancement by chitosan and its derivatives. *Advanced drug delivery reviews* 52 (2), 117-126

500 Ciccolini, J. et al. (2000) Enhanced antitumor activity of 5-fluorouracil in combination with 2'-deoxyinosine in human colorectal cell lines and human colon tumor xenografts. *Clinical Cancer Research* 6 (4), 1529-1535

501 Glazer, R.I. and Lloyd, L.S. (1982) Association of cell lethality with incorporation of 5-fluorouracil and 5-fluorouridine into nuclear RNA in human colon carcinoma cells in culture. *Mol Pharmacol* 21 (2), 468-473

502 Narishetty, S.T. and Panchagnula, R. (2005) Effect of L-menthol and 1,8-cineole on phase behavior and molecular organization of SC lipids and skin permeation of zidovudine. *J Control Release* 102 (1), 59-70

503 Narishetty, S.T. and Panchagnula, R. (2004) Transdermal delivery of zidovudine: effect of terpenes and their mechanism of action. *J Control Release* 95 (3), 367-379

504 Shitara, Y. et al. (2004) Gemfibrozil and its glucuronide inhibit the organic anion transporting polypeptide 2 (OATP2/OATP1B1:SLC21A6)-mediated hepatic uptake and CYP2C8-mediated metabolism of cerivastatin: analysis of the mechanism of the clinically relevant drug-drug interaction between cerivastatin and gemfibrozil. *J Pharmacol Exp Ther* 311 (1), 228-236

505 Fujino, H. et al. (2003) Studies on the interaction between fibrates and statins using human hepatic microsomes. *Arzneimittelforschung* 53 (10), 701-707

506 Prueksaritanont, T. et al. (2002) Effects of fibrates on metabolism of statins in human hepatocytes. *Drug Metab Dispos* 30 (11), 1280-1287

507 Minotti, G. et al. (2001) Paclitaxel and docetaxel enhance the metabolism of doxorubicin to toxic species in human myocardium. *Clin Cancer Res* 7 (6), 1511-1515

508 Menez, C. et al. (2007) Physicochemical characterization of molecular assemblies of miltefosine and amphotericin B. *Mol Pharm* 4 (2), 281-288

509 Menez, C. et al. (2006) Interaction between miltefosine and amphotericin B: consequences for their activities towards intestinal epithelial cells and Leishmania donovani promastigotes in vitro. *Antimicrob Agents Chemother* 50 (11), 3793-3800

510 Siddik, Z.H. (2003) Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene* 22 (47), 7265-7279

511 Zicca, A. et al. (2002) Reduction of cisplatin hepatotoxicity by procainamide hydrochloride in rats. *Eur J Pharmacol* 442 (3), 265-272

512 Kaminsky, L.S. and Zhang, Z.Y. (1997) Human P450 metabolism of warfarin. *Pharmacol Ther* 73 (1), 67-74

513 Ngui, J.S. et al. (2001) In vitro stimulation of warfarin metabolism by quinidine: increases in the formation of 4'- and 10-hydroxywarfarin. *Drug Metab Dispos* 29 (6), 877-886

514 Matsuura, M. et al. (1980) Combined antibacterial activity of amoxicillin with clavulanic acid against ampicillin-resistant strains. *Antimicrob Agents Chemother* 17 (6), 908-911

515 Nasher, M.A. and Hay, R.J. (1998) Synergy of antibiotics against Streptomyces somaliensis isolates in vitro. *J Antimicrob Chemother* 41 (2), 281-284

516 Cohen, S.G. and Criep, L.H. (1949) Observations on the symptomatic treatment of chronic vascular headache with cafergone (ergotamine tartrate and caffeine). *N Engl J Med* 241 (23), 896-900

517 Stein, E.A. et al. (1996) Efficacy and Tolerability of Low-dose Simvastatin and Niacin, Alone and in Combination, in Patients With Combined Hyperlipidemia: A Prospective Trial. *J Cardiovasc Pharmacol Ther* 1 (2), 107-116

518 Loehrer, P.J., Sr. et al. (1988) Cisplatin plus etoposide in small cell lung cancer. *Semin Oncol* 15 (3 Suppl 3), 2-8

519 Vendsborg, P.B. and Bach-Mortensen, N. (1977) Fat cell size and blood lactate in humans. *Scand J Clin Lab Invest* 37 (4), 317-320

520 Fletcher, D. et al. (1997) Isobolographic analysis of interactions between intravenous morphine, propacetamol, and diclofenac in carrageenin-injected rats. *Anesthesiology* 87 (2), 317-326

521 Pace, E. et al. (2004) Synergistic effects of fluticasone propionate and salmeterol on in vitro T-cell activation and apoptosis in asthma. *J Allergy Clin Immunol* 114 (5), 1216-1223

522 Greenwood, D. and O'Grady, F. (1976) Activity and interaction of trimethoprim and sulphamethoxazole against Escherichia coli. *J Clin Pathol* 29 (2), 162-166

523 Rolinson, G.N. (1980) Effect of beta-lactam antibiotics on bacterial cell growth rate. *J Gen Microbiol* 120 (2), 317-323

524 Cole, M. (1982) Biochemistry and action of clavulanic acid. *Scott Med J* 27 Spec No., S10-16

525 Brogden, R.N. et al. (1981) Amoxycillin/clavulanic acid: a review of its antibacterial activity, pharmacokinetics and therapeutic use. *Drugs* 22 (5), 337-362

526 Nials, A.T. et al. (1993) Investigations into factors determining the duration of action of the beta 2-adrenoceptor agonist, salmeterol. *Br J Pharmacol* 108 (2), 507-515

527 Mamani-Matsuda, M. et al. (2004) Long-acting beta2-adrenergic formoterol and salmeterol induce the apoptosis of B-chronic lymphocytic leukaemia cells. *Br J Haematol* 124 (2), 141-150

528 Meltzer, E.O. (1997) The pharmacological basis for the treatment of perennial allergic rhinitis and non-allergic rhinitis with topical corticosteroids. *Allergy* 52 (36 Suppl), 33-40

529 Zhang, X. et al. (2000) Enhancement of human eosinophil apoptosis by fluticasone propionate, budesonide, and beclomethasone. *Eur J Pharmacol* 406 (3), 325-332

530 Barnes, P.J. (2002) Scientific rationale for inhaled combination therapy with long-acting beta2-agonists and corticosteroids. *Eur Respir J* 19 (1), 182-191

531 Voeller, D. et al. (1994) Interaction of Pneumocystis carinii dihydropteroate synthase with sulfonamides and diaminodiphenyl sulfone (dapsone). *J Infect Dis* 169 (2), 456-459

532 Meekins, C.V. et al. (1994) Immunochemical analysis of sulfonamide drug allergy: identification of sulfamethoxazole-substituted human serum proteins. *J Allergy Clin Immunol* 94 (6 Pt 1), 1017-1024

533 Brumfitt, W. and Hamilton-Miller, J.M. (1993) Reassessment of the rationale for the combinations of sulphonamides with diaminopyrimidines. *J Chemother* 5 (6), 465-469

534 Lowe, P.A. and Malcolm, A.D. (1976) Rifampicin binding as a probe for subunit interactions in Escherchia coli RNA polymerase. *Biochim Biophys Acta* 454 (1), 129-137

535 Lee-Huang, S. et al. (1974) Inhibition of polypeptide chain initiation in Escherichia coli by elongation factor G. *Proc Natl Acad Sci U S A* 71 (8), 2928-2931

536 Biebricher, C.K. and Druminski, M. (1980) Inhibition of RNA polymerase activity by the Escherichia coli protein biosynthesis elongation factor Ts. *Proc Natl Acad Sci U S A* 77 (2), 866-869

537 Dinos, G.P. et al. (2003) Erythromycin, roxithromycin, and clarithromycin: use of slow-binding kinetics to compare their in vitro interaction with a bacterial ribosomal complex active in peptide bond formation. *Mol Pharmacol* 63 (3), 617-623

538 Rojo, F. et al. (1984) Analysis of the different molecular forms of penicillin-

binding protein 1B in Escherichia coli ponB mutants lysogenized with specialized transducing lambda (ponB+) bacteriophages. *Eur J Biochem* 144 (3), 571-576

539 Villalon, C.M. et al. (1999) Canine external carotid vasoconstriction to methysergide, ergotamine and dihydroergotamine: role of 5-HT1B/1D receptors and alpha2-adrenoceptors. *Br J Pharmacol* 126 (3), 585-594

540 Badia, A. et al. (1988) Effects of ergotamine on cardiovascular catecholamine receptors in the pithed rat. *Gen Pharmacol* 19 (3), 475-481

541 Boulenger, J.P. et al. (1982) Effects of caffeine and theophylline on adenosine and benzodiazepine receptors in human brain. *Neurosci Lett* 30 (2), 161-166

542 Mukhopadhyay, S. and Poddar, M.K. (1995) Caffeine-induced locomotor activity: possible involvement of GABAergic-dopaminergic-adenosinergic interaction. *Neurochem Res* 20 (1), 39-44

543 Levin, R.M. et al. (1981) Quantitative analysis of the effects of caffeine on sperm motility and cyclic adenosine 3',5'-monophosphate (AMP) phosphodiesterase. *Fertil Steril* 36 (6), 798-802

544 Anderson, J.R. et al. (1981) Effect of caffeine on ergotamine absorption from rat small intestine. *J Pharm Sci* 70 (6), 651-657

545 Ganji, S.H. et al. (2003) Niacin and cholesterol: role in cardiovascular disease (review). *J Nutr Biochem* 14 (6), 298-305

546 Plosker, G.L. and McTavish, D. (1995) Simvastatin. A reappraisal of its pharmacology and therapeutic efficacy in hypercholesterolaemia. *Drugs* 50 (2), 334-363

547 Maccubbin, A.E. et al. (1991) A cyclophosphamide/DNA phosphoester adduct formed in vitro and in vivo. *Cancer Res* 51 (3), 886-892

548 Mehta, J.R. et al. (1980) Alkylation of guanosine and deoxyguanosine by phosphoramide mustard. *Cancer Res* 40 (11), 4183-4186

549 Haller, D.G. (1997) Trimetrexate: experience with solid tumors. *Semin Oncol* 24 (5 Suppl 18), S18-71-S18-76

550 Pinedo, H.M. and Peters, G.F. (1988) Fluorouracil: biochemistry and pharmacology. *J Clin Oncol* 6 (10), 1653-1664

551 Sun, X.X. et al. (2007) 5-fluorouracil activation of p53 involves an MDM2-ribosomal protein interaction. *J Biol Chem* 282 (11), 8052-8059

552 Fernandes, D.J. and Bertino, J.R. (1980) 5-fluorouracil-methotrexate synergy: enhancement of 5-fluorodeoxyridylate binding to thymidylate synthase by dihydropteroylpolyglutamates. *Proc Natl Acad Sci U S A* 77 (10), 5663-5667

553 Schwieler, L. et al. (2005) Prostaglandin-mediated control of rat brain kynurenic acid synthesis--opposite actions by COX-1 and COX-2 isoforms. *J Neural Transm* 112 (7), 863-872

554 Bertolini, A. et al. (2006) Paracetamol: new vistas of an old drug. *CNS Drug Rev* 12 (3-4), 250-275

555 Ouellet, M. and Percival, M.D. (2001) Mechanism of acetaminophen inhibition of cyclooxygenase isoforms. *Arch Biochem Biophys* 387 (2), 273-280

556 Hinz, B. et al. (2008) Acetaminophen (paracetamol) is a selective cyclooxygenase-2 inhibitor in man. *FASEB J* 22 (2), 383-390

557 Ganter, B. and Giroux, C.N. (2008) Emerging applications of network and pathway analysis in drug discovery and development. *Curr Opin Drug Discov Devel* 11 (1), 86-94

558 Eckstein, N. et al. (2008) Epidermal growth factor receptor pathway analysis identifies amphiregulin as a key factor for cisplatin resistance of human breast

cancer cells. *J Biol Chem* 283 (2), 739-750

559 Ganter, B. et al. (2008) Pathway analysis tools and toxicogenomics reference databases for risk assessment. *Pharmacogenomics* 9 (1), 35-54

560 Apic, G. et al. (2005) Illuminating drug discovery with biological pathways. *FEBS Lett* 579 (8), 1872-1877

561 Davidov, E. et al. (2003) Advancing drug discovery through systems biology. *Drug Discov Today* 8 (4), 175-183

562 Huang, S. (2002) Rational drug discovery: what can we learn from regulatory networks? *Drug Discov Today* 7 (20 Suppl), S163-169

563 Nahta, R. and Esteva, F.J. (2007) Trastuzumab: triumphs and tribulations. *Oncogene* 26 (25), 3637-3643

564 Pietras, R.J. et al. (1998) Remission of human breast cancer xenografts on therapy with humanized monoclonal antibody to HER-2 receptor and DNA-reactive drugs. *Oncogene* 17 (17), 2235-2249

565 Le, X.F. et al. (2005) Genes affecting the cell cycle, growth, maintenance, and drug sensitivity are preferentially regulated by anti-HER2 antibody through phosphatidylinositol 3-kinase-AKT signaling. *J Biol Chem* 280 (3), 2092-2104

566 Lee, S. et al. (2002) Enhanced sensitization to taxol-induced apoptosis by herceptin pretreatment in ErbB2-overexpressing breast cancer cells. *Cancer Res* 62 (20), 5703-5710

567 Bild, A.H. et al. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439 (7074), 353-357

568 Cheok, M.H. and Evans, W.E. (2006) Acute lymphoblastic leukaemia: a model for the pharmacogenomics of cancer therapy. *Nat Rev Cancer* 6 (2), 117-129

569 Lee, J.K. et al. (2007) A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. *Proc Natl Acad Sci U S A* 104 (32), 13086-13091

570 Gerhold, D.L. et al. (2002) Better therapeutics through microarrays. *Nat Genet* 32 Suppl, 547-551

571 Rickardson, L. et al. (2005) Identification of molecular mechanisms for cellular drug resistance by combining drug activity and gene expression profiles. *Br J Cancer* 93 (4), 483-492

572 den Boer, M.L. and Pieters, R. (2007) Microarray-based identification of new targets for specific therapies in pediatric leukemia. *Curr Drug Targets* 8 (6), 761-764

573 Wirth, G.J. et al. (2006) Microarrays of 41 human tumor cell lines for the characterization of new molecular targets: expression patterns of cathepsin B and the transferrin receptor. *Oncology* 71 (1-2), 86-94

574 Andre, F. et al. (2006) DNA arrays as predictors of efficacy of adjuvant/neoadjuvant chemotherapy in breast cancer patients: current data and issues on study design. *Biochim Biophys Acta* 1766 (2), 197-204

575 Ho, C. and Laskin, J. (2009) EGFR-directed therapies to treat non-small-cell lung cancer. *Expert Opin Investig Drugs* 18 (8), 1133-1145

576 Sharma, S.V. et al. (2007) Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 7 (3), 169-181

577 Stemke-Hale, K. et al. (2008) An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res* 68 (15), 6084-6091

578 Linardou, H. et al. (2009) Somatic EGFR mutations and efficacy of tyrosine kinase

inhibitors in NSCLC. *Nat Rev Clin Oncol* 6 (6), 352-366

579 Chen, F.L. et al. (2008) Acquired resistance to small molecule ErbB2 tyrosine kinase inhibitors. *Clin Cancer Res* 14 (21), 6730-6734

580 Nahta, R. et al. (2006) Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer. *Nat Clin Pract Oncol* 3 (5), 269-280

581 Cameron, D.A. and Stein, S. (2008) Drug Insight: intracellular inhibitors of HER2--clinical development of lapatinib in breast cancer. *Nat Clin Pract Oncol* 5 (9), 512-520

582 Knight, Z.A. et al. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer* 10 (2), 130-137

583 Oprea, T.I. et al. (2009) A crowdsourcing evaluation of the NIH chemical probes. *Nat Chem Biol* 5 (7), 441-447

584 Sartore-Bianchi, A. et al. (2009) PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res* 69 (5), 1851-1857

585 Laurent-Puig, P. et al. (2009) Mutations and response to epidermal growth factor receptor inhibitors. *Clin Cancer Res* 15 (4), 1133-1139

586 Xia, W. et al. (2006) A model of acquired autoresistance to a potent ErbB2 tyrosine kinase inhibitor and a therapeutic strategy to prevent its onset in breast cancer. *Proc Natl Acad Sci U S A* 103 (20), 7795-7800

587 Liu, L. et al. (2009) Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. *Cancer Res* 69 (17), 6871-6878

588 Eichhorn, P.J. et al. (2008) Phosphatidylinositol 3-kinase hyperactivation results in lapatinib resistance that is reversed by the mTOR/phosphatidylinositol 3-kinase inhibitor NVP-BEZ235. *Cancer Res* 68 (22), 9221-9230

589 Zhou, B.B. et al. (2006) Targeting ADAM-mediated ligand cleavage to inhibit HER3 and EGFR pathways in non-small cell lung cancer. *Cancer Cell* 10 (1), 39-50

590 Sos, M.L. et al. (2009) Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. *J Clin Invest* 119 (6), 1727-1740

591 Forbes, S.A. et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* 38 (Database issue), D652-657

592 Thomas, R.K. et al. (2007) High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 39 (3), 347-351

593 Amann, J. et al. (2005) Aberrant epidermal growth factor receptor signaling and enhanced sensitivity to EGFR inhibitors in lung cancer. *Cancer Res* 65 (1), 226-235

594 Gandhi, J. et al. (2009) Alterations in genes of the EGFR signaling pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in lung cancer cell lines. *PLoS One* 4 (2), e4576

595 Kao, J. et al. (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* 4 (7), e6146

596 Neve, R.M. et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 10 (6), 515-527

597 Hoeflich, K.P. et al. (2009) In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. *Clin Cancer Res* 15 (14), 4649-4664

598 Finn, R.S. et al. (2009) PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Res* 11 (5), R77

599 Junttila, T.T. et al. (2009) Ligand-independent HER2/HER3/PI3K complex is disrupted by trastuzumab and is effectively inhibited by the PI3K inhibitor GDC-0941. *Cancer Cell* 15 (5), 429-440

600 Kataoka, Y. et al. Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines. *Ann Oncol* 21 (2), 255-262

601 Kawa, S. et al. (2009) A novel mouse monoclonal antibody targeting ErbB2 suppresses breast cancer growth. *Biochem Biophys Res Commun* 384 (3), 329-333

602 Konecny, G.E. et al. (2006) Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res* 66 (3), 1630-1639

603 Koninki, K. et al. Multiple molecular mechanisms underlying trastuzumab and lapatinib resistance in JIMT-1 breast cancer cells. *Cancer Lett*

604 Rusnak, D.W. et al. (2007) Assessment of epidermal growth factor receptor (EGFR, ErbB1) and HER2 (ErbB2) protein expression levels and response to lapatinib (Tykerb, GW572016) in an expanded panel of human normal and tumour cell lines. *Cell Prolif* 40 (4), 580-594

605 Kadara, H. et al. (2009) Identification of gene signatures and molecular markers for human lung cancer prognosis using an in vitro lung carcinogenesis system. *Cancer Prev Res (Phila Pa)* 2 (8), 702-711

606 Ince, T.A. et al. (2007) Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* 12 (2), 160-170

607 Eckford, P.D. and Sharom, F.J. (2009) ABC efflux pump-based resistance to chemotherapy drugs. *Chem Rev* 109 (7), 2989-3011

608 Kitazaki, T. et al. (2005) Gefitinib, an EGFR tyrosine kinase inhibitor, directly inhibits the function of P-glycoprotein in multidrug resistant cancer cells. *Lung Cancer* 49 (3), 337-343

609 Noguchi, K. et al. (2009) Substrate-dependent bidirectional modulation of P-glycoprotein-mediated drug resistance by erlotinib. *Cancer Sci* 100 (9), 1701-1707

610 Polli, J.W. et al. (2008) The role of efflux and uptake transporters in [N-{3-chloro-4-[(3-fluorobenzyl)oxy]phenyl}-6-[5-({[2-(methylsulfonyl)ethy l]amino}methyl)-2-furyl]-4-quinazolinamine (GW572016, lapatinib) disposition and drug interactions. *Drug Metab Dispos* 36 (4), 695-701

611 Imai, K. and Takaoka, A. (2006) Comparing antibody and small-molecule therapies for cancer. *Nat Rev Cancer* 6 (9), 714-727

612 Pollack, J.R. et al. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* 99 (20), 12963-12968

613 Perren, A. et al. (1999) Immunohistochemical evidence of loss of PTEN expression in primary ductal adenocarcinomas of the breast. *Am J Pathol* 155 (4), 1253-1260

614 Li, J. et al. (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 275 (5308), 1943-1947

615 Beroukhim, R. et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104 (50), 20007-20012

616 Redon, R. et al. (2006) Global variation in copy number in the human genome.

Nature 444 (7118), 444-454

617 Rennstam, K. et al. (2007) Cytogenetic characterization and gene expression profiling of the trastuzumab-resistant breast cancer cell line JIMT-1. *Cancer Genet Cytogenet* 172 (2), 95-106

618 Draghici, S. (2002) Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov Today* 7 (11 Suppl), S55-63

619 Uramoto, H. and Mitsudomi, T. (2007) Which biomarker predicts benefit from EGFR-TKI treatment for patients with lung cancer? *Br J Cancer* 96 (6), 857-863

620 Laurent-Puig, P. et al. (2009) Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J Clin Oncol* 27 (35), 5924-5930

621 Kadota, M. et al. (2009) Identification of novel gene amplifications in breast cancer and coexistence of gene amplification with an activating mutation of PIK3CA. *Cancer Res* 69 (18), 7357-7365

622 Shendure, J. and Ji, H. (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26 (10), 1135-1145

623 Pariset, L. et al. (2009) Microarrays and high-throughput transcriptomic analysis in species with incomplete availability of genomic sequences. *N Biotechnol* 25 (5), 272-279

624 Cooper, G.M. et al. (2008) Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat Genet* 40 (10), 1199-1203

625 Finn, O.J. (2003) Cancer vaccines: between the idea and the reality. *Nat Rev Immunol* 3 (8), 630-641

626 Pardoll, D. (2003) Does the immune system see tumors as foreign or self? *Annu Rev Immunol* 21, 807-839

627 Viola, A. and Lanzavecchia, A. (1996) T cell activation determined by T cell receptor number and tunable thresholds. *Science* 273 (5271), 104-106

628 Francis, T. (2003) Mechanisms of Tumor Escape from the Immune Response. *Routledge,USA*

629 Singh, S. et al. (1992) Stroma is critical for preventing or permitting immunological destruction of antigenic cancer cells. *J Exp Med* 175 (1), 139-146

630 Dunn, G.P. et al. (2004) The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* 21 (2), 137-148

631 Zitvogel, L. et al. (2006) Cancer despite immunosurveillance: immunoselection and immunosubversion. *Nat Rev Immunol* 6 (10), 715-727

632 Greiner, J. et al. (2000) Simultaneous expression of different immunogenic antigens in acute myeloid leukemia. *Exp Hematol* 28 (12), 1413-1422

633 Forrest, W.F. and Cavet, G. (2007) Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science* 317 (5844), 1500; author reply 1500

634 Rubin, A.F. and Green, P. (2007) Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science* 317 (5844), 1500

635 Wood, L.D. et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318 (5853), 1108-1113

636 Berke, G. (1994) The binding and lysis of target cells by cytotoxic lymphocytes: molecular and cellular aspects. *Annu Rev Immunol* 12, 735-773

637 Rammensee, H.G. et al. (1993) Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol* 11, 213-244

638 Zinkernagel, R.M. and Doherty, P.C. (1997) The discovery of MHC restriction.

Immunol Today 18 (1), 14-17

639 York, I.A. and Rock, K.L. (1996) Antigen processing and presentation by the class I major histocompatibility complex. *Annu Rev Immunol* 14, 369-396

640 Kawashima, I. et al. (1999) Identification of HLA-A3-restricted cytotoxic T lymphocyte epitopes from carcinoembryonic antigen and HER-2/neu by primary in vitro immunization with peptide-pulsed dendritic cells. *Cancer Res* 59 (2), 431-435

641 Kesmir, C. et al. (2002) Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 15 (4), 287-296

642 Muthusamy, V. et al. (2006) Amplification of CDK4 and MDM2 in malignant melanoma. *Genes Chromosomes Cancer* 45 (5), 447-454

643 Okamoto, I. et al. (2005) Seven novel and stable translocations associated with oncogenic gene expression in malignant melanoma. *Neoplasia* 7 (4), 303-311

644 Stecca, B. et al. (2007) Melanomas require HEDGEHOG-GLI signaling regulated by interactions between GLI1 and the RAS-MEK/AKT pathways. *Proc Natl Acad Sci U S A* 104 (14), 5895-5900

645 Ein-Dor, L. et al. (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* 103 (15), 5923-5928

646 Cruz, J.A. and Wishart, D.S. (2006) Applications of Machine Learning in Cancer Predictionand Prognosis. *Cancer Informatics* 2, 59-78

647 Salomon, D.S. et al. (1995) Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit Rev Oncol Hematol* 19 (3), 183-232

648 Ochs, J.S. (2004) Rationale and clinical basis for combining gefitinib (IRESSA, ZD1839) with radiation therapy for solid tumors. *Int J Radiat Oncol Biol Phys* 58 (3), 941-949

649 Jemal, A. et al. (2008) Cancer statistics, 2008. *CA Cancer J Clin* 58 (2), 71-96

650 Jemal, A. et al. (2007) Cancer statistics, 2007. *CA Cancer J Clin* 57 (1), 43-66

651 http://www.merck.com/mmpe/sec05/ch062/ch062b.html#sec05-ch062-ch062b-1405. Lung Carcinoma: Tumors of the Lungs, Online edition. Merck Manual Professional Edition

652 Herbst, R.S. et al. (2008) Lung cancer. *N Engl J Med* 359 (13), 1367-1380

653 Huber, R.M. and Stratakis, D.F. (2004) Molecular oncology--perspectives in lung cancer. *Lung Cancer* 45 Suppl 2, S209-213

654 Sorensen, J.B. et al. (1993) Interobserver variability in histopathologic subtyping and grading of pulmonary adenocarcinoma. *Cancer* 71 (10), 2971-2976

655 Gail, M.H. et al. (1984) Prognostic factors in patients with resected stage I non-small cell lung cancer. A report from the Lung Cancer Study Group. *Cancer* 54 (9), 1802-1813

656 Takise, A. et al. (1988) Histopathologic prognostic factors in adenocarcinomas of the peripheral lung less than 2 cm in diameter. *Cancer* 61 (10), 2083-2088

657 Okada, M. et al. (1999) Evaluation of TMN classification for lung carcinoma with ipsilateral intrapulmonary metastasis. *Ann Thorac Surg* 68 (2), 326-330; discussion 331

658 Harpole, D.H., Jr. et al. (1995) A prognostic model of recurrence and death in stage I non-small cell lung cancer utilizing presentation, histopathology, and oncoprotein expression. *Cancer Res* 55 (1), 51-56

659 Sequist, L.V. et al. (2007) Molecular predictors of response to epidermal growth factor receptor antagonists in non-small-cell lung cancer. *J Clin Oncol* 25 (5), 587-595

660 Scagliotti, G.V. et al. (2008) Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* 26 (21), 3543-3551

661 Guo, L. et al. (2006) Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. *Clin Cancer Res* 12 (11 Pt 1), 3344-3354

662 Beer, D.G. et al. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8 (8), 816-824

663 Bhattacharjee, A. et al. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 98 (24), 13790-13795

664 Edgerton, E. et al. (2007) Data Mining for Gene Networks Relevant to Poor Prognosis in Lung Cancer Via Backward-Chaining Rule Induction. *Cancer Informatics* 2, 93-114

665 Lu, Y. et al. (2006) A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med* 3 (12), e467

666 Chen, H.Y. et al. (2007) A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356 (1), 11-20

667 Xu, J. et al. (2005) Survival analysis of microarray expression data by transformation models. *Comput Biol Chem* 29 (2), 91-94

668 http://www.genome.wi.mit.edu/MPR/lung.

669 http://www.xlstat.com/en/support/tutorials/km.htm.

670 Joo, Y.E. et al. (2002) Expression of vascular endothelial growth factor and p53 in pancreatic carcinomas. *Korean J Intern Med* 17 (3), 153-159

671 Strohmeyer, D. et al. (2000) Vascular endothelial growth factor and its correlation with angiogenesis and p53 expression in prostate cancer. *Prostate* 45 (3), 216-224

672 Maeda, K. et al. (1998) Expression of p53 and vascular endothelial growth factor associated with tumor angiogenesis and prognosis in gastric cancer. *Oncology* 55 (6), 594-599

673 Liu, D.H. et al. (2001) Expression of vascular endothelial growth factor and its role in oncogenesis of human gastric carcinoma. *World J Gastroenterol* 7 (4), 500-505

674 Lee, J.S. et al. (2002) Expression of vascular endothelial growth factor in adenocarcinomas of the uterine cervix and its relation to angiogenesis and p53 and c-erbB-2 protein expression. *Gynecol Oncol* 85 (3), 469-475

675 Gills, J.J. et al. (2004) Targeting aberrant signal transduction pathways in lung cancer. *Cancer Biol Ther* 3 (2), 147-155

676 Muraoka, K. et al. (1993) Enhanced expression of a tumor-cell-derived collagenase-stimulatory factor in urothelial carcinoma: its usefulness as a tumor marker for bladder cancers. *Int J Cancer* 55 (1), 19-26

677 Caudroy, S. et al. (2002) EMMPRIN-mediated MMP regulation in tumor and endothelial cells. *Clin Exp Metastasis* 19 (8), 697-702

678 Tang, Y. et al. (2005) Extracellular matrix metalloproteinase inducer stimulates tumor angiogenesis by elevating vascular endothelial cell growth factor and matrix metalloproteinases. *Cancer Res* 65 (8), 3193-3199

679 Yan, L. et al. (2005) Roles of the multifunctional glycoprotein, emmprin (basigin; CD147), in tumour progression. *Thromb Haemost* 93 (2), 199-204

680 Klein, C.A. et al. (2002) Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol* 20 (4), 387-392

681 Strieter, R.M. et al. (2004) CXC chemokines: angiogenesis, immunoangiostasis, and metastases in lung cancer. *Ann N Y Acad Sci* 1028, 351-360

682 Strieter, R.M. et al. (2005) CXC chemokines in angiogenesis relevant to chronic fibroproliferation. *Curr Drug Targets Inflamm Allergy* 4 (1), 23-26

683 Tzouvelekis, A. et al. (2006) Angiogenesis in interstitial lung diseases: a pathogenetic hallmark or a bystander? *Respir Res* 7, 82

684 Strieter, R.M. et al. (2004) *Chemokines: Angiogenesis and Metastases in Lung Cancer*, John Wiley & Sons

685 Schuller, H.M. (2007) Neurotransmitter receptor-mediated signaling pathways as modulators of carcinogenesis. *Prog Exp Tumor Res* 39, 45-63

686 Ho, Y.S. et al. (2005) Tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) induces cell proliferation in normal human bronchial epithelial cells through NFkappaB activation and cyclin D1 up-regulation. *Toxicol Appl Pharmacol* 205 (2), 133-148

687 Hakomori, S. (1989) Aberrant glycosylation in tumors and tumor-associated carbohydrate antigens. *Adv Cancer Res* 52, 257-331

688 Friederichs, J. et al. (2000) The CD24/P-selectin binding pathway initiates lung arrest of human A125 adenocarcinoma cells. *Cancer Res* 60 (23), 6714-6722

689 Martin-Satue, M. et al. (1998) Enhanced expression of alpha(1,3)-fucosyltransferase genes correlates with E-selectin-mediated adhesion and metastatic potential of human lung adenocarcinoma cells. *Cancer Res* 58 (7), 1544-1550

690 Ohyama, C. et al. (1999) Dual roles of sialyl Lewis X oligosaccharides in tumor metastasis and rejection by natural killer cells. *Embo J* 18 (6), 1516-1525

691 Grzmil, M. et al. (2004) Up-regulated expression of the MAT-8 gene in prostate cancer and its siRNA-mediated inhibition of expression induces a decrease in proliferation of human prostate carcinoma cells. *Int J Oncol* 24 (1), 97-105

692 Kayed, H. et al. (2006) FXYD3 is overexpressed in pancreatic ductal adenocarcinoma and influences pancreatic cancer cell growth. *Int J Cancer* 118 (1), 43-54

693 Kim, J.H. et al. (2006) Phospholipase D1 regulates cell migration in a lipase activity-independent manner. *J Biol Chem* 281 (23), 15747-15756

694 Zhong, M. et al. (2003) Phospholipase D prevents apoptosis in v-Src-transformed rat fibroblasts and MDA-MB-231 breast cancer cells. *Biochem Biophys Res Commun* 302 (3), 615-619

695 Kahlina, K. et al. (2004) p68 DEAD box RNA helicase expression in keratinocytes. Regulation, nucleolar localization, and functional connection to proliferation and vascular endothelial growth factor gene expression. *J Biol Chem* 279 (43), 44872-44882

696 Kvissel, A.K. et al. (2007) Androgen dependent regulation of protein kinase A subunits in prostate cancer cells. *Cell Signal* 19 (2), 401-409

697 Burns, J.M. et al. (2006) A novel chemokine receptor for SDF-1 and I-TAC involved in cell survival, cell adhesion, and tumor development. *J Exp Med* 203 (9), 2201-2213

698 Strieter, R.M. et al. (2005) CXC Chemokines in Angiogenesis Relevant to Chronic Fibroproliferation. *Current Drug Targets - Inflammation & Allergy* 4 (1), 23-26

699 Oue, N. et al. (2005) Expression and localization of Reg IV in human neoplastic and non-neoplastic tissues: Reg IV expression is associated with intestinal and neuroendocrine differentiation in gastric adenocarcinoma. *J Pathol* 207 (2), 185-198

700 Sekikawa, A. et al. (2005) Possible role of REG Ialpha protein in ulcerative colitis

and colitic cancer. *Gut* 54 (10), 1437-1444

701 Jang, C.Y. et al. (2004) RpS3, a DNA repair endonuclease and ribosomal protein, is involved in apoptosis. *FEBS Lett* 560 (1-3), 81-85

702 Kim, S.H. and Kim, J. (2006) Reduction of invasion in human fibrosarcoma cells by ribosomal protein S3 in conjunction with Nm23-H1 and ERK. *Biochim Biophys Acta* 1763 (8), 823-832

703 Robert, C. et al. (1999) Expression of plasminogen activator inhibitors 1 and 2 in lung cancer and their role in tumor progression. *Clin Cancer Res* 5 (8), 2094-2102

704 Almholt, K. et al. (2003) Metastasis of transgenic breast cancer in plasminogen activator inhibitor-1 gene-deficient mice. *Oncogene* 22 (28), 4389-4397

705 Speleman, L. et al. (2007) Prognostic value of plasminogen activator inhibitor-1 in head and neck squamous cell carcinoma. *Head Neck* 29 (4), 341-350

706 Gil-Bazo, I. et al. (2006) [New prognostic and predictive factors in advanced colorectal cancer]. *Med Clin (Barc)* 126 (14), 541-548

707 Goldman, N.A. et al. (2006) GLUT1 and GLUT8 in endometrium and endometrial adenocarcinoma. *Mod Pathol* 19 (11), 1429-1436

708 Tesfaigzi, Y. et al. (2003) SPRR1B overexpression enhances entry of cells into the G0 phase of the cell cycle. *Am J Physiol Lung Cell Mol Physiol* 285 (4), L889-898

709 Patterson, T. et al. (2001) Mechanism of repression of squamous differentiation marker, SPRR1B, in malignant bronchial epithelial cells: role of critical TRE-sites and its transacting factors. *Oncogene* 20 (5), 634-644

710 Zhong, H. and Bowen, J.P. (2006) Antiangiogenesis drug design: multiple pathways targeting tumor vasculature. *Curr Med Chem* 13 (8), 849-862

711 Kyu-Ho Han, E. et al. (2000) Modulation of drug resistance by alpha-tubulin in paclitaxel-resistant human lung cancer cell lines. *Eur J Cancer* 36 (12), 1565-1571

712 Dumontet, C. and Sikic, B.I. (1999) Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. *J Clin Oncol* 17 (3), 1061-1070

713 Kommagani, R. et al. (2006) Identification of vitamin D receptor as a target of p63. *Oncogene* 25 (26), 3745-3751

714 Chen, X. et al. (2002) TTD: Therapeutic Target Database. *Nucleic Acids Res* 30 (1), 412-415

715 Futreal, P.A. et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4 (3), 177-183

716 Vogelstein, B. and Kinzler, K.W. (2004) Cancer genes and the pathways they control. *Nat Med* 10 (8), 789-799

717 de Castro Junior, G. et al. (2006) Angiogenesis and cancer: A cross-talk between basic science and clinical trials (the "do ut des" paradigm). *Crit Rev Oncol Hematol* 59 (1), 40-50

718 Mancuso, A. and Sternberg, C.N. (2005) Colorectal cancer and antiangiogenic therapy: what can be expected in clinical practice? *Crit Rev Oncol Hematol* 55 (1), 67-81

719 Irish, J.M. et al. (2006) Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat Rev Cancer* 6 (2), 146-155

720 Muller, A.J. and Scherle, P.A. (2006) Targeting the mechanisms of tumoral immune tolerance with small-molecule inhibitors. *Nat Rev Cancer* 6 (8), 613-625

721 Ilhan, N. et al. (2004) Functional significance of vascular endothelial growth factor and its receptor (receptor-1) in various lung cancer types. *Clin Biochem* 37 (9), 840-845

722 Dudek, A.Z. and Mahaseth, H. (2005) Circulating angiogenic cytokines in patients with advanced non-small cell lung cancer: correlation with treatment response and survival. *Cancer Invest* 23 (3), 193-200

723 Kaiser, U. et al. (1996) Expression of vitamin D receptor in lung cancer. *J Cancer Res Clin Oncol* 122 (6), 356-359

724 Cooper, R. et al. (2003) Glucose transporter-1 (GLUT-1): a potential marker of prognosis in rectal carcinoma? *Br J Cancer* 89 (5), 870-876

725 Noh, D.Y. et al. (2000) Overexpression of phospholipase D1 in human breast cancer tissues. *Cancer Lett* 161 (2), 207-214

726 Zhao, Y. et al. (2000) Increased activity and intranuclear expression of phospholipase D2 in human renal cancer. *Biochem Biophys Res Commun* 278 (1), 140-143

727 Oka, M. et al. (2003) Protein kinase C alpha associates with phospholipase D1 and enhances basal phospholipase D activity in a protein phosphorylation-independent manner in human melanoma cells. *J Invest Dermatol* 121 (1), 69-76

728 Madjd, Z. et al. (2005) High expression of Lewis y/b antigens is associated with decreased survival in lymph node negative breast carcinomas. *Breast Cancer Res* 7 (5), R780-787

729 Castello, R. et al. (2006) Plasminogen activator inhibitor-1 4G/5G polymorphism in breast cancer patients and its association with tissue PAI-1 levels and tumor severity. *Thromb Res* 117 (5), 487-492

730 Bhuvarahamurthy, V. et al. (2004) In situ gene expression of urokinase-type plasminogen activator and its receptor in transitional cell carcinoma of the human bladder. *Oncol Rep* 12 (4), 909-913

731 Shetty, S. and Idell, S. (1999) Posttranscriptional regulation of urokinase receptor gene expression in human lung carcinoma and mesothelioma cells in vitro. *Mol Cell Biochem* 199 (1-2), 189-200

732 Pedersen, H. et al. (1994) Urokinase and plasminogen activator inhibitor type 1 in pulmonary adenocarcinoma. *Cancer Res* 54 (1), 120-123

733 Yonemura, Y. et al. (2003) REG gene expression is associated with the infiltrating growth of gastric carcinoma. *Cancer* 98 (7), 1394-1400

734 Dhar, D.K. et al. (2004) Expression of regenerating gene I in gastric adenocarcinomas: correlation with tumor differentiation status and patient survival. *Cancer* 100 (6), 1130-1136

735 Krug, M. and Hilgeroth, A. (2008) Recent advances in the development of multi-kinase inhibitors. *Mini Rev Med Chem* 8 (13), 1312-1327

736 Gill, A.L. et al. (2007) A comparison of physicochemical property profiles of marketed oral drugs and orally bioavailable anti-cancer protein kinase inhibitors in clinical development. *Curr Top Med Chem* 7 (14), 1408-1422

737 Shoichet, B.K. (2004) Virtual screening of chemical libraries. *Nature* 432 (7019), 862-865

738 Yamane, S. et al. (2008) Proinflammatory role of amphiregulin, an epidermal growth factor family member whose expression is augmented in rheumatoid arthritis patients. *J Inflamm (Lond)* 5, 5

739 Ma, X.H. et al. (2008) Evaluation of virtual screening performance of support vector machines trained by sparsely distributed active compounds. *J Chem Inf Model* 48 (6), 1227-1237

740 Gozalbes, R. et al. (2008) Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. *J Med Chem* 51 (11), 3124-

3132

741 Deng, X.Q. et al. (2008) Pharmacophore modelling and virtual screening for identification of new Aurora-A kinase inhibitors. *Chem Biol Drug Des* 71 (6), 533-539

742 Deanda, F. et al. (2008) Kinase-Targeted Library Design through the Application of the PharmPrint Methodology. *J Chem Inf Model* 48 (12), 2395-2403

743 Briem, H. and Gunther, J. (2005) Classifying "kinase inhibitor-likeness" by using machine-learning methods. *Chembiochem* 6 (3), 558-566

744 Gundla, R. et al. (2008) Discovery of novel small-molecule inhibitors of human epidermal growth factor receptor-2: combined ligand and target-based approach. *J Med Chem* 51 (12), 3367-3377

745 Prado-Prado, F.J. et al. (2008) Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg Med Chem*

746 Zhang, X. and Fernandez, A. (2008) In silico drug profiling of the human kinome based on a molecular marker for cross reactivity. *Mol Pharm* 5 (5), 728-738

747 Bigioni, M. et al. (2007) Antitumour effect of combination treatment with Sabarubicin (MEN 10755) and cis-platin (DDP) in human lung tumour xenograft. *Cancer Chemother Pharmacol*

748 Robert, J. and Larsen, A.K. (1998) Drug resistance to topoisomerase II inhibitors. *Biochimie* 80 (3), 247-254

749 Hoeijmakers, J.H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411 (6835), 366-374

750 Meshnick, S.R. (2002) Artemisinin: mechanisms of action, resistance and toxicity. *Int J Parasitol* 32 (13), 1655-1660

751 Atamna, H. et al. (1996) Mode of antimalarial effect of methylene blue and some of its analogues on Plasmodium falciparum in culture and their inhibition of P. vinckei petteri and P. yoelii nigeriensis in vivo. *Biochem Pharmacol* 51 (5), 693-700

752 Farber, P.M. et al. (1998) Recombinant Plasmodium falciparum glutathione reductase is inhibited by the antimalarial dye methylene blue. *FEBS Lett* 422 (3), 311-314

753 Nascimento, C.G. and Branco, L.G. (2007) Role of the peripheral heme oxygenase-carbon monoxide pathway on the nociceptive response of rats to the formalin test: evidence for a cGMP signaling pathway. *Eur J Pharmacol* 556 (1-3), 55-61

754 Akoachere, M. et al. (2005) In vitro assessment of methylene blue on chloroquine-sensitive and -resistant Plasmodium falciparum strains reveals synergistic action with artemisinins. *Antimicrob Agents Chemother* 49 (11), 4592-4597

755 Meshnick, S.R. et al. (1993) Iron-dependent free radical generation from the antimalarial agent artemisinin (qinghaosu). *Antimicrob Agents Chemother* 37 (5), 1108-1114

756 Ginsburg, H. et al. (1998) Inhibition of glutathione-dependent degradation of heme by chloroquine and amodiaquine as a possible basis for their antimalarial mode of action. *Biochem Pharmacol* 56 (10), 1305-1313

757 Raymond, E. et al. (2000) Epidermal growth factor receptor tyrosine kinase as a target for anticancer therapy. *Drugs* 60 Suppl 1, 15-23; discussion 41-12

758 Rusthoven, J.J. et al. (1999) Multitargeted antifolate LY231514 as first-line chemotherapy for patients with advanced non-small-cell lung cancer: A phase II study. National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* 17

(4), 1194

759 Smit, E.F. et al. (2003) ALIMTA (pemetrexed disodium) as second-line treatment of non-small-cell lung cancer: a phase II study. *Ann Oncol* 14 (3), 455-460

760 Vogelzang, N.J. et al. (2003) Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma. *J Clin Oncol* 21 (14), 2636-2644

761 Giovannetti, E. et al. (2008) Molecular mechanisms underlying the synergistic interaction of erlotinib, an epidermal growth factor receptor tyrosine kinase inhibitor, with the multitargeted antifolate pemetrexed in non-small-cell lung cancer cells. *Mol Pharmacol* 73 (4), 1290-1300

762 Becker, J.C. et al. (2006) Molecularly targeted therapy for melanoma: current reality and future options. *Cancer* 107 (10), 2317-2327

763 Ishii, Y. et al. (2007) Targeting the ubiquitin-proteasome pathway in cancer therapy. *Anticancer Agents Med Chem* 7 (3), 359-365

764 Marks, P. et al. (2001) Histone deacetylases and cancer: causes and therapies. *Nat Rev Cancer* 1 (3), 194-202

765 Xu, Y. et al. (2005) The histone deacetylase inhibitor suberoylanilide hydroxamic acid down-regulates expression levels of Bcr-abl, c-Myc and HDAC3 in chronic myeloid leukemia cell lines. *Int J Mol Med* 15 (1), 169-172

766 Dai, Y. et al. (2008) Interactions between bortezomib and romidepsin and belinostat in chronic lymphocytic leukemia cells. *Clin Cancer Res* 14 (2), 549-558

767 Adjei, A.A. et al. (2000) A Phase I trial of the farnesyl transferase inhibitor SCH66336: evidence for biological and clinical activity. *Cancer Res* 60 (7), 1871-1877

768 David, E. et al. (2005) The combination of the farnesyl transferase inhibitor lonafarnib and the proteasome inhibitor bortezomib induces synergistic apoptosis in human myeloma cells that is associated with down-regulation of p-AKT. *Blood* 106 (13), 4322-4329

769 Shao, J. et al. (2000) Oncogenic Ras-mediated cell growth arrest and apoptosis are associated with increased ubiquitin-dependent cyclin D1 degradation. *J Biol Chem* 275 (30), 22916-22924

770 Bundred, N.J. (2007) Bs24 apoptosis: why surgeons need to understand it. *ANZ J Surg* 77 Suppl 1, A6

771 Coradini, D. et al. (1994) Activity of tamoxifen and new antiestrogens on estrogen receptor positive and negative breast cancer cells. *Anticancer Res* 14 (3A), 1059-1064

772 Argiris, A. et al. (2004) Synergistic interactions between tamoxifen and trastuzumab (Herceptin). *Clinical Cancer Research* 10 (4), 1409-1420

773 Osborne, C.K. and Schiff, R. (2003) Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer. *Breast* 12 (6), 362-367

774 Schiff, R. et al. (2004) Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance. *Clin Cancer Res* 10 (1 Pt 2), 331S-336S

775 Sawyers, C.L. (2003) Will mTOR inhibitors make it as cancer drugs? *Cancer Cell* 4 (5), 343-348

776 Sebolt-Leopold, J.S. and Herrera, R. (2004) Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer* 4 (12), 937-947

777 Legrier, M.E. et al. (2007) Targeting protein translation in human non small cell

lung cancer via combined MEK and mammalian target of rapamycin suppression. *Cancer Res* 67 (23), 11300-11308

778 Marzec, M. et al. (2007) Oncogenic tyrosine kinase NPM/ALK induces activation of the rapamycin-sensitive mTOR signaling pathway. *Oncogene* 26 (38), 5606-5614

779 Deininger, M.W. and Druker, B.J. (2003) Specific targeted therapy of chronic myelogenous leukemia with imatinib. *Pharmacol Rev* 55 (3), 401-423

780 Gianni, M. et al. (2001) Tyrosine kinase inhibitor STI571 potentiates the pharmacologic activity of retinoic acid in acute promyelocytic leukemia cells: effects on the degradation of RARalpha and PML-RARalpha. *Blood* 97 (10), 3234-3243

781 de Bono, J.S. et al. (2003) Farnesyltransferase inhibitors and their potential in the treatment of breast carcinoma. *Semin Oncol* 30 (5 Suppl 16), 79-92

782 Radujkovic, A. et al. (2006) Combination treatment of imatinib-sensitive and -resistant BCR-ABL-positive CML cells with imatinib and farnesyltransferase inhibitors. *Anticancer Res* 26 (3A), 2169-2177

783 Hoover, R.R. et al. (2002) Overcoming STI571 resistance with the farnesyl transferase inhibitor SCH66336. *Blood* 100 (3), 1068-1071

784 Goga, A. et al. (1995) Alternative signals to RAS for hematopoietic transformation by the BCR-ABL oncogene. *Cell* 82 (6), 981-988

785 McCubrey, J.A. et al. (2008) Targeting survival cascades induced by activation of Ras/Raf/MEK/ERK, PI3K/PTEN/Akt/mTOR and Jak/STAT pathways for effective leukemia therapy. *Leukemia*

786 Kim, S.Y. et al. (2008) Trastuzumab inhibits the growth of human gastric cancer cell lines with HER2 amplification synergistically with cisplatin. *Int J Oncol* 32 (1), 89-95

787 Shah, N.P. et al. (2004) Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science* 305 (5682), 399-401

788 Walport, M.J. and Isenberg, D.A. (1992) Research priorities in rheumatology. *Br J Rheumatol* 31 (8), 505-506

789 Odgerel, T. et al. (2007) The FLT3 inhibitor PKC412 exerts differential cell cycle effects on leukemic cells depending on the presence of FLT3 mutations. *Oncogene*

790 Fabbro, D. et al. (2000) PKC412--a protein kinase inhibitor with a broad therapeutic potential. *Anticancer Drug Des* 15 (1), 17-28

791 Gleixner, K.V. et al. (2007) Synergistic growth-inhibitory effects of two tyrosine kinase inhibitors, dasatinib and PKC412, on neoplastic mast cells expressing the D816V-mutated oncogenic variant of KIT. *Haematologica* 92 (11), 1451-1459

792 Vichalkovski, A. et al. (2006) Tyrosine kinase modulation of protein kinase C activity regulates G protein-linked Ca2+ signaling in leukemic hematopoietic cells. *Cell Calcium* 39 (6), 517-528

793 Qi, X. and Mochly-Rosen, D. (2008) The PKC{delta} -Abl complex communicates ER stress to the mitochondria - an essential step in subsequent apoptosis. *J Cell Sci* 121 (Pt 6), 804-813

794 Trudel, S. et al. (2007) The Bcl-2 family protein inhibitor, ABT-737, has substantial antimyeloma activity and shows synergistic effect with dexamethasone and melphalan. *Clin Cancer Res* 13 (2 Pt 1), 621-629

795 Laane, E. et al. (2007) Dexamethasone-induced apoptosis in acute lymphoblastic leukemia involves differential regulation of Bcl-2 family members. *Haematologica* 92 (11), 1460-1469

796 Gil, L. et al. (2007) Activity of bortezomib in adult de novo and relapsed acute myeloid leukemia. *Anticancer Res* 27 (6B), 4021-4025

797 Cardoso, F. et al. (2006) Bortezomib (PS-341, Velcade) increases the efficacy of trastuzumab (Herceptin) in HER-2-positive breast cancer cells in a synergistic manner. *Mol Cancer Ther* 5 (12), 3042-3051

798 Biswas, D.K. et al. (2000) Epidermal growth factor-induced nuclear factor kappa B activation: A major pathway of cell-cycle progression in estrogen-receptor negative breast cancer cells. *Proc Natl Acad Sci U S A* 97 (15), 8542-8547

799 Yu, C. et al. (2003) Histone deacetylase inhibitors promote STI571-mediated apoptosis in STI571-sensitive and -resistant Bcr/Abl+ human myeloid leukemia cells. *Cancer Res* 63 (9), 2118-2126

800 Hofmann, W.K. et al. (2002) Ph(+) acute lymphoblastic leukemia resistant to the tyrosine kinase inhibitor STI571 has a unique BCR-ABL gene mutation. *Blood* 99 (5), 1860-1862

801 Cunningham, J.T. et al. (2007) mTOR controls mitochondrial oxidative function through a YY1-PGC-1alpha transcriptional complex. *Nature* 450 (7170), 736-740

802 Xu, R.H. et al. (2005) Synergistic effect of targeting mTOR by rapamycin and depleting ATP by inhibition of glycolysis in lymphoma and leukemia cells. *Leukemia* 19 (12), 2153-2158

803 Nau, P.N. et al. (2002) Metabolic adaptation of the fetal and postnatal ovine heart: regulatory role of hypoxia-inducible factors and nuclear respiratory factor-1. *Pediatr Res* 52 (2), 269-278

804 Wetzler, M. et al. (2007) Synergism between arsenic trioxide and heat shock protein 90 inhibitors on signal transducer and activator of transcription protein 3 activity--pharmacodynamic drug-drug interaction modeling. *Clin Cancer Res* 13 (7), 2261-2270

805 Schwock, J. et al. (2007) Efficacy of Hsp90 inhibition for induction of apoptosis and inhibition of growth in cervical carcinoma cells in vitro and in vivo. *Cancer Chemother Pharmacol*

806 Chatterjee, M. et al. (2007) STAT3 and MAPK signaling maintain overexpression of heat shock proteins 90alpha and beta in multiple myeloma cells, which critically contribute to tumor-cell survival. *Blood* 109 (2), 720-728

807 Lu, J. et al. (2007) Targeting thioredoxin reductase is a basis for cancer therapy by arsenic trioxide. *Proc Natl Acad Sci U S A* 104 (30), 12288-12293

808 Wang, E. et al. (2001) The farnesyl protein transferase inhibitor SCH66336 is a potent inhibitor of MDR1 product P-glycoprotein. *Cancer Res* 61 (20), 7525-7529

809 Ghosal, A. et al. (2006) Identification of human liver cytochrome P450 enzymes responsible for the metabolism of lonafarnib (Sarasar). *Drug Metab Dispos* 34 (4), 628-635

810 Shi, B. et al. (2000) The farnesyl protein transferase inhibitor SCH66336 synergizes with taxanes in vitro and enhances their antitumor activity in vivo. *Cancer Chemother Pharmacol* 46 (5), 387-393

811 Pan, L. et al. (2007) Histone deacetylase inhibitor trichostatin a potentiates doxorubicin-induced apoptosis by up-regulating PTEN expression. *Cancer* 109 (8), 1676-1688

812 Florenes, V.A. et al. (2004) Deacetylase inhibition in malignant melanomas: impact on cell cycle regulation and survival. *Melanoma Res* 14 (3), 173-181

813 Gartel, A.L. et al. (1996) p21--negative regulator of the cell cycle. *Proc Soc Exp Biol Med* 213 (2), 138-149

814 Dowdy, S.C. et al. (2006) Histone deacetylase inhibitors and paclitaxel cause synergistic effects on apoptosis and microtubule stabilization in papillary serous endometrial cancer cells. *Mol Cancer Ther* 5 (11), 2767-2776

815 Hur, J. et al. (2006) Regulation of expression of BIK proapoptotic protein in human breast cancer cells: p53-dependent induction of BIK mRNA by fulvestrant and proteasomal degradation of BIK protein. *Cancer Res* 66 (20), 10153-10161

816 Primeau, M. et al. (2003) Synergistic antineoplastic action of DNA methylation inhibitor 5-AZA-2'-deoxycytidine and histone deacetylase inhibitor depsipeptide on human breast carcinoma cells. *International Journal of Cancer* 103 (2), 177-184

817 Falugi, C. et al. (2003) Increasing complexity of farnesyltransferase inhibitors activity: role in chromosome instability. *Curr Cancer Drug Targets* 3 (2), 109-118

818 Russo, P. et al. (2002) RPR-115135, a farnesyltransferase inhibitor, increases 5-FU- cytotoxicity in ten human colon cancer cell lines: role of p53. *Int J Cancer* 100 (3), 266-275

819 Ogata, M. et al. (2005) Mechanism of action of dipropofol and synergistic action with other antibacterial agents in vitro. *Biol Pharm Bull* 28 (9), 1773-1775

820 Boger, D.L. (2001) Vancomycin, teicoplanin, and ramoplanin: synthetic and mechanistic studies. *Med Res Rev* 21 (5), 356-381

821 Qiu, Y. et al. (2007) The farnesyltransferase inhibitor R115777 up-regulates the expression of death receptor 5 and enhances TRAIL-induced apoptosis in human lung cancer cells. *Cancer Res* 67 (10), 4973-4980

822 Medeiros, B.C. et al. (2007) The farnesyl transferase inhibitor, tipifarnib, is a potent inhibitor of the MDR1 gene product, P-glycoprotein, and demonstrates significant cytotoxic synergism against human leukemia cell lines. *Leukemia* 21 (4), 739-746

823 Caraglia, M. et al. (2004) The farnesyl transferase inhibitor R115777 (Zarnestra) synergistically enhances growth inhibition and apoptosis induced on epidermoid cancer cells by Zoledronic acid (Zometa) and Pamidronate. *Oncogene* 23 (41), 6900-6913

824 LaVallee, T.M. et al. (2003) 2-methoxyestradiol up-regulates death receptor 5 and induces apoptosis through activation of the extrinsic pathway. *Cancer Res* 63 (2), 468-475

825 Debernardis, D. et al. (1996) Interactions between taxol and camptothecin. *Anticancer Drugs* 7 (5), 531-534

826 Takebayashi, Y. et al. (1999) Poisoning of human DNA topoisomerase I by ecteinascidin 743, an anticancer drug that selectively alkylates DNA in the minor groove. *Proc Natl Acad Sci U S A* 96 (13), 7196-7201

827 Fayette, J. et al. (2006) ET-743: a novel agent with activity in soft-tissue sarcomas. *Curr Opin Oncol* 18 (4), 347-353

828 Takebayashi, Y. et al. (2001) Ecteinascidin 743 induces protein-linked DNA breaks in human colon carcinoma HCT116 cells and is cytotoxic independently of topoisomerase I expression. *Clin Cancer Res* 7 (1), 185-191

829 Beumer, J.H. et al. (2007) Trabectedin (ET-743, Yondelis) is a substrate for P-glycoprotein, but only high expression of P-glycoprotein confers the multidrug resistance phenotype. *Invest New Drugs* 25 (1), 1-7

830 Takahashi, N. et al. (2002) Sequence-dependent synergistic cytotoxicity of ecteinascidin-743 and paclitaxel in human breast cancer cell lines in vitro and in vivo. *Cancer Res* 62 (23), 6909-6915

831 Tedesco, K.L. and Rybak, M.J. (2004) Daptomycin. *Pharmacotherapy* 24 (1), 41-57

832 Silverman, J.A. et al. (2003) Correlation of daptomycin bactericidal activity and membrane depolarization in Staphylococcus aureus. *Antimicrob Agents Chemother* 47 (8), 2538-2544

833 Boaretti, M. and Canepari, P. (2000) Purification of daptomycin binding proteins (DBPs) from the membrane of Enterococcus hirae. *New Microbiol* 23 (3), 305-317

834 Wegrzyn, A. et al. (1998) Differential inhibition of transcription from sigma70- and sigma32-dependent promoters by rifampicin. *FEBS Lett* 440 (1-2), 172-174

835 Galvao, M. (1990) Role of angiotensin-converting enzyme inhibitors in congestive heart failure. *Heart Lung* 19 (5 Pt 1), 505-511

836 Patick, A.K. et al. (1997) Activities of the human immunodeficiency virus type 1 (HIV-1) protease inhibitor nelfinavir mesylate in combination with reverse transcriptase and protease inhibitors against acute HIV-1 infection in vitro. *Antimicrob Agents Chemother* 41 (10), 2159-2164

837 Molla, A. et al. (2002) In vitro antiviral interaction of lopinavir with other protease inhibitors. *Antimicrob Agents Chemother* 46 (7), 2249-2253

838 Sobrado, M. et al. (2003) Combined nimodipine and citicoline reduce infarct size, attenuate apoptosis and increase bcl-2 expression after focal cerebral ischemia. *Neuroscience* 118 (1), 107-113

839 Hansen, M.R. et al. (2007) Overexpression of Bcl-2 or Bcl-xL prevents spiral ganglion neuron death and inhibits neurite growth. *Dev Neurobiol* 67 (3), 316-325

840 Rong, Y. and Distelhorst, C.W. (2007) Bcl-2 Protein Family Members: Versatile Regulators of Calcium Signaling in Cell Survival and Apoptosis. *Annu Rev Physiol*

841 Yagami, T. et al. (2003) Prostaglandin E2 rescues cortical neurons from amyloid beta protein-induced apoptosis. *Brain Res* 959 (2), 328-335

842 Tham, S.M. et al. (2005) Synergistic and additive interactions of the cannabinoid agonist CP55,940 with mu opioid receptor and alpha2-adrenoceptor agonists in acute pain models in mice. *Br J Pharmacol* 144 (6), 875-884

843 Blednov, Y.A. et al. (2003) A pervasive mechanism for analgesia: activation of GIRK2 channels. *Proc Natl Acad Sci U S A* 100 (1), 277-282

844 Paris, A. and Tonner, P.H. (2005) Dexmedetomidine in anaesthesia. *Curr Opin Anaesthesiol* 18 (4), 412-418

845 Hara, K. et al. (2005) The effects of anesthetics and ethanol on alpha2 adrenoceptor subtypes expressed with G protein-coupled inwardly rectifying potassium channels in Xenopus oocytes. *Anesth Analg* 101 (5), 1381-1388

846 Raehal, K.M. and Bohn, L.M. (2005) Mu opioid receptor regulation and opiate responsiveness. *Aaps J* 7 (3), E587-591

847 Marker, C.L. et al. (2005) Spinal G-protein-gated potassium channels contribute in a dose-dependent manner to the analgesic effect of mu- and delta- but not kappa-opioids. *J Neurosci* 25 (14), 3551-3559

848 Tsuura, Y. et al. (1994) Nitric oxide opens ATP-sensitive K+ channels through suppression of phosphofructokinase activity and inhibits glucose-induced insulin release in pancreatic beta cells. *J Gen Physiol* 104 (6), 1079-1098

849 Slater, A.F. (1993) Chloroquine: mechanism of drug action and resistance in Plasmodium falciparum. *Pharmacol Ther* 57 (2-3), 203-235

850 Alvarez, S. et al. (2007) Structure, function and modulation of retinoic acid receptor beta, a tumor suppressor. *Int J Biochem Cell Biol*

851 Lee, M.O. and Kang, H.J. (2002) Role of coactivators and corepressors in the

induction of the RARbeta gene in human colon cancer cells. *Biol Pharm Bull* 25 (10), 1298-1302

852 Morris, M.I. and Villmann, M. (2006) Echinocandins in the management of invasive fungal infections, part 1. *Am J Health Syst Pharm* 63 (18), 1693-1703

853 Karlowsky, J.A. et al. (2006) In vitro interactions of anidulafungin with azole antifungals, amphotericin B and 5-fluorocytosine against Candida species. *Int J Antimicrob Agents* 27 (2), 174-177

854 Kuefer, R. et al. (2007) Antagonistic effects of sodium butyrate and N-(4-hydroxyphenyl)-retinamide on prostate cancer. *Neoplasia* 9 (3), 246-253

855 Shimada, K. et al. (2002) Contributions of mitogen-activated protein kinase and nuclear factor kappa B to N-(4-hydroxyphenyl)retinamide-induced apoptosis in prostate cancer cells. *Mol Carcinog* 35 (3), 127-137

856 Pankey, G.A. and Ashcraft, D.S. (2005) In vitro synergy of ciprofloxacin and gatifloxacin against ciprofloxacin-resistant Pseudomonas aeruginosa. *Antimicrob Agents Chemother* 49 (7), 2959-2964

857 Duncan, R. et al. (2005) Polymer-drug conjugates: towards a novel approach for the treatment of endrocine-related cancer. *Endocr Relat Cancer* 12 Suppl 1, S189-199

858 Tahara, H. et al. (2006) Inhibition of oat3-mediated renal uptake as a mechanism for drug-drug interaction between fexofenadine and probenecid. *Drug Metab Dispos* 34 (5), 743-747

859 Nakajima, Y. et al. (2004) Mechanism of the drug interaction between valproic acid and carbapenem antibiotics in monkeys and rats. *Drug Metab Dispos* 32 (12), 1383-1391

860 Tsai, J.C. et al. (1996) Metabolic approaches to enhance transdermal drug delivery. 1. Effect of lipid synthesis inhibitors. *J Pharm Sci* 85 (6), 643-648

861 Babita, K. and Tiwary, A.K. (2005) Transcutaneous delivery of levodopa: enhancement by fatty acid synthesis inhibition. *Mol Pharm* 2 (1), 57-63

862 Koller, W.C. and Rueda, M.G. (1998) Mechanism of action of dopaminergic agents in Parkinson's disease. *Neurology* 50 (6 Suppl 6), S11-14; discussion S44-18

863 Tomita, M. et al. (1995) Absorption-enhancing mechanism of sodium caprate and decanoylcarnitine in Caco-2 cells. *J Pharmacol Exp Ther* 272 (2), 739-743

864 Motlekar, N.A. et al. (2005) Oral delivery of low-molecular-weight heparin using sodium caprate as absorption enhancer reaches therapeutic levels. *J Drug Target* 13 (10), 573-583

865 Verhaeghe, R. (1998) The use of low-molecular-weight heparins in cardiovascular disease. *Acta Cardiol* 53 (1), 15-21

866 Kanamitsu, S.I. et al. (2000) Prediction of in vivo drug-drug interactions based on mechanism-based inhibition from in vitro data: inhibition of 5-fluorouracil metabolism by (E)-5-(2-Bromovinyl)uracil. *Drug Metab Dispos* 28 (4), 467-474

867 Podder, H. et al. (2001) Pharmacokinetic interactions augment toxicities of sirolimus/cyclosporine combinations. *J Am Soc Nephrol* 12 (5), 1059-1071

868 Hayashi, K. et al. (2000) Potentiation of ganciclovir toxicity in the herpes simplex virus thymidine kinase/ganciclovir administration system by ponicidin. *Cancer Gene Ther* 7 (1), 45-52

869 Okuda, T. et al. (2004) Drug interaction between methamphetamine and antihistamines: behavioral changes and tissue concentrations of methamphetamine in rats. *Eur J Pharmacol* 505 (1-3), 135-144

870 Rothman, R.B. et al. (2006) Dual dopamine-5-HT releasers: potential treatment

agents for cocaine addiction. *Trends Pharmacol Sci* 27 (12), 612-618

871 Hayashi, K. et al. (2006) The role of a HSV thymidine kinase stimulating substance, scopadulciol, in improving the efficacy of cancer gene therapy. *J Gene Med* 8 (8), 1056-1067

872 Simonson, S.G. et al. (2004) Rosuvastatin pharmacokinetics in heart transplant recipients administered an antirejection regimen including cyclosporine. *Clinical Pharmacology and Therapeutics* 76 (2), 167-177

873 Yamazaki, M. et al. (2005) Effects of fibrates on human organic anion-transporting polypeptide 1B1-, multidrug resistance protein 2- and P-glycoprotein-mediated transport. *Xenobiotica* 35 (7), 737-753

874 Schneck, D.W. et al. (2004) The effect of gemfibrozil on the pharmacokinetics of rosuvastatin. *Clin Pharmacol Ther* 75 (5), 455-463

875 Higley, B. et al. (1982) Pyrimidine nucleoside phosphorylase activity in tumour and matched normal gastrointestinal mucosa. *Gut* 23 (12), 1072-1076

876 Meyers, M. et al. (2001) Role of the hMLH1 DNA mismatch repair protein in fluoropyrimidine-mediated cell death and cell cycle responses. *Cancer Res* 61 (13), 5193-5201

877 Cowen, R.L. et al. (2004) Hypoxia targeted gene therapy to increase the efficacy of tirapazamine as an adjuvant to radiotherapy: reversing tumor radioresistance and effecting cure. *Cancer Res* 64 (4), 1396-1402

878 Wang, Q. et al. (2007) Monocarboxylate transporter (MCT) mediates the transport of gamma-hydroxybutyrate in human kidney HK-2 cells. *Pharm Res* 24 (6), 1067-1078

879 Wang, Q. and Morris, M.E. (2007) Flavonoids modulate monocarboxylate transporter-1-mediated transport of gamma-hydroxybutyrate in vitro and in vivo. *Drug Metab Dispos* 35 (2), 201-208

880 Ngui, J.S. et al. (2000) Cytochrome P450 3A4-mediated interaction of diclofenac and quinidine. *Drug Metab Dispos* 28 (9), 1043-1050

881 Mackenzie, P.I. (2000) Identification of uridine diphosphate glucuronosyltransferases involved in the metabolism and clearance of mycophenolic acid. *Ther Drug Monit* 22 (1), 10-13

882 Kuypers, D.R. et al. (2005) Drug interaction between mycophenolate mofetil and rifampin: possible induction of uridine diphosphate-glucuronosyltransferase. *Clin Pharmacol Ther* 78 (1), 81-88

883 Dickinson, R.G. et al. (1984) pH-dependent rearrangement of the biosynthetic ester glucuronide of valproic acid to beta-glucuronidase-resistant forms. *Drug Metab Dispos* 12 (2), 247-252

884 Perucca, E. (2002) Pharmacological and therapeutic properties of valproate: a summary after 35 years of clinical experience. *CNS Drugs* 16 (10), 695-714

885 Hynes, N.E. and Schlange, T. (2006) Targeting ADAMS and ERBBs in lung cancer. *Cancer Cell* 10 (1), 7-11

886 Bos, J.L. (1988) The ras gene family and human carcinogenesis. *Mutat Res* 195 (3), 255-271

887 McDermott, U. et al. (2007) Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc Natl Acad Sci U S A* 104 (50), 19936-19941

888 Bader, A.G. et al. (2006) Cancer-specific mutations in PIK3CA are oncogenic in vivo. *Proc Natl Acad Sci U S A* 103 (5), 1475-1479

889 Gymnopoulos, M. et al. (2007) Rare cancer-specific mutations in PIK3CA show

gain of function. *Proc Natl Acad Sci U S A* 104 (13), 5569-5574

890 Forgacs, E. et al. (1998) Mutation analysis of the PTEN/MMAC1 gene in lung cancer. *Oncogene* 17 (12), 1557-1565

891 Hollestelle, A. et al. (2007) Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol Cancer Res* 5 (2), 195-201

892 Jonsson, G. et al. (2007) High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. *Genes Chromosomes Cancer* 46 (6), 543-558

APPENDICES

Appendix Table S1 Literature reported pharmacodynamically synergistic drug combinations due to anti-counteractive actions, in which synergy has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed.

Combination target relationship	**Drug A (mechanism of actions related to synergy)**	**Drug B (mechanism of actions related to synergy)**	**Reported synergistic effect**	**Synergism determination method**	**Possible mechanism of synergism in anti-counteractive actions**
Different targets of the same pathway	17-AAG (heat-shock protein antagonist, induced cell cycle inhibition and apoptosis by inhibiting NF-kappaB, AP-1 and PI3K/Akt pathways[376], Hsp90/FLT3 inhibitor[480])	Arsenic trioxide (degraded aberrant PML-retinoic acid receptor alpha fusion protein, generated reactive oxygen species, and activated Akt survival pathway[437])	Synergistic anticancer effect[437]	Median dose effect analysis (Calcusym)	Arsenic trioxide's anticancer generation of reactive oxygen species is partially off-set by its own counteractive activation of Akt survival pathway[376]. 17-AAG abrogated arsenic trioxide's activation of Akt survival pathway[376] to reduce the counteractive effect
	Oxaliplatin (DNA adduct, preferably bind to major groove of GG, AG and TACT sites, complex conformation different from that of cisplatin[408], caused DNA strand break and non-DNA initiated apoptosis[409])	Irinotecan (DNA topoisomerase I inhibitor, increased EGFR phosphorylation in Lovo & WiDR cells[410])	Synergistic anticancer effect in AZ-521 and NUGC-4 cells, additive effect in MKN-45 cells[411]	Median drug effect analysis	Effect of oxaliplatin's DNA adduct formation[408] may be partially reduced by certain mutant DNA topoisomerase I acting on DNA adduct to generate different topoisomers[412]. Irinotecan inhibition of DNA topoisomerase I[410] partially off-sets this counteractive activity
Different targets of the same pathway that regulated the same target	Cisplatin (DNA inter- and intra-strand adduct, preferably bind to the major groove of GG, AG and TACT sites[381] thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis[510])	Trabectedin (bind covalently to central G in the minor groove of selected DNA pyrimidine-G-G and purine-G-C triplets[385], formed unusual DNA replication intermediates thereby inhibited DNA replication[386], interacted with DNA and DNA repair systems in a way different from cisplatin[384])	Synergistic antitumor activity[384]	Interaction index method of Berebbaum	Trabectedin inhibition of DNA replication[386] reduced the counteractive activity of DNA polymerase mediated mutagenic translesional bypass replication across cisplatin-DNA adducts[382]
	Cisplatin (DNA inter- and intra-strand adduct, preferably bind to the major groove of GG, AG and	Topotecan (topoisomerase I inhibitor, interacted with DNA, stabilized a covalent	Synergistic cytotoxic activity[378-380]	Multi-drug effect equation, combination	Topotecan blocking of DNA replication[383] reduced the counteractive activity of mutagenic translesional bypass replication across cisplatin-DNA adducts[382]

	TACT sites[381] thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis[510])	topoisomerase-DNA complex, thereby blocked DNA replication forks[383])		index, median-drug effect method	
	Cisplatin (DNA inter- and intra-strand adduct, preferably bind to the major groove of GG, AG and TACT sites[381] thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis [510])	Sabarubicin (topoisomerase II inhibitor[17])	Synergistic cytotoxic effect in tumour cell lines NSCLC H460 and SCLC GLC4[747]	Combination index	Sabarubicin blocking of DNA replication[17] reduced the counteractive activity of mutagenic translesional bypass replication across cisplatin-DNA adducts[382]
	DL-cycloserine (bacterial cell wall synthesis inhibitor[413])	Epigallocatechin gallate (disrupted integrity of bacterial cell wall via direct binding to peptidoglycan[413])	Synergistic effect on bacterial cell wall[413]	Fractional inhibitory concentration index	Cell wall alteration may induce counteractive cell wall synthesis to restore cell wall integrity[414], DL-cycloserine inhibition of cell wall synthesis hindered the restoration thereby enhanced Epigallocatechin gallate's cell wall disruption activity
	Gefitinib (EGFR tyrosine kinase inhibitor, induced cyclin-dependent kinase inhibitors, inhibited p27 and p21, decreased MMP-2 and MMP-9 enzyme activity[422])	Irinotecan (DNA topoisomerase I inhibitor, increased EGFR phosphorylation in Lovo & WiDR cells[410])	Synergistic inhibitory effect on colorectal cancer Lovo & WiDR cells[410]	Combination index	Irinotecan produced anticancer effect via DNA topoisomerase inhibition, but promoted proliferation by increased phosphorylation of EGFR in certain cell types[410]. Gefitinib produced anticancer effect via EFFR tyrosine kinase inhibition and others [422], which offsets the counteractive effect of increased EGFR phosphorylation
Different targets of related pathways	17-AAG (heat-shock protein antagonist, induced cell cycle inhibition and apoptosis by inhibiting NF-kappaB, AP-1 and PI3K/Akt pathways[376])	Etoposide (topoisomerase II inhibitor, induced DNA double-strand breaks during DNA replication[748], increased expression of DNA repair-related protein Rad51[749])	Synergism between etoposide and 17-AAG in leukemia cells[480]	Combination index method (CalcuSyn by BioSoft)	The effect of etoposide's DNA strand break is partially offset by its own counteractive increase of expression of DNA repair-related protein Rad51[749]. Higher levels of Rad51 and its interacting partner Chk1 are associated with presence of FLT3[480]. Inhibition of Hsp90/FLT3 by 17-AAG may reduce Rad51 and Chk1 to reduce the counteractive effect
	Artemisinin (interacted with heme to mediate its decomposition into free radicals that alkylate essential malarial proteins[750])	Methylene blue (inhibited heme polymerization[751]; selectively inhibited glutathione reductase[752]; soluble guanylate cyclase inhibitor[753]	Synergistic antimalarial effects in both chloroquine - sensitive and - resistant P.	isobologram method	Artemisinin's antimalarial activity possibly arise from its interaction with heme that facilitates heme conversion into free radical[755], which can be off-set by parasite's counteractive actions of heme polymerization into insoluble hemozoin and heme degradation by glutathione[756]. These counteractive

			falciparum strains[754]		actions are partially reduced by methylene blue's inhibition of heme polymerization[751] and glutathione reductase[752], resulting in synergistic antimalarial effect
	Erlotinib (EGFR tyrosine kinase inhibitor[757])	Pemetrexed (dihydrofolate reductase, thymidylate synthase and glycinamide ribonucleotide formyl transferase inhibitor[758-760];increased EGFR phosphorylation and reduced Akt phosphorylation[761])	Synergistic cytotoxicity in all cells[761]	Combination-index method	Erlotinib's inhibition of EGFR may trigger activation of Akt-mediated negative-feedback signaling, leading to increased membrane HER3 expression and reduced HER3 phosphatase activity, thereby evading EGFR inhibition[31]. This counteractive action can be partially reduced by pemetrexed's reduction of Akt phosphorylation[761], leading to synergistic effect
	Bortezomib (proteasome inhibitor, protected pro-apoptotic pathways by inhibiting proteasome degradation of P53[762], inhibited NF-kappaB and induced endoplasmic reticulum stress[763])	Sodium butyrate or suberoylanilide hydroxamic acid (Histone deacetylase inhibitor, promoted histone acetylation and chromatin structure relaxation[764], down-regulated expression levels of Bcr-abl, c-Myc and HDAC3[765]; mediated RelA acetylation and NF-kappaB activation[766]	Synergistic apoptosis induction in human multiple myeloma cells[41]	Combination Index	HDAC inhibitor's pro-apoptotic down-regulation of Bcr-abl is partially offset by its mediation of NF-kappaB activation [763]. This counteractive action can be partially reduced by bortezomib's pro-apoptotic inhibition of NF-kappaB[766]
	Lonafarnib (farnesyl transferase inhibitor, inhibited Ras farnesylation[767])	Bortezomib (proteasome inhibitor, protected pro-apoptotic pathways by inhibiting proteasome degradation of P53[762], inhibited NF-kappaB and induced endoplasmic reticulum stress[763])	Synergistic myeloma- cell death activity[768]	Isobolographic analysis	Bortezomib produced anticancer effect by inhibiting proteasome degradation of P53[762]. But protesome inhibition reduced ubiquitin-dependent cyclin D1 degradation, which hindered Ras-mediated cell growth arrest and apoptosis[769] thereby reduced bortezomib's anticancer effect. This counteractive activity can be partially offset by lonafamib inhibition of Ras farnesylation[767] that subsequently induced apoptosis by activating the pro-apoptotic protein BAD in BCL2 family[770]
	Tamoxifen (estrogen receptor antagonist[771])	Trastuzumab (herceptin) (anti-HER-2/neu antibody[563])	Synergistic growth inhibition in ER-positive, HER-2/neu -	Combination index	ER crosstalks with EGFR and HER-2/neu[774], signaling via EGFR and HER-2/neu can activate ER and its coactivator AIB1, ER of cell membrane can activate EGFR/HER-2[773]. Anti-HER-2/neu antibody

			overexpressing BT-474 breast tumor cells[772,773]		trastuzumab[563] stopped HER-2/neu induced activation of ER and AIB1. ER antagonist tamoxifen[771] stopped ER induced activation of EGFR/HER-2. Use of both drugs reduced the counteractive crosstalks
	Rapamycin or deforolimus (mTOR inhibitor[775])	CI-1040 or PD0325901 (MEK inhibitor[776])	Synergistic antitumor efficacy in animal models of human lung cancer and in K-RAS mutant, non-V600EB-RAF, B-RAFV600E mutant cell lines[777]	Cell proliferation assay and combination index method of Chou and Talalay	Effects of the inhibition of mTOR by rapamycin or its analogue deforolimus may be partially offset by NPM/ALK-induced mTOR activation that is transduced through the MEK-ERK signaling pathway[778]. This countractive action may be reduced by CI-1040 or PD0325901's inhibition of MEK[776]
	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395] distorted mitosis to trigger apoptosis[415], induced p53 and CDK inhibitors[416], activated caspase-10, caspases-8, -6, and -3, leading to apoptosis[417], activated ERK [418] and CDK2[419], activated p38 MAP kinase and p53[420])	NU6140 (CDK inhibitor, down-regulated antiapoptotic protein survivin[421])	Synergistic apoptotic response[421]	Median drug effect analysis	Use of both drugs promoted complementary apoptosis activities via triple actions of surviving down-regulation by NU6140[421], microtubule stabilization[395] and caspase activation[417] by paclitaxel. Paclitaxel's promotion of apoptosis may be partially offset by its counteractive pro-growth activation of ERK[418] and CDK2[419], which may be partially reduced by NU6140's inhibition of CDK[421]
Different targets of cross-talking pathways	Gefitinib (EGFR tyrosine kinase inhibitor, induced cyclin-dependent kinase inhibitors p27 and p21, decreased MMP-2 and MMP-9 enzyme activity[422])	Taxane (disrupted microtubule by binding to beta-tubulin[423], induced tumor suppressor gene p53 and cyclin-dependent kinase inhibitors P21, down regulated Bcl-2, leading to apoptosis[416])	Strong synergistic effect in breast cancer MCF7/ADR cells[424]	Combination index	Taxane produced anticancer effect by inducing apoptosis[416] and microtubule disruption[423]. Crosstalk between EGFR and hypoxia-inducible factor-1alpha pathways increased resistance to apoptosis by up-regulating survivin[28]. Gefitinib produced anticancer effect via EFFR tyrosine kinase inhibition, which offsets the counteractive EGFR-hypoxia crosstalk in resisting taxane's pro-apoptosis activity
	Gleevec (selective inhibitor of c-Abl, p210bcr-abl, c-Kit, and	L744,832 or LB42918 (farnesyltransferase inhibitor,	Synergistically promoted	Median dose effect analysis	Gleevec inhibition of Abl may leed to selection of resistant mutatons in Bcr-Abl[783], some of the Bcr-Abl

	PDGF-R tyrosine kinases[779,780])	inhibited Ras farnesylation[781])	apoptosis in different imatinib-sensitive and -resistant BCR-ABL-positive CML cells[782]	method of Chou and Talalay	mutants bind to Ras associated proteins to activate an alternative Ras mediated tansformation[784] and survival[785] signal. The survival signal involves activation of survival cascades via Ras/Raf/MEK/ERK signaling[785]. This counteractive alternative signal may be partially blocked by using farnesyltransferase inhibitor to inhibit Ras farnesylation
	Cisplatin (DNA inter- and intra-strand adduct, preferably bind to the major groove of GG, AG and TACT sites[381] thereby inhibited DNA polymerization and induced DNA damage to trigger apoptosis[510])	Trastuzumab (herceptin) (anti-HER-2/neu antibody[563])	Synergistic growth inhibition in SNU-216 as an HER2-amplified cell line among gastric cancer cell lines[786]	Isobologram analysis	Cisplatin formed DNA adduct to induce DNA damage and apoptosis, which may be attenuated by DNA repair systems in certain cell types[510]. This counteractive DNA repair action may be partially reduced by herceptin's anti-HER2 activitity that suppressed DNA repair pathway known to crosstalk to HER2[564] and inhibited PI3K-AKT pathway[565] to enhance apoptosis[566]
	Dasatinib(inhibitor of c-abl,src, fyn, lck and kit[787,788])	PKC412 (inhibitor of Flt[789], PKC, VEGFR2, PDGFR, c-kit[790])	Synergistic apoptotic effects in HMC-1.2 cells[791]	Combination index	Inhibition of c-abl and src enhances catalytic activity of some PKC subtypes and their binding to Bcr/Abl in specific cell types[792]. Complex of abl with certain PKC subtypes enable communication of ER stress to mitochondria, which is an essential step in subsequent apoptosis [793]. This possible counteractive action against dasatinib's inhibition of c-abl and src[77] may be partially alleviated by PKC412's inhibition of PKC[80]
Different targets in the same pathway that crosstalks via other pathway	Gefitinib (EGFR tyrosine kinase inhibitor, induced cyclin-dependent kinase inhibitors p27 and p21, decreased MMP-2 and MMP-9 enzyme activity[422])	PD98059 (MEK inhibiton[425])	Synergistic antitumor effect in breast cancer MDA-MB-361 cells186	Combination index, isobolographic analysis	An autocrine growth loop critical for tumor growth is formed in EGFR-Ras-Raf-MEK-ERK network such that activated MEK activates ERK which upregulates EGFR ligands which promotes the autocrine growth loop[426]. This loop produced counteractive activity against gefitinib or PD98059 by reducing the effect of MEK or EGFR tyrosine kinase inhibition. Simultaneous use of both drugs helps disrupting this autocrine growth loop, thereby enhancing each other's effect
Same target (different sites)	AZT (HIV-1 reverse transcriptase inhibitor[427])	Non-nucleoside HIV-1 reverse transcriptase inhibitor[428]	Antiviral synergism[429]	Isobolographic analysis, Yonetani & Theorell plot	AZT resistance is partly due to phosphorolytical removal of the AZT-terminated primer[430], NNRTI inhibited RT catalyzed phosphorolysis, thereby reduced AZT resistance[429]

Appendix Table S2 Literature reported pharmacodynamically synergistic drug combinations due to complementary actions, in which synergy has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed.

Combination target relation	**Drug A (mechanism of actions related to synergy)**	**Drug B (mechanism of actions related to synergy)**	**Reported synergistic effect**	**Synergism determination method**	**Possible mechanism of synergism in promoting complementary actions**
Different targets of the related pathways that regulate the same targets	17-AAG (Inhibited Hsp90/FLT3[480], which degraded ALK and Akt, dephosphorylated ERK, downregulated cyclin D1, CDK4, and CDK6 in ALCL cells[388], heat-shock protein antagonist, induced cell cycle inhibition and apoptosis by inhibiting NF-kappaB, AP-1 and PI3K/Akt pathways[376])	U0126 (MEK1/2 inhibitor)	17-AAG synergizes with U0126 in ALCL irrespective of ALK expression[388]	Combination index and isobologram from Chou-Talalay method (Calcusyn by Biosoft)	Activated ERK promotes ALCL cell survival. HSP90 is abundantly expressed in ALCL cells. 17-AAG produced its effect on ALCL cells by inhibiting Hsp90/FLT3 which dephosphorylated ERK [388]. Such an action is complemented by U0126's inhibition of MEK1/2 which inhibited ERK[388]
	ABT-737(Bcl-2 family proteins Bcl-2, Bcl-xL inhibitor[794])	Dexamethasone (down-regulatied Bcl-2 and Bcl-xL[795])	Synergistic effect in inducing myeloma cell death[794]	Combination Index	ABT-738's inhibition of Bcl-2 and Bcl-xL[794] is complemented by dexamethasone 's down-regulation of Bcl-2 and Bcl-xL[795]
	Azithromycin (hindered bacterial protein synthesis by binding to 50S component of 70S ribosomal subunit[431])	Ceftazidime (blocked penicillin-binding proteins and thus bacterial cell wall synthesis[432])	Synergistic antibacterial effect[433]	Checkerboard method, fractional inhibitory concentration	Hindered protein synthesis by azithromycin[431] reduces penicillin-binding proteins to complement ceftazidime's blocking of penicillin-binding proteins[432]
	Bortezomib(inhibited proteasome and NF-kappaB[796])	Trastuzumab (herceptin) (anti-HER-2/neu antibody[563])	Synergistic apoptosis effect in HER-2 positive breast cancer cell lines[797]	Combination Index	Bortezomib's inhibition of NF-kappaB[796] is complemented by herceptin's inhibition of HER-2 receptor[797] that subsequently blocks EGF-induced NF-kappaB activation[798]
	Gleevec (selective inhibitor of c-Abl, p210bcr-abl, c-Kit, and PDGF-R tyrosine kinases[779,780])	Histone deacetylase inhibitor (promoted histone acetylation and chromatin structure relaxation[764]; down-regulated Bcr-abl, c-Myc and HDAC3[765])	Synergistically induced apoptosis in STI571-resistant K562 and LAMA 84 cells[799]	Combination index	Gleevec 's pro-apoptotic inhibition of Abl[780,800] may be partially complemented by Histone deacetylase inhibitor's down-regulation of Bcr-abl[765]
	Rapamycin or deforolimus (mTOR inhibitor[775];targeted transcription	3-BrOP (inhibited glycolysis by inactivating hexokinase, a	Synergistically impacted energy	Combination Index	mTOR inhibition by rapamycin further compromised the ability of cells to uptake glucose when the

	factor yy1 to down-regulate mitochondrial transcriptional regulators PGC-1alpha, oestrogen-related receptor alpha and nuclear respiratory factors[801] (18046414)	key enzyme in the glycolytic pathway[802])	metabolism in cancer cells[802]		glycolytic pathway is inhibited by 3-BrOP[802], which is partly due to the down-regulation of nuclear respiratory factors by 3-BrOP[801] that down-regulated glycolytic and mitochondrial oxidative proteins[803]
	Triclosan (E.coli fabI inhibitor, antimicrobial activity[348])	Antisense drug Ec106fabI targeting mRNA of E.coli fabI[348]	Some combinations of protein inhibitor and antisense drug of shared genetic target satisfy conservative definition of antimicrobial synergy[348]	Checkerboard method, fractional inhibitory concentration indices	Joint inhibition and reduction of bacterial protein
	Celecoxib (COX-2 inhibitor, inactivated protein kinase Akt to stop its suppression of apoptosis, it also inhibited ER Ca2+ ATPase[397])	Emodin (tyrosine kinase inhibitor [398], down-regulated protein kinase Akt via inhibition of components of the PI3K pathway to reduce AKT suppression of apoptosis[399])	Synergistically suppressed growth of certain tumor cells[396]	Isobolographic analysis, fractional inhibition method, Zhang method	In addition to its antitumor activity via tyrosine kinase inhibition, emodin down-regulated Akt[399] to complement celecoxib's inactivation of Akt[397] to reduce Akt's suppression of apoptosis
Different targets of the related pathways that regulate the same process	17-DMAG (Inhibited Hsp90, which prevented stabilization of "client" cancer targets such as mutated p53, Raf-1, ErbB2, and other signaling proteins[804], thereby induced apoptosis and growth arrest in certain carcinoma cells[805]. Attenuated STAT3 and phospho-ERK level[806]	Arsenic trioxide (inhibited thioredoxin reductase leading to apoptosis, which is the basis for its anticancer activity[807], down-regulated constitutive STAT3 activity in AML cells[804]	ATO and Hsp90 inhibitor 17-DMAG showed synergistic interactions in inhibiting constitutive STAT3 activity and inducing cell death, in spite of a concurrent synergistic up-regulation of HSP70[804]	Isobologram	Both drugs complement each other's activity by inducing apoptosis via Hsp90[804] and thioredoxin reductase inhibition[807]. Moreover, both drugs downregulated the constitutive STAT3[804,806], which are overexpressed in 50% of AML cases.
	Aplidin (induced apoptosis by activating and clustering death receptors FasL[434], activating JNK, EGFR, Src, and	Cytarabine (DNA binder[437], inhibited synthesome associated DNA polymerase alpha activity[438], inhibited	Aplidin synergizes with cytarabine in exhibiting anticancer activities in leukemia	Chou-Talelay combination index (Calcusym Biosoft)	Both drugs complement each other's activity by inducing apoptosis via each of the two major cascades of apoptosis pathway, aplidin activated and clusterd death receptors of FasL [434] which subsequently

	p38MAPK[435], inhibited VEGF release and secretion[436])	RNA synthesis and DNA repair that lead to increased cellular stress and reduced survival protein Mcl-l which subsequently activate caspase and apoptosis [439])	and lymphoma models in vitro and in vivo[440]		activates the receptor-mediated extrinsic cascade[441], cytarabine increased cellular stress and reduced survival protein Mcl-1[439] which subsequently triggers the mitochondrial intrinsic cascade[441].
Different targets of the same pathway that regulate the same target	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis [415] and induce p53 and CDK inhibitors[416])	Lonafarnib (farnesyl transferase inhibitor, inhibited Ras farnesylation, microtubule associated alpha-tubulin deacetylase[394], and P-gp[808], metabolized by CYP3A4 and CYP3A5[809])	Synergistically inhibited deacetylating activity of tubulin decaetylase[810]	Thin plate spline method	Both drugs complement each other's microtubule stabilization effects through enhanced acetylation activity of alpha-tubulin by paclitaxel[395] and reduced deacetylation activity of alpha-tubulin deacetylase by lonafarnib[394]
	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	Tubacin (histone deacetylase 6 inhibitor, inhibited microtubule associated alpha-tubulin deacetylase activity[442])	Synergistically enhanced tubulin acetylation[394]	Combination index (Calcusym)	Both drugs complement each other's microtubule stabilization effects through enhanced acetylation activity of alpha-tubulin by paclitaxel[395] and reduced deacetylation activity of alpha-tubulin deacetylase by tubacin[442]
	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	Trichostatin (histone deacetylase inhibitor, inhibited microtubule associated alpha-tubulin deacetylase activity [394], acetylated core histones at PTEN promoter thereby induced PTEN transcription leading to enhanced apoptosis[811], induced cyclin-dependent kinase inhibitor p21[812] thereby induced G1 arrest and blocked entry into S phase[813])	Synergistic effects on apoptosis and microtubule stabilization[814]	Individual/combination response preprocessing and comparative analysis	These drugs complement each other by two actions. One jointly promotes apoptosis by triggering it via aberrant mitosis (paclitaxel)[415] and by enhancing it via upregulating PTEN (trichostatin)[811]. The other involves microtubule stabilization by enhanced acetylation activity of alpha-tubulin (paclitaxel)[395] and reduced deacetylation activity of alpha-tubulin deacetylase (trichostatin [394]
Different targets of	5-AZA-2'-deoxycytidine (DNA	Fluorouracil (metabolite	Synergistic effect of		5-AZA-2'-deoxycytidine inhibition of DNA

related pathways that regulate the same target	methyltransferase -1 and -3B inhibitor, stopped silencing of the pro-apoptotic BI [387])	inhibited thymidylate synthase that stopped DNA synthesis[550], stabilized and activated P53 by blocking MDM2 feedback inhibition through ribosomal proteins[551])	5-Aza-2'-deoxycytidine and 5-fluorouracil on drug-resistant tumors		methyltransferase -1 and -3B stopped silencing of tumor suppressor gene, pro-apoptotic BIK, in cancer cells[387]. Fluorouracil stabilized and activated P53 [551], activation of P53 upregulated BIK[815] which complements 5-AZA-2'-deoxycytidine' un-silencing of BIK
	5-AZA-2'-deoxycytidine (DNA methyltransferase -1 and -3B inhibitor, stopped silencing of the pro-apoptotic BIK[387])	Depsipeptide (histone deacetylase inhibitor, induced the pro-apoptotic BIK[387])	Synergistic antineoplastic effect[816]	Valeriote & Lin's comparative analysis method	5-AZA-2'-deoxycytidine inhibition of DNA methyltransferase -1 and -3B stopped silencing of tumor suppressor gene, pro-apoptotic BIK, in cancer cells to complement depsipeptide's induction of the same gene[387]
	Gefitinib (EGFR tyrosine kinase inhibitor, induced cyclin-dependent kinase inhibitors p27 and p21, decreased MMP-2 and MMP-9 enzyme activity[422])	ST1926 (activated MAP kinases p38 and JNK, released cytochrome c, activated caspase proteolytic cascad[443])	Synergistic modulation of survival signaling pathways[444]	Combination Index	Gefitinib 's inhibition of EGFR is complemented by ST1926's activation of MAP kinases p38[443] that subsequently mediates internalization of EGFR[445], and by ST1926's activation of caspase proteolytic cascade[443]
Different targets of related pathways	Fluorouracil (metabolite inhibited thymidylate synthase that stopped DNA synthesis[550], stabilized and activated P53 by blocking MDM2 feedback inhibition through ribosomal proteins[551])	RPR-115135 (farnesyl transferase inhibitor, inhibited Ras farnesylation[817])	Synergistic cytotoxic effect[818]	Combination index	Joint tumor suppressive (via fluorouracil stabilization of P53[551]) and antiproliferative (via RPR-115135 inhibition of Ras farnestlation[817]) actions
	CP55940 (cannabinoid agonist, elicited analgesic effects in acute and chronic pain states via spinal and supraspinal pathways[391])	Dexmedetomidine (alpha2 adrenoceptor agonist, activated endogenous nonrapid eye movement sleep-promoting pathways[454])	Significant antinociception synergy in some cases	Isobolographic analysis	Cannabinoid agonist modulated spinal and supraspinal pathways[391] that regulate pain[390] , dexmedetomidine promoted sleepiness[454] that sustains reduction in spike activity of spinoreticular tract neurons[389]
	Dipropofol (inhibited bacterial protein synthesis or amino acid incorporation[819])	Vancomycin (blocked transglycosylation and transpeptidation reactions in polymerization of bacterial cell wall peptidoglycan, thereby inhibited cell wall	Synergism against vancomycin resistant bacterial strains[819]	Checkerboard method	Hindered protein synthesis by dipropofol[819] might reduce cell-wall synthesis proteins and thus complement vancomycin's inhibition of cell wall

		biosynthesis[820])			
	Tipifarnib (farnesyl transferase inhibitor, inhibited Ras farnesylation, upregulated death receptor 5, a p53 target gene and receptor of TRAIL[821], inhibited P-gp[822])	Zoledronic acid (activated caspase 3 and fragmented PARP to induce apoptosis, reduced Ras activity and antagonized its stimulation by EGF[823])	Strong synergism in growth inhibition and apoptosis[823]	Combination index (Calcusym)	Joint anti-growth activities via tipifarnib inhibition of Ras farnesylation[821] and zoledronic acid reduction of Ras activity[823], joint apoptosis activities via tipifarnib upregulation of death receptor 5[821] and zoledronic acid activation of caspase 3[823], upregulation of death receptor helps to activate capspases[824])
	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395] distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416], activated caspase-10, caspases-8, -6, and -3, leading to apoptosis[417], activated ERK[418] which in turn activates CDK2[419], activated p38 MAP kinase and p53[420])	NU6140 (CDK inhibitor, down-regulated antiapoptotic protein survivin[421])	Synergistic apoptotic response[421]	Median drug effect analysis	Use of both drugs promoted complementary apoptosis activities via triple actions of surviving down-regulation by NU6140[421], microtubule stabilization[395] and caspase activation[417] by paclitaxel. Paclitaxel's promotion of apoptosis may be partially offset by its counteractive pro-growth activation of ERK [418] and CDK2[419], which may be partially reduced by NU6140 via its inhibition of CDK[421]
	Sildenafil (phosphodiesterase-5 inhibitor[446])	Iloprost (prostacyclin receptor agonist leading to vascular relaxation[447], activated phospholipase C [448] that promoted VEGF-induced skin vasorelaxation [449], self-regulated endothelial cell adhesion molecules[450])	Synergistic action to cause strong pulmonary vasodilatation[451]	Dose effect curve surpassed that of individual drug alone combined	Sildenafil produced vasodilation activity by inhibiting phosphodiresterase-5[446], iloprost produced vasodilation activity by agonizing prostacyclin receptor[447] and by activating phospholipase C[448]. Targeting of multiple vasodilatation regulation pathways NO/cGMP[452], MaxiK channel -mediated relaxation[453], and phospholipase C[448] contribute to the synergistic actions.
Different target subtypes of related pathways	Dexmedetomidine (alpha2A receptor agonist, produced antinociceptive effect via an endogenous sleep-promoting pathway[454])	ST-91 (agonist of alpha2 receptor of other subtypes, produced antinociceptive effect via upraspinal receptors and at both spinal and brainstem levels of the acoustic startle response pathway[455])	Synergistic antinociceptive action[345,456]	Isobolographic analysis	ST-91 modulated spinal and supraspinal pathways[455] that regulate pain[390] , dexmedetomidine promoted sleepiness[454] that sustains reduction in spike activity of spinoreticular tract neurons[389]

Same target (different states)	Mycophenolate mofetil (inosine monophosphate dehydrogenase inhibitor, drug metabolite mycophenolic acid binds to the site of nicotinamide adenine dinucleotide cofactor[393])	Mizoribine (inosine monophosphate dehydrogenase inhibitor, drug metabolite mizoribine monophosphate binds to the enzyme in transition state having a new conformation[457])	Mild synergistic suppression of T and B cell proliferation[458]	Median drug effect analysis, Combination index	Simultaneous inhibition of enzyme reactant-state and transition state have the advantage of covering more conformational space for the inhibitors to better compete with natural substrates for the binding sites.
Same target (overlapping binding sites)	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	Discodermolide (stabilized microtubule dynamics and enhanced microtubule polymer mass[459] resulting in aberrant mitosis that triggers apoptosis [415] and induced p53 and CDK inhibitors[416], retained antiproliferative activity against carcinoma cells resistant to paclitaxel due to beta-tubulin mutations[460])	Antiproliferative synergy[461]	Combination index	Explanation 1: Binding sites of both drugs overlapping, certain mutations resistant to one drug are ineffective against the other, thereby covering more diverse range of mutant types[60,340,462]. Explanation 2: One drug binds and induces conformational change in tubulin that increases the binding affinity of the other[60,463]. Explanation 3: These drugs may differentially bind to or affect different tubulin subtypes, microtubule architectures, or microtubule regulators, thereby covering more diverse range of microtubule dynamics[57,60,463,464]
Same target (different binding sites)	Paclitaxel (stabilized microtubules via alpha-tubulin acetylation[395], distorted mitosis to trigger apoptosis[415] and induce p53 and CDK inhibitors[416])	Peloruside A (binds at a different site from that of paclitaxel, stabilized microtubules via binding to a unique site on the tubulin alpha, beta heterodimer[465])	Peloruside A synergizes with paclitaxel to enhance the antimitotic action of the drugs[465]	Berenbaum's combination index	Explanation 1: Binding sites of both drugs overlapping, certain mutations resistant to one drug are ineffective against the other, thereby covering more diverse range of mutant types[60,340,462]. Explanation 2: One drug binds and induces conformational change in tubulin that increases the binding affinity of the other[60,463]. Explanation 3: These drugs may differentially bind to or affect different tubulin subtypes, microtubule architectures, or microtubule regulators, thereby covering more diverse range of microtubule dynamics[57,60,435,463]
	Paclitaxel (external DNA binder with partial helix stabilization without altering B-form, binds to A-T, G-C bases and the backbone PO(2) groups[392], interacted	Trabectedin (formed DNA adduct at the central G in minor groove of pyrimidine-G-G and purine-G-C triplets that stabilizes duplex DNA to	Synergistic cytotoxicity[830]	Isobolographic analysis, Chou-Talalay equation	Both drugs enhance each other's effect by two actions: (1) binding to different sites of DNA at mutually compatible conformation, thereby complement each other on their blocking of DNA polymerase and transcription processes[385,392], (2) these bindings

	with DNA topoisomerase I[825])	hamper strand separation and stall replication and transcription forks[385], induced topoisomerase I mediated protein-linked DNA breaks[826], traped protein from the nucleotide-excision repair system resulting in DNA damage[827], induced transient p53 elevation[828], and it is a P-gp substrate[829])			facilitated interaction with DNA topoisomerase I[825] and its DNA breaking actions[826].

Appendix Table S3 Literature reported pharmacodynamically synergistic drug combinations due to facilitating actions, in which synergy has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed.

Combination target relationship	**Drug A (mechanism of actions related to synergy)**	**Drug B (mechanism of actions related to synergy)**	**Reported synergistic effect [Ref]**	**Synergism determination method**	**Possible mechanism of synergism in promoting facilitating actions**
Different targets of related pathways	Ampicillin (blocked penicillin-binding protein 2A and thus bacterial cell wall synthesis[466])	Daptomycin (disrupted bacterial membrane structure[467])	Significant antibacterial synergy[347]	Checkerboard method, fractional inhibitory concentration	Most penicilling-binding proteins are associated with membrane[468], membrane disruption by daptomyci [467] likely hinders the functions of penicilling-binding proteins and further expose them to ampicillin binding
	Gentamicin (targeted bacterial ribosome, caused misreading of the genetic code and inhibited translocation, thereby disrupted protein synthesis[401])	Vancomycin (inhibited bacterial cell wall peptidoglycan synthesis[402], altered permeability of cell membrane and selectively inhibited ribonucleic acid synthesis[403])	Synergistic action against penicillin-resistant bacterial strains[400]	Checkerboard method, fractional inhibitory concentration indices	Vancomycin altered membrane permeability[403] thereby enhanced gentamicin penetration into bacterial cells and its bioavailability
	Daptomycin (disrupted bacterial membrane function without penetrating into the cytoplasm[831], depolarized membrane[832], and inhibited lipoteichoic acid synthesis[833])	Rifampicin (interfered with bacterial nucleic acid synthesis by binding to the beta subunit of prokaryotic RNA polymerases[834])	Significant antibacterial synergy[347]	Checkerboard method, fractional inhibitory concentration	Depolarization of bacterial membrane by daptomycin[832] enhanced rifamapicin penetration into bacterial cells and thus its bioavailability
Different targets of related pathways that regulate the same target	BQ-123 (Endothelin A receptor antagonist, mediated vasodilatation[405])	Enalapril (angiotensin converting enzyme inhibitor, up-regulated Endothelin B[406], vasodilation is mediated by both ACE inhibition[835] and Endothelin B1 upregulation[407])	Synergistic endothelium-dependent vasodilation enhancing actions[404]	Randomized, double-blind, crossover studies	Enalapril up-regulated ETB as well as inhibited ACE leading to vasodilation[406,407] , BQ-123 antagonism of ETA caused vasodilation[405] and displaced endogenous ET-1 from ETA onto upregulated ETB to enhance its activity by effectively increasing ETB agonist concentration[404]
	Candesartan-cilexetil (angiotensin AT1 receptor antagonist[469])	Ramipril (angiotensin converting enzyme inhibitor[470], reduced angiotensin II formation[471])	Synergistically reduced systolic blood pressure[472]	Dose-response curve shifted 6.6-fold leftwards compared to hypothetic additive curve	Candesartan-cilexetil reduced systolic blood pressure by antagonizing angiotensin AT1 receptor[469], ramipril reduced systolic blood pressure by inhibiting angiotensin converting

					enzyme[470], ramipril inhibited AT1 receptor agonist formation[471] thereby facilitating the action of candesartan-cilexetil by reducing AT1 agonist-antagonist competition
Same target	Saquinavir (HIV protease inhibitor[836], Pgp substrate[837])	Lopinavir (HIV protease inhibitor (9835517), inhibited Pgp in CACO-2 cells[837])	Synergistic inhibition of HIV1 replication in MT4 cells[837]	Combination indices	As a Pgp substrate, HIV protease inhibitor saquinavir may be removed by Pgp mediated drug efflux, making it less available for HIV protease inhibition[837]. Inhibition of Pgp by another HIV protease inhibitor lopinavir facilitates the therapeutic action of saquinavir by blocking its efflux[837], leading to synergistic action.

Appendix Table S4 Literature reported pharmacodynamically additive drug combinations, in which additive action has been determined by well established synergy/additive analysis methods and its molecular mechanism has been revealed

Action type	**Combination target relationship**	**Drug A (mechanism of action related to additive effect)**	**Drug B (mechanism of action related to additive effect)**	**Reported additive effect [Ref]**	**Additism determination method**	**Possible mechanism of additive effect**
Equivalent or overlapping actions	Different targets of the same pathways that regulate the same target	Citicoline (increased BCL-2 expression[838], regulated Bcl-2 antiapoptotic pathway for neuron cell survival [839]. Bcl-2 interacted with IP(3) receptor Ca(2+) channels on ER, regulating their opening in response to IP(3) and thus inhibiting IP(3)-mediated Ca(2+) signals that induce apoptosis[840])	Nimodipine (L-type voltage sensitive calcium channel blocker [838], reduced Ca(2+) influx to rescue cortical neurons from apoptosis[841])	Additive neuroprotective effect[838]	Measurement of infarct volume	Citicoline upregulated Bcl-2[838], thereby enhanced inhibition of IP(3)-mediated Ca(2+) signals which reduced apoptosis[840].Nimodipine reduced Ca(2+) influx to rescue cortical neurons from apoptosis[841].
		CP55940 (cannabinoid antagonist[842], coupling to postsynaptic GIRK2 potassium channels[843])	Dexmedetomidine (alpha2 adrenoceptor agonist[844], coupling to postsynaptic GIRK2 potassium channels[845])	Additive antinociceptive actions in some cases and synergistic actions in other cases[842]	Isobolographic analysis	Both produced therapeutic actions via coupling to postsynaptic GIRK2 potassium channels[843,845]
		CP55940 (cannabinoid antagonist[842], coupling to postsynaptic GIRK2 potassium channels[843])	Morphine (mu opioid receptor agonist[846], coupling to postsynaptic GIRK2 potassium channels[847])	Additive antinociceptive actions in some cases and synergistic actions in other cases[842]	Isobolographic analysis	Both via their coupling to postsynaptic GIRK2 potassium channel [843,847]
		Diazoxide (ATP-sensitive K+ channel activator[481], enhanced ATPase activity of channel regulatory subunits sulphonylurea	Sodium nitroprusside (activated ATP-sensitive K+ channel[481], acted as the donor of nitric oxide which subsequently opened	Additive antinociceptive effect[485]	ANOVA synergism & dose effect data analysis	Diazoxide enhanced ATPase activity of channel regulatory subunits[482], Sodium nitroprusside opened the channel by acting as the donor of nitric oxide[848].

		receptors[482])	channel[848])			
		Diazoxide (ATP-sensitive K+ channel opener[481], enhanced ATPase activity of channel regulatory subunits sulphonylurea receptors[482])	Dibutyryl-cGMP (activated ATP-sensitive K+ channel[481], activated channel via a cGMP-dependent protein kinase[483, 484])	Additive antinociceptive effect[485]	ANOVA synergism & dose effect data analysis	Diazoxide enhanced ATPase activity of channel regulatory subunits[482], Dibutyryl-cGMP activated channel via a cGMP-dependent protein kinase [483,484]
		Methylene blue (inhibited heme polymerization[751]; selectively inhibited glutathione reductase[752]; soluble guanylate cyclase inhibitor[753]	Mefloquine and quinine (heme polymerase inhibitor[849])	Additive antimalarial effect in chloroquine-sensitive and -resistant P. falciparum strains[754]	Isobolographic analysis	Both acted redundantly at the heme polymerization pathway
		Retinoic acid (activated and up-regulated retinoic acid receptor beta, a tumor suppressor that promote apoptosis[850])	Trichostatin A (histone deacetylase inhibitor, reactivated retinoic acid receptor beta mRNA expression[851])	Additive inhibition of cell proliferation[473]	ANOVA synergism & dose effect data analysis	Retinoic acid activated and up-regulated the target[850], Trichostatin A up-regulated the target[851]
	Same target (different sites with direct contact with agonist site)	Propofol (interacted with GABA A receptor, acting on at TM3 segment of the beta2 subunit[486])	Sevoflurane (interacted with GABA A receptor at Ser270 of the alpha1 and alpha2 subunits[487])	Additive action in producing consciousness and movement to skin incision during general anesthesia[488]	Dixon up-down method	Propofol binds to TM3 segment of the beta2 subunit[486], Sevoflurane binds to Ser270 of the alpha1 subunit[487]. As agonist binding site is located between alpha1 and beta2 subunits[489], both drugs likely hinder agonist activity, thereby producing mutually substitutable actions.
	Same target (same site):	Ampicillin (blocked penicillin-binding protein 2A and thus bacterial cell wall synthesis)[466]	Imipenem (inhibited penicillin-binding protein -1A, -1B, -2, -4 and -5 and thus bacterial cell wall synthesis)[490]	Additive antibacterial effect[347]	Checkerboard method, fractional inhibitory concentration	Both acted at the same active site of penicillin-binding protein 2A[491] but at relatively high MICs of ≥32μg/ml [466]. The relatively high MICs make it less likely for both drugs to saturate target

						sites, thereby maintaining additive antibacterial effect.
Independent actions	Different targets of unrelated pathways	Anidulafungin (inhibitor of beta-(1,3)-dglucan synthase, an essential component of fungal cell wall)[852]	Amphotericin B (formed ion channels in fungal membranes)[495]	Additive antifungal effect[853]	Checkerboard method, fractional inhibitory concentration	Anidulafungin disrupted cell wall formation during reproductive cycle[852], Amphotericin B disrupted membranes of mature fungi[495]. They act at different stages and different sites in non-interfering manner.
		Artemisinin (interfered with parasite transport proteins PfATP6, disrupted parasite mitochondrial function, modulated host immune function)[492]	Curcumin (generated ROS and down-regulated PfGCN5 histone acetyltransferase activity, producing cytotoxicity for malaria parasites)[261]	Additive antimalarial activities[493]	Fractional inhibitory concentrations	Artemisinin blocked calcium transport to endoplasmic reticulum[492], Curcumin induced DNA damage and histone hypoacetylation[261]. They act at different sites in non-interfering manner.
	Same target (different sites)	Doxorubicin (DNA intercalator[475], preferred AT regions[475])	Trabectedin (formed covalent guanine adduct at specific sites in DNA minor grove[476], interacted with DNA repair system)	Additive anticancer effect[474]	Isobolographic analysis	Both bind to DNA in non-interfering manner, one preferred AT regions[475], the other alkylated guanines[476]. Recent progresses in designing dual platinum- intercalator conjugates[477] suggested that it is possible for both drugs to act without hindering each other's binding mode
Independent actions at dosages significantly lower than MICs, complementary actions at higher dosages	Different targets of unrelated pathways	Azithromycin (hindered bacterial protein synthesis by binding to 50S component of 70S ribosomal subunit[431])	Imipenem (inhibited penicillin-binding protein -1A, -1B, -2, -4 and -5 and thus bacterial cell wall synthesis)[490]	Additive antibacterial effect[433]	Checkerboard method, fractional inhibitory concentration	Azithromycin hindered bacterial protein synthesis[431] at MIC of 0.12μg/ml[494]. Imipenem blocked bacterial cell wall formation[490] at MICs of ≥32μg/ml[466]. At dosages significantly lower than MICs for both drugs, azithromycin's reduction of penicillin-binding proteins[490] may be insufficient for imipenem to saturate these proteins, allowing its unhindered inhibition of these proteins[490], thereby these actions proceed in non-interfering manner

Appendix Table S5 Literature reported pharmacodynamically antagonistic drug combinations in 2000-2006, in which antagonism has been determined by established methods and its molecular mechanism has been revealed. The antagonism of the listed drug combinations is due to interfering actions of the partner drugs in each combination.

Combination target relation	**Drug A (mechanism of action related to antagonism)**	**Drug B (mechanism of action related to antagonism)**	**Reported antagonistic effect [Ref]**	**Antagonism determination method**	**Possible mechanism of antagonism of interfering actions**
Different targets of related pathways that regulate the same target	17-AAG (heat-shock protein antagonist, induced cell cycle inhibition and apoptosis by inhibiting NF-kappaB, AP-1 and PI3K/Akt pathways[376], Hsp90/FLT3 inhibitor[480])	Cytarabine (DNA binder [437], inhibited synthesome associated DNA polymerase alpha activity[438])	17-AAG antagonized the cytotoxic activity of cytarabine[437]	Median dose effect analysis (Calcusym)	17-AAG altered the condition necessary for cytarabine to produce its activity. It induced G1 arrest, which subsequently prevented cytarabine incorporation into cellular DN[437]
	Amphotericin B (formed ion channels in fungal membranes[495])	Ravuconazole (inhibited biosynthesis of ergosterol, a component of fungal cell membrane[496])	Antagonism in experimental invasive pulmonary aspergillosis[497,498]	Loewe additivity-based drug-interaction model	Amphotericin B can form ion channels more easily in the presence of ergosterol [495], ravuconazole inhibition of ergosterol synthesis[496] can therefore reduce the activity of amphotericin B in forming ion channels[495]
Same target	4-HPR (Reduced ERK activity to inhibit prostate cancer cells[854], activated JNK to induce apoptosis[855], suppressed IκBα phosphorylation which inhibited NF-κB activation and downregulated antiapoptotic genes[854])	Sodium butyrate (Reduced ERK activity to inhibit prostate cancer cells[854], activated JNK to induce apoptosis[855] suppressed IκBα phosphorylation which inhibited NF-κB activation and downregulated antiapoptotic genes[854])	Sodium butyrate and 4-HPR administered together antagonize effects of each other on prostate cancer[854]	Isobologram	Co-administration of 4-HPR and sodium butyrate possibly affected each other's actions in suppressing IκBα phosphorylation, thereby reduced their inhibitory effects on NF-κB activation and antiapoptotic gene expression. NF-κB activation also downregulated JNK phosphorylation leading to inhibition of apoptosis in prostate cells[854].

	Aminophylline (adenosine receptor antagonist, phosphodiesterase inhibitor, released intracellular calcium[478])	Theophylline (released intracellular calcium, adenosine receptor antagonist, phosphodiesterase inhibitor[478])	Antagonism of inhibitory adenosine autoreceptors and release of intracellular calcium[478]	Quantal release measurement	Adenosine receptor antagonist binding may be associated with non-unique binding site conformations [479]. Aminophylline binding may lock the receptor into a unique conformation that hinder theophylline binding, thereby producing antagonistic effect

Appendix Table S6 Literature reported pharmacokinetically potentiative drug combinations, in which potentiative effect has been determined by established methods and its molecular mechanism has been revealed

Biochemical class of potentiative effect	**Drug A (therapeutic or toxic effects and mechanism of actions)**	**Drug B (mechanism of action related to potentiative effect)**	**Reported potentiative effect [Ref]**	**Possible mechanism of potentiative actions**
Positive regulation of drug transport or permeation	AZT (anti-HIV, HIV-1 reverse transcriptase inhibitor)	1,8-Cineole (formed hydrogen bonds with lipid head groups of stratum corneum lipids[502])	Enhanced cross-skin permeation of AZT[503]	Enabled drug transport across skin possibly by disrupting absorption barrier via binding to lipid head groups
	Ciprofloxacin (antibacterial, inhibited DNA gyrase, an enzyme specific and essential for all bacteria)	Gatifloxacin (inhibited efflux pump of ciprofloxacin[856])	Synergistic antibacterial action against pseudomonas aeruginosa via efflux pump inhibition[856]	Avoided drug excretion by inhibiting efflux pump of ciprofloxacin
	Doxorubicin (anticancer by DNA intercalation)	HPMA copolymer (formed conjugate with anthracycline[857])	Polymer anthracycline conjugation enabled bypass of multi-drug resistance[857]	Enabled drug absorption by avoiding its efflux via drug-polymer conjugate formation
	Fexofenadine (nonsedating antihistamine, H1-antagonist, renal uptake by hOAT3 transporter)	Probenecid (inhibited hOAT3 transporter uptake of fexofenadine[858])	Increased plasma concentration of fexofenadine due to inhibition of its renal elimination[859]	Avoided drug excretion by inhibiting hOAT3 transporter uptake of fexofenadine
	Levodopa (dopaminergic agent in Parkinson's disease)	Fatty acid synthesis inhibitor (selective inhibition of fatty acid synthesis delays barrier recovery rates after barrier perturbation of drugs[860])	Enhanced transcutaneous delivery of levodopa[860-862]	Delayed recovering of drug transport barrier by inhibiting the synthesis of barrier components
	Low molecular weight heparin (antithrombotic, antithrombin binder to inhibit activated coagulation factors)	Chitosan (absorption enhancer, reversibly interacted with components of tight junctions, leading to widening of paracellular routes and increased permeability of peptide drugs across	Oral drug absorption enhancement[499]	Disrupted drug absorption barrier across mucosal epithelia via interaction with barrier components

		mucosal epithelia[499])		
	Low molecular weight heparin (antithrombotic, antithrombin binder to inhibit activated coagulation factors)	Sodium caprate (absorption enhancer, released calcium from intracellular stores via activation of phospholipase C in plasma membrane, which induced contraction of calmodulin-dependent actin microfilament, followed by dilatation of the paracellular route[863])	Sodium caprate acts as a relatively safe and efficient absorption enhancer of low molecular weight heparin[863-865]	Disrupted drug absorption barrier at plasma membrane
Enhanced drug distribution or localization	Fluorouracil (anticancer, metabolized by thymidine phosphorylase and others, metabolite inhibited thymidylate synthase that stopped DNA synthesis[550], metabolite stabilized P53 due to RNA-directed effects)	Sorivudine (antiviral, metabolized into (E)-5-(2-Bromovinyl)uracil by thymidine phosphorylase, which subsequently inactivates this enzyme by irreversible binding[866])	Enhanced toxic activity of fluorouracil by elevating its plasma concentrations[866]	Enhanced level of drug in plasma by metabolism and uptake inhibition
	Cerivastatin (cholesterol-lowering, HMG-CoA reductase inhibitor)	Gemfibrozil (inhibited CYP2C8 mediated metabolism of statins, inhibited OATP2 mediated uptake of cerivastatin[504])	Increased plasma concentration of statins by inhibiting their metabolism and uptake[504-506]	Enhanced level of drug in plasma by metabolism and uptake inhibition
	Cyclosporine (immunosuppressive via calcineurin antagonism, induced nephrotoxicity by decreasing renal blood flow, generating reactive free radicals, and inducing vasoconstriction and apoptosis)	Sirolimus (increased cyclosporine concentrations in whole blood and, particularly, in kidney[867])	Exacerbated renal dysfunction by cyclosporine[867]	Enhanced level of drug in tissue of toxic action
	HSV thymidine kinase gene and ganciclovir (anticancer gene therapy drug combination)	Ponicidin (significantly accumulated the phosphorylated metabolites of ganciclovir and suppressed the extracellular release of ganciclovir[868])	Potentiated ganciclovir cytotoxicity[868]	Enhanced level of prodrug metabolite
	Methamphetamine (psychomotor stimulant by targeting biogenic amine transporters)	D-chlorpheniramine (increased plasma and brain tissue concentrations of methamphetamine[869])	Significantly potentiated methamphetamine-induced psychomotor activation[869,870]	Enhanced level of drug in tissue of therapeutic action
	HSV thymidine kinase gene and ganciclovir (anticancer gene therapy	Scopadulciol (stimulated HSV thymidine kinase activity, increased	Improved efficacy of cancer gene therapy via enhanced activity and increased level	Enhanced level of prodrug metabolite

	drug combination)	levels of ganciclovir metabolite[871])	of prodrug metabolite[871]	
	Rosuvastatin (decreased levels of atherogenic lipoproteins in patients with or at high risk of cardiovascular disease, HMG-CoA inhibitor)	Cyclosporine (Inhibited OATP-C mediated uptake, rosuvastatin is a substrate of this process[872])	Significantly increased systemic exposure of rosuvastatin[872]	Enhanced level of drug by inhibiting drug uptake
	Rosuvastatin (decreased levels of atherogenic lipoproteins in patients with or at high risk of cardiovascular disease, HMG-CoA inhibitor)	Gemfibrozil (Inhibited OATP2 mediated uptake, rosuvastatin is a substrate of this process[873])	Increased plasma concentrations of rosuvastatin[874]	Enhanced level of drug by inhibiting drug uptake
Enhanced Drug metabolism	Fluorouracil (anticancer, metabolized by thymidine phosphorylase and others [875], metabolite inhibited thymidylate synthase that stopped DNA synthesis[550], metabolite stabilized P53 due to RNA-directed effects[876])	2'-deoxyinosine (modulator that enhances thymidine phosphorylase activity[500])	Enhanced antitumor activity of fluorouracil in human colorectal cell lines and colon tumor xenografts[500]	Enhanced metabolism of prodrug into active metabolite
	Doxorubicin (anticancer by DNA intercalation, converted to doxorubicinol by NADPH-dependent aldo/keto or carbonyl reductases [507], which produced cardiotoxicity by mediating transition from reversible to irreversible damage)	Paclitaxel (stimulated enzymatic activity of NADPH-dependent aldo/keto or carbonyl reductases[507])	Enhanced cardiotoxicity by increasing metabolism of doxorubicin into toxic metabolit[507]	Enhanced metabolism of drug into toxic metabolite
	Tirapazamine and cisplatin (Tirapazamine produced anti-cancer effect and potentiated cisplatin anticancer activities when metabolized by P450R into toxic free radical[877], cisplatin is a DNA adduct)	Adenoviral delivery of human P450R genes (enhanced tirapazamine metabolizing enzyme expression and metabolism of prodrug into active metabolite[877])	Targeted gene prodrug therapy increased efficacy of tirapazamine[877]	Enhanced metabolism of prodrug into active metabolite

Appendix Table S7 Literature reported pharmacokinetically reductive drug combinations, in which reductive effect has been determined by established methods and its molecular mechanism has been revealed.

Biochemical class of reductive effect	Drug A (therapeutic or toxic effects and mechanism of actions)	Drug B (mechanism of action related to reductive effect)	Reported reductive effect [Ref]	Possible mechanism of reductive actions
Drug transport and permeation	Amphotericin B (antileishmanial, formed aggregate with miltefosine[508])	Miltefosine (antileishmanial, formed aggregate with amphotericin B[508])	Reduced miltefosine-induced paracellular permeability enhancement in Caco-2 cell monolayers, inhibited uptake of both drugs, decreased transepithelial transport of both drugs[509]	Reduced drug permeability and transport
	Gamma-hydroxybutyrate (drug of abuse, increased dopamine concentration, MCT1 transporter mediated its disposition and renal reabsorption[878])	Luteolin (exhibited MCT1 transporter mediated uptake of gamma-hydroxybutyrate[879])	Significantly increased renal and total clearances of gamma-hydroxybutyrate[878]	Enhanced drug excretion
Drug distribution and localization	Cisplatin (DNA inter- and intra-strand adduct)	Procainamide hydrochloride (formed cisplatin-procainamide complex[511])	Reduced cisplatin-induced hepatotoxicity via formation of less toxic platinum complex, leading to inactivation of cisplatin or its highly toxic metabolites and to a different subcellular distribution of platinum[511]	Reduced level of toxic drug by formation of less toxic complex and rearrangement of its subcellular distribution
Drug metabolism	Warfarin (anticoagulant and antithrombotic, affected coagulation proteins that act sequentially to produce thrombin, metabolized by CYP3A4[512])	Quinidine (stimulated CYP3A4 mediated metabolism of warfarin[513])	Reduced anticoagulanet effect of warfarin by stimulating its metabolism[513]	Enhanced metabolism of active drug into inactive metabolite
	Diclofenac (anti-inflammatory, metabolized into 5-hydroxylated by cytochrome P450 CYP3A4[880])	Quinidine (stimulated CYP3A4 mediated metabolism of diclofenac[880])	Increased diclofenac clearance and reduced its plasma concentration by enhanced metabolism[880]	Reduced level of drug by enhanced metabolism
	Mycophenolate mofetil (immunosuppressive, a prodrug whose metabolite mycophenolic acid is a potent and reversible uncompetitive inhibitor of inosine monophosphate dehydrogenase, metabolized by	Rifampin (induced expression of gastrointestinal uridine diphosphate-glucuronosyltransferases[882])	Drug interaction leads to underexposure and loss of clinical efficacy of mycophenolate mofetil by induction of renal, hepatic, and gastrointestinal uridine diphosphate-glucuronosyltransferases and organic anion transporters[882]	Reduced level of drug by enhanced metabolism

	gastrointestinal uridine diphosphate-glucuronosyltransferases[881])			
	Valproic acid (antiepileptic, increased gabaergic transmission, reduced release and/or effects of excitatory amino acids, blocked voltage-gated sodium channels, modulated dopaminergic and serotoninergic transmission, metabolized into valproic acid glucuronide[883])	Carbapenem antibiotics (inhibited the hydrolytic enzyme involved in the hydrolysis of valproic acid glucuronide to valproic acid, resulting in a decrease of plasma concentration of valproic acid[859])	Caused seizures in epileptic patients due to lowered plasma levels of valproic acid[859,884]	Reduced level of drug in plasma by metabolism inhibition

Appendix Table S8 Drug related sensitizing/resistant mutations of EGFR and cancer related activating mutations of EGFR, PIK3CA, RAS, and BRAF, and inactivation mutations of PTEN.

Disease	Type of Mutation	Percentage of 85 NSCLC Cell-lines or 40 Breast Cancer Cell-lines with This Type of Mutation	Specific Mutations (Number of NSCLC or Breast Cancer Cell-lines with This Mutation)
NSCLC	Gefitinib , erlotinib , and lapatinib sensitizing mutation of EGFR[576]	11.7%	E746_A750del (4) / E746_A750del, T751A(1) / E746_T751del, I ins(1) / L747_E749del, A750P(1) / L747_S752del, P753S(1) / L858R(2)
	Gefitinib , erlotinib , and lapatinib resistant mutation of EGFR[576]	2.4%	T790M (2)
	Gefitinib and erlotinib resistant mutation of HER2[885]	1.2%	G776VC (1)
	Activating mutation of KRAS[886]	32.9%	G12A (1) / G12C (9) / G12D (3) / G12R (1) / G12S (1) / G12V (4) / G13C (2) / G13D (4) / Q61H (2) / Q61K (1)
	Activating mutation of NRAS[886]	5.9%	Q61K (3) / Q61L (1) / Q61R (1)
	Activating mutation of BRAF [887]	7.1%	G466V(1) / G469A(3) / L597V(1) / V600E(1)
	Activating mutation PIK3CA [888,889]	4.7%	E542K (1) / E545K (2) / H1047R(1)
	Inactivating mutation PTEN[890]	4.7%	H61R(1) / G251C(1) / R233*(2)
Breast Cancer	Activating mutation PIK3CA[889]	41.9%	C420R(2) / E542K(2) / E545K(2) / H1047L(1) / H1047R(6) / P539 (1)
	Inactivating mutation PTEN[891]	12.9%	A72fsX(1) / C136Y(1) / D92H(1) / V275fs*(1)

,

Appendix Table S9 Cancer related and drug related specific mutations in 85 NSCLC and 40 breast cancer cell-lines.

Cell-lines	Disease	Mutated Gene[591,592]	Type of Mutation	Mutation Details	
				Amino Acid	Nucleotide
A427	NSCLC	KRAS	Activating mutation	G12D	35G>A
A549	NSCLC	KRAS	Activating mutation	G12S	34G>A
Calu1	NSCLC	KRAS	Activating mutation	G12C	34G>T
Calu3	NSCLC	ND			
Calu6	NSCLC	KRAS	Activating mutation	Q61K	181C>A
Colo699	NSCLC	ND *			
DV90	NSCLC	KRAS	Activating mutation	G13D	38G>A
EKVX	NSCLC	ND			
H1155	NSCLC	KRAS	Activating mutation	Q61H	183A>T
H1155	NSCLC	PTEN	Inactivating mutation	R233*	697C>T
H1299	NSCLC	NRAS	Activating mutation	Q61K	181C>A
H1355	NSCLC	KRAS	Activating mutation	G13C	37G>T
H1355	NSCLC	BRAF	Activating mutation	G469A	1406G>C
H1395	NSCLC	BRAF	Activating mutation	G469A	1406G>C
H1437	NSCLC	ND			
H1563	NSCLC	PIK3CA*	Activating mutation	E542K	1624G>A
H1568	NSCLC	ND			
H157	NSCLC	KRAS	Activating mutation	G12R	34G>C
H157	NSCLC	PTEN	Inactivating mutation	G251C	751G>T
H157	NSCLC	PTEN	Inactivating mutation	H61R	182A>G
H1648	NSCLC	ND			
H1650	NSCLC	EGFR	EGFR sensitizing mutation	E746_A750del	2235_2249del15
H1666	NSCLC	BRAF	Activating mutation	G466V	1397G>T
H1734	NSCLC	KRAS	Activating mutation	G13C	37G>T

H1755	NSCLC	BRAF	Activating mutation	G469A	1406G>C
H1770	NSCLC	ND			
H1781	NSCLC	ERBB2*	gefitinib and erlotinib resistant mutation	G776VC	
H1792	NSCLC	KRAS	Activating mutation	G12C	34G>T
H1819	NSCLC	ND			
H1838	NSCLC	ND			
H1915	NSCLC	ND*			
H1944	NSCLC	KRAS*	Activating mutation	G13D	38G>A
H1975	NSCLC	EGFR	EGFR-I sensitizing mutation	L858R	2573T>G
H1975	NSCLC	EGFR	EGFR-I resistant mutation	T790M	2369C>T
H1993	NSCLC	ND			
H2009	NSCLC	KRAS	Activating mutation	G12A	35G>C
H2030	NSCLC	KRAS	Activating mutation	G12C	34G>T
H2052	NSCLC	ND			
H2077	NSCLC	ND*			
H2087	NSCLC	BRAF	Activating mutation	L597V	1789C>G
H2087	NSCLC	NRAS	Activating mutation	Q61K	181C>A
H2110	NSCLC	ND			
H2122	NSCLC	KRAS	Activating mutation	G12C	34G>T
H2126	NSCLC	ND			
H2172	NSCLC	ND*			
H2228	NSCLC	ND			
H23	NSCLC	KRAS	Activating mutation	G12C	34G>T
H23	NSCLC	PTEN	Inactivating mutation	R233*	697C>T
H2347	NSCLC	NRAS	Activating mutation	Q61R	182A>G
H2444	NSCLC	KRAS*	Activating mutation	G12V	

H28	NSCLC	ND			
H2882	NSCLC	ND			
H2887	NSCLC	KRAS*	Activating mutation	G12V	
H3122	NSCLC	ND			
H322	NSCLC	ND			
H3255	NSCLC	EGFR	EGFR-I sensitizing mutation	L858R	34G>T
H358	NSCLC	KRAS	Activating mutation	G12C	34G>T
H441	NSCLC	KRAS	Activating mutation	G12V	35G>T
H460	NSCLC	PIK3CA	Activating mutation	E545K	1633G>A
H460	NSCLC	KRAS	Activating mutation	Q61H	183A>T
H520	NSCLC	ND			
H522	NSCLC	ND			
H596	NSCLC	PIK3CA	Activating mutation	E545K	1633G>A
H647	NSCLC	KRAS	Activating mutation	G13D	38G>A
H661	NSCLC	ND			
H820	NSCLC	EGFR*	EGFR-I sensitizing mutation	E746_T751del, I ins	
H820	NSCLC	EGFR*	EGFR-I resistnat mutation	T790M	2369C>T
HCC1171	NSCLC	KRAS*	Activating mutation	G12C	
HCC1195	NSCLC	NRAS*	Activating mutation	Q61L	
HCC1359	NSCLC	ND*			
HCC15	NSCLC	NRAS*	Activating mutation	Q61K	
HCC1833	NSCLC	ND*			
HCC193	NSCLC	ND*			
HCC2279	NSCLC	EGFR*	EGFR-I sensitizing mutation	E746_A750del	2235_2249del15
HCC2429	NSCLC	ND*			
HCC2450	NSCLC	PIK3CK*	Activating mutation	H1047R	3140A>G
HCC2935	NSCLC	EGFR*	EGFR-I sensitizing mutation	E746_A750del, T751A	

HCC364	NSCLC	BRAF	Activating mutation	V600E	1799T>A
HCC366	NSCLC	ND*			
HCC4006	NSCLC	EGFR*	EGFR-I sensitizing mutation	L747_E749del, A750P	
HCC44	NSCLC	KRAS*	Activating mutation	G12C	
HCC461	NSCLC	KRAS*	Activating mutation	G12D	
HCC515	NSCLC	KRAS*	Activating mutation	G13D	
HCC78	NSCLC	ND*			
HCC827	NSCLC	EGFR*	EGFR-I sensitizing mutation	E746_A750del	2235_2249del15
HCC95	NSCLC	ND*			
HOP62	NSCLC	KRAS	Activating mutation	G12C	34G>T
HOP92	NSCLC	ND			
LCLC103H	NSCLC	ND			
LCLC97TM	NSCLC	KRAS	Activating mutation	G12V	35G>T
LouNH91	NSCLC	EGFR*	EGFR-I sensitizing mutation	L747_S752del, P753S	
PC9	NSCLC	EGFR*	EGFR-I sensitizing mutation	E746_A750del	2235_2249del15
SKLU1	NSCLC	KRAS*	Activating mutation	G12D	35G>A
AU565	Breast cancer	ND			
BT20	Breast cancer	PIK3CA	Activating mutation	P539R	1616C>G
BT20	Breast cancer	PIK3CA	Activating mutation	H1047R	3140A>G
BT474	Breast cancer	ND			
BT549	Breast cancer	PTEN	Inactivating mutation	V275fs*1	823delG
CAL-51	Breast cancer	PIK3CA	Activating mutation	E542K	1624G>A
CAMA1	Breast cancer	PTEN	Inactivating mutation	D92H	274G>C
EFM19	Breast cancer	PIK3CA	Activating mutation	H1047L	3140A>T
EFM19	Breast cancer	PIK3CA*	Activating mutation	H1047R	3140A>G
EFM192A	Breast cancer	PIK3CA*	Activating mutation	C420R	
HCC1143	Breast cancer	ND			

HCC1395	Breast cancer	ND			
HCC1419	Breast cancer	ND			
HCC1954	Breast cancer	PIK3CA*	Activating mutation	H1047R	3140A>G
HCC70	Breast cancer	ND			
HS578T	Breast cancer	HRAS	Activating mutation	G12D	c35G>A
JIMT-1	Breast cancer	PIK3CA[#]	Activating mutation	C420R	
KPL1	Breast cancer	ND*			
MCF7	Breast cancer	PIK3CA	Activating mutation	E545K	c633G>A
MDA-MB-157	Breast cancer	ND			
MDA-MB-175VII	Breast cancer	ND			
MDA-MB-231	Breast cancer	BRAF	Activating mutation	G464V	1391G>T
MDA-MB-231	Breast cancer	KRAS	Activating mutation	G13D	38G>A
MDA-MB-361	Breast cancer	PIK3CA	Activating mutation	E545K	1633G>A
MDA-MB-361	Breast cancer	PIK3CA	Activating mutation	K567R	1700A>G
MDA-MB-415	Breast cancer	PTEN	Inactivating mutation	C136Y	407G>A
MDA-MB-435s	Breast cancer	BRAF	Activating mutation	V600E	1799T>A
MDA-MB-436	Breast cancer	ND			
MDA-MB-453	Breast cancer	PTEN	Inactivating mutation	E307K	919G>A
MDA-MB-453	Breast cancer	PIK3CA	Activating mutation	H1047R	3140A>G
MDA-MB-468	Breast cancer	PTEN	Inactivating mutation	A72fsX5	253+1G>T
SK-BR-3	Breast cancer	ND			
T47D	Breast cancer	PIK3CA	Activating mutation	H1047R	3140A>G
UACC812	Breast cancer	ND*			
UACC893	Breast cancer	PIK3CA	Activating mutation	H1047R	3140A>G
ZR-75-1	Breast cancer	ND			

ZR-75-30	Breast cancer	ND			

* Mutation was only reported in Ref [592]; # PIK3CA mutation of JIMT-1 was reported by Ref [892]

Abbreviations: ND, no sensitizing/resistant/activating mutation was detected according to COSMIC database and Ref 4.

Appendix Table S10 Sensitivity data of NSCLC cell-lines treated with gefitinib, erlotinib, and lapatinib

NSCLC Cell-line	Sensitivity of Cell-line to Gefitinib Inhibition #	Reported Potency (IC50) of Gefitinib Inhibition (μM)		Sensitivity of Cell-line to Erlotinib Inhibition #	Reported Potency (IC50/ED50) of Erlotinib Inhibition (μM)		Sensitivity of Cell-line to Lapatinib Inhibition #	Reported Potency (ED50) of Lapatinib Inhibition (μM)
		Ref [594]	Ref [593]		Ref [594]	Ref [590]		Ref [590]
A427				R		1.24	R	9.4406
A549	R	25		R	60	10	R	10
Calu1	R		41	R		10	R	10
Calu3	S	0.78		-	1.29	0.7	S	0.1679
Calu6	R		34	R		9.65	R	2.7542
Colo699				R		4.26	R	5.8884
DV90				R		3.95	R	1.4125
EKVX				R		10	R	10
H1155	R	183		R	8.63			
H1299	R	26.4		R	41.9	10	R	10
H1355	R	325		R	27	3.31	R	5.6885
H1395	R	71		R	10.5	5.05	R	6.6834
H1437	R	62		R	12.5	10	R	10
H1563				R		10	R	10
H1568				R		1.08	R	2.541
H157	R	115		R	128	10	R	10
H1648	R	36.7		R	34	7.77	S	0.9441
H1650	R	11.7		R	15	2.13	R	3.8905
H1666	R	180		R	13	3.31	S	0.5957
H1734				R		3.79	R	4.3652
H1755				R		7.5	R	10

H1770	R	160		R	111	10		
H1781	R	19		R	44	2.54	R	2.9174
H1792				R		10	R	10
H1819	R	19		R	6.3	3.92	S	0.7328
H1838				R		3.47	R	10
H1915				R		10	R	10
H1944				R		1.83	R	10
H1975	R	25		R	33	10	R	10
H1993	R	17.9		R	5.2	8.06	R	4.3152
H2009	R	33.2		R	25.8	10	R	10
H2030				R		4.95	R	5.0119
H2052				R		8.98	R	10
H2077				R		10	R	10
H2087	R	18.4		R	9.9	10	R	10
H2110				R		4.5	R	2.7861
H2122	R	35		R	76.8	10	R	10
H2126	R	21.4		R	13	10	R	10
H2172				R		10	R	8.9125
H2228				R		10	R	10
H23				R		10	R	5.6234
H2347	R	60		R	5.2	10	R	5.9566
H2444				R		4.22	R	7.6736
H28				R		10	R	1.6032
H2882	R	19.2		R	66	10	R	5.1286
H2887	R	110		R	101	10	R	10
H3122				R		10	R	10
H322	R	120		R	56	2.21	R	2.4831

H322M				R		1.29	S	0.4416
H3255	S	0.089		S	0.129	0.02	S	0.309
H358	R	12.5		R	6.2	1.11	R	1.6032
H441	R	15.7		R	7.1	3.61	R	10
H460	R	16.9		R	72	10	R	3.3113
H520	R	13.6		R		10	R	6.8391
H522				R		5.83	R	8.7096
H596				R		1.2	R	10
H647				R		10	R	10
H661				R		10	R	10
H820	R	3		R	7.1	10	R	10
HCC1171	R	127		R	160	10	R	10
HCC1195	R	27.6		R	175	10	NA	
HCC1359	R	65		R	88	10	R	10
HCC15	R	52		R	100	10	R	10
HCC1833				R		10	R	2.6915
HCC193	R	21.1		R	20.5	10	R	1.7378
HCC2279	S	0.0479		S	0.093	0.01	R	10
HCC2429				R		10	R	5.9566
HCC2450				R		10	R	10
HCC2935	S	0.11		S	0.163	0.07	S	0.2344
HCC364				R		4.19	R	10
HCC366	R	30		R	11	0.99	R	10
HCC4006	S	0.23		S	0.124	0.04	S	0.537
HCC44	R	57.8		R	28	10	R	10
HCC461	R	13.9		R	16	9.04	R	10
HCC515	R	120		R	154	1.85	R	9.5499

HCC78	R	81		R	21.2	10	R	4.1687
HCC827	S	0.04		S	0.0388	0.02	S	0.7943
HCC95	R	24		R	18.4	10	R	3.2359
HOP62				R		10	R	5.4325
HOP92				R		10	R	10
LCLC103H				R		10	R	10
LCLC97TM				R		5.26	R	7.3282
LouNH91				R		3.05	R	5.1286
PC9	S	0.0309		S		0.02	R	1.4962
SKLU1				R		10	R	10

* A cell-line with IC50≤ 1 μmol/L for gefitinib, erlotinib, and lapatinib was considered to be sensitive (S) to a given drug[583], otherwise it was considered as resistant (R) to the drug. - : cell-line with inconsistent sensitivity data, which is not included in this study.

Appendix Table S11 Sensitivity data of breast cancer cell-lines treated with trastuzumab and lapatinib

Breast Cancer Cell-lines	Sensitivity of Cell-line to Trastuzumab inhibition[#]	Reported Potency of Trastuzumab Inhibition (%25 growth inhibition of cancer cell at 10µg/ml or ED50 ≤1µg/ml)				Sensitivity of Cell-line to Lapatinib Inhibition[#]	Reported Potency(IC50) of Lapatinib Inhibition (µM)		
		Ref [599]	Ref [600]	Ref [601]	Ref [598]		Ref [602]	Ref [604]	Ref [603]
BT20	R			R		R	9.8		
BT474	S	S	S	S	S	S	0.022	0.025	
BT549	R		R			R		6.35	
CAL51						R	1.2		
CAMA1	R		R			R	8.3		
EFM19						R	4.6		
EFM192A	R				R	R	1.1		
HCC1143	R			R					
HCC1395	R			R					
HCC1419	S	S	S						
HCC1954	R	R	R						
HCC70	R		R						
Hs578T						R		5.11	
JIMT-1	R	R				R			>1.5
KPL1						R	5.4		
MCF7	R	R	R	R		R	7.7	4.82	
MDA-MB-157						R	6.3		
MDA-MB-175VII						S	0.012		
MDA-MB-231	R		R	R		R	18.6	7.01	
MDA-MB-361	R		R		R	R	0.99		
MDA-MB-415	R		R						
MDA-MB-435s	R		R	R		R	8.5		
MDA-MB-436	R			R					
MDA-MB-453	S		S			-	3.9	0.079	
MDA-MB-468	R			R		R	4.7	2.32	
SK-BR-3	S	S	S			S	0.037	0.032	
T47D	R		R	R		R	1.9	4.83	
UACC812	S		S			S	0.01		
UACC893	R		R			S	0.433		

ZR75-1	-		R	S		R	9.9	3.01	
ZR75-30	S	S	S						

A cell-line with IC50≤ 1 µmol/L for lapatinib or %25 growth inhibition of cancer cells at 10µg/ml for Trastuzumab was considered to be sensitive (S) to a given drug[583], otherwise it was considered as resistant (R) to the drug. - : cell-line with inconsistent sensitivity data, which is not included in this study.

Appendix Table S13 The clinical information of 86 lung adenocarcinoma samples from Beer et al [662]

Sample ID	cluster ID[1]	Age	Sex	Tumor stage. either 1 or 3	T (tumor size)	N (nodal status)	Survival times (month)[2]	Patient's survival status	classification (tumor histological type)[3]	Tumor differentiation	p53 nuclear accumulation status	12/13th codon K-ras mutation status	Smoking [4]
AD2	Cluster 1	65.6	F	1	1	0	91.8	alive	BD	Poor	+	-	48
AD5	Cluster 1	62	F	1	2	0	108.2	alive	BA	Well	-	+	positive
L01	Cluster 1	76.7	M	1	2	0	47	alive	BD/CC	Poor	-	-	100
L06	Cluster 1	57.9	F	1	1	0	91.9	alive	BD	Poor	-	+	NA
L26	Cluster 1	61.4	M	1	2	0	17.7	alive	BD	Poor	-	+	90
L33	Cluster 1	53.5	F	3	4	0	29.4	alive	BD	Moderate	-	-	23
L43	Cluster 1	50.6	F	1	2	0	78.5	alive	BD	Moderate	-	-	57
L56	Cluster 1	60.2	M	1	1	0	61.8	alive	BD/CC	Moderate	+	-	90
L62	Cluster 1	52.3	F	3	3	2	52.4	alive	BD	Moderate	-	-	none
L83	Cluster 1	62	F	1	2	0	30.6	alive	BA/mucinous	Well	-	-	none
L91	Cluster 1	63.7	M	3	2	2	6.1	alive	BD/mucinous	Poor	-	-	30
L92	Cluster 1	55.4	M	3	4	0	8.5	alive	BD	Poor	-	-	50
AD10	Cluster 1	65	M	1	1	0	84.1	death	BD	Moderate	-	NA	60
L04	Cluster 1	51.7	M	1	2	0	45.8	death	BD	Poor	-	-	50
L13	Cluster 1	67.1	M	1	1	0	79.5	death	BD	Moderate	+	+	25
L19	Cluster 1	56.5	M	3	3	2	9.6	death	BD	Moderate	-	+	40
L34	Cluster 1	77.2	M	3	1	2	14.9	death	BD	Moderate	+	-	45
L36	Cluster 1	69.7	M	3	1	2	7.2	death	BD/PA	Moderate	-	+	25
L37	Cluster 1	64.4	M	3	1	2	2.6	death	BD	Poor	-	+	84
L41	Cluster 1	73.1	F	1	2	0	8.4	death	BD/CC	Poor	-	+	26
L54	Cluster 1	45.8	F	3	3	1	4	death	BD	Poor	+	+	75
L40	Cluster 1	54.9	F	3	1	2	20.1	death	BD	Moderate	-	-	7.5
L80	Cluster 1	68.2	F	1	2	0	10.1	death	BD/mucinous	Moderate	+	+	50
L61	Cluster 1	63.1	F	1	2	0	20.6	death	BD	Moderate	-	-	30
L95	Cluster 1	72	F	3	2	2	5.4	death	BD	Poor	-	+	50
L96	Cluster 1	64	F	3	3	1	21.2	death	BD	Moderate	-	+	50
AD7	Cluster 2	56	M	1	1	0	68.1	alive	BD	Moderate	+	-	80
L02	Cluster 2	63.2	M	1	1	0	39.1	alive	BD	Poor	-	-	27

L09	Cluster 2	48.2	F	1	1	0	98.7	alive	BD	Moderate	-	+	none
L101	Cluster 2	46.3	F	1	1	0	40	alive	B/A/mucinous	Well	-	-	NA
L103	Cluster 2	84.6	F	1	1	0	30.8	alive	B/A	Well	-	-	none
L104	Cluster 2	68.5	F	1	1	0	24.4	alive	B/A	Well	-	-	5
L105	Cluster 2	74.2	F	1	1	0	28.3	alive	B/A with PA	Well	-	+	75
L108	Cluster 2	61	F	1	1	0	19.5	alive	B/A	Well	-	+	100
L111	Cluster 2	54.9	F	1	1	0	1.5	alive	B/A	Well	-	+	40
L12	Cluster 2	44.6	F	1	1	0	85.2	alive	BD	Moderate	-	-	15
L18	Cluster 2	82.5	F	1	1	0	48.2	alive	BD	Well	-	-	none
L23	Cluster 2	62.2	M	3	2	2	15.1	alive	BD/PA	Moderate	-	+	20
L25	Cluster 2	62.6	F	1	2	0	14.5	alive	BD	Well	-	+	50
L27	Cluster 2	70	M	1	1	0	21.1	alive	BD	Poor	+	-	60
L38	Cluster 2	78.5	F	3	4	2	10	alive	BD	Poor	+	+	2
L42	Cluster 2	76	F	1	1	0	63.4	alive	BD	Well	-	-	40
L46	Cluster 2	60.4	M	1	2	0	82.4	alive	BD	Poor	-	+	160
L47	Cluster 2	60	M	1	2	0	60.5	alive	BD	Moderate	-	+	27
L48	Cluster 2	42.8	M	1	1	0	77.8	alive	BD	Moderate	-	-	60
L52	Cluster 2	67.3	M	1	1	0	65.4	alive	BA	Well	-	-	30
L57	Cluster 2	73.6	F	1	2	0	54.8	alive	BD/PA	Moderate	-	+	50
L65	Cluster 2	59.6	M	1	1	0	52.9	alive	BD	Moderate	-	-	60
L78	Cluster 2	75.6	F	1	1	0	36.5	alive	BD	Moderate	-	+	108
L82	Cluster 2	69.2	F	1	1	0	34.1	alive	BA/BD	Well	-	-	40
L85	Cluster 2	60.2	M	1	1	0	26.8	alive	BD/mucinous	Moderate	-	+	60
L97	Cluster 2	63.6	F	1	1	0	4.9	alive	B/A	Well	-	+	34
L50	Cluster 2	72.1	M	1	1	0	19	death	BD/PA	Moderate	+	+	100
AD3	Cluster 3	59.5	F	1	2	0	93.7	alive	BD	Moderate	-	-	positive
AD8	Cluster 3	75	M	1	2	0	34.2	alive	BD	Moderate	-	-	14
L05	Cluster 3	54.6	F	1	1	0	110.6	alive	BD/CC	Moderate	-	-	29
L08	Cluster 3	59.9	F	1	1	0	107.9	alive	BD	Moderate	-	+	80
L102	Cluster 3	74.6	F	1	1	0	40	alive	BD	Moderate	-	-	50
L106	Cluster 3	82.8	F	1	1	0	25.3	alive	B/A	Well	-	-	none
L107	Cluster 3	59.4	F	1	1	0	13	alive	BD	well/mod.	-	+	none

L17	Cluster 3	40.9	F	1	2	0	83.7	alive	BD/PA	Moderate	-	+	15
L22	Cluster 3	65.6	M	1	1	0	12.5	alive	BD	Moderate	+	-	90
L30	Cluster 3	51.8	F	1	1	0	20.2	alive	BD	Moderate	+	-	20
L31	Cluster 3	62.1	F	1	1	0	25.2	alive	BA/mucinous	Well	-	-	20
L49	Cluster 3	65.8	F	1	1	0	70.7	alive	BD	Moderate	-	+	20
L59	Cluster 3	71.5	F	3	2	2	54.6	alive	BD/PA	Moderate	-	+	25
L64	Cluster 3	65.4	M	1	2	0	48.1	alive	BD	Moderate	+	-	12
L76	Cluster 3	46.2	M	1	1	0	87.7	alive	BD	Poor	-	+	50
L81	Cluster 3	58.4	M	1	1	0	36	alive	BA	Well	-	-	90
L84	Cluster 3	66.8	F	1	2	0	32.2	alive	BD	Poor	-	-	15
L86	Cluster 3	62.7	F	1	1	0	10.1	alive	B/A	Well	-	-	45
L87	Cluster 3	66.3	M	1	1	0	10.4	alive	BD	Moderate	+	-	18
L88	Cluster 3	52.9	F	1	1	0	8.3	alive	BD	Poor	+	+	60
L89	Cluster 3	58.8	M	3	2	2	12.2	alive	BD	Moderate	NA	+	48
L99	Cluster 3	73.8	M	1	2	0	4.5	alive	B/A/mucinous	Well	-	+	55
L100	Cluster 3	72.9	F	1	1	0	43.8	censored	B/A	Well	-	-	2.5
L24	Cluster 3	84.5	F	1	1	0	1.6	censored	BD	Poor	-	-	75
AD6	Cluster 3	66.2	M	1	2	0	34.6	death	BA	Well	-	+	NA
L11	Cluster 3	68.2	F	1	2	0	34.7	death	BA	Well	-	+	none
L20	Cluster 3	79.8	M	1	2	0	19.9	death	BA	Well	-	-	30
L35	Cluster 3	64.4	M	3	2	2	28.2	death	BD	Moderate	+	+	4
L45	Cluster 3	74.9	F	1	1	0	29.6	death	BD	Poor	-	+	30
L53	Cluster 3	58.5	F	3	2	2	16.6	death	BD/PA	Moderate	-	-	none
L79	Cluster 3	49	F	1	2	0	8.7	death	BD	Poor	-	-	60
L90	Cluster 3	63.8	F	1	1	0	5.8	death	BD/PA	Moderate	-	-	100
L94	Cluster 3	72	M	3	3	2	2.4	death	BD/mucinous	Moderate	-	-	50

[1]These clusters are obtained from hierarchical cluster analysis of the 86 samples and 21 survival marker genes share by 10 signatures
[2]This is patient's survival time from operation date to death or last follow up as of May 2001
[3]BD: bronchial derived; BA: bronchial alveolar; CC: clear cell; PA: papillary; Note that some tumors contained a mixture of two histological types
[4]Patient smoking history in packs per year

Appendix Table S14 The clinical information of 84 lung adenocarcinoma samples from Bhattacharjee et al [663]

Sample ID	Cluster ID[1]	Age	Sex	Stage:AJCC TNM	Stage Summary	Survival time (month)[2]	Patient's status*	Clinical Path (type diameter features) [3]	Path II[4]	Site of elapse/ metastasis	Smoking[5]
AD111	Cluster 3	76	F	T1NxMx	IA	72.4	1	ad 2.0 m-p			40
AD115	Cluster 2	70	F	T2N1M0	IIB	21.9	3	ad 6.5 m	adm/adw	lung, LN	75
AD118	Cluster 3	69	M	T1N0Mx	IA	49.6	3	ad 2.5 m	adm	lung, LN	25
AD120	Cluster 3	68	M	T2N0Mx	IB	38.9	3	ad 8.0 m	adm	bone	54
AD122	Cluster 1	73	F	T2N1Mx	IIB	33.9	3	ad 5.0 m	adm	lung	0
AD123	Cluster 1	60	M	T3N0Mx	IIB	74	1	ad 5.0 m	adm,pap		126
AD127	Cluster 3	65	F	T1N2Mx	IIIA	8.2	3	ad 1.8 p	adp	LN	69
AD130	Cluster 1	75	M	T2N1Mx	IIB	7.1	d	ad 15.0 BAC	BAC		100
AD136	Cluster 1	66	F	T2N0Mx	IB	31.4	1	ad 4.0 m	adm		100
AD159	Cluster 1	71	M	T2N1Mx	IIB	19.7	d	ad 5.5 m-p	adw,acinar		80
AD162	Cluster 3	75	F	T2N0Mx	IB	41.7	1	ad 3.5 m	admod,acinar		60
AD164	Cluster 1	68	M	T3N0Mx	IIB	15	3	ad 4.5 p	adpoor, acinar	LN	80
AD167	Cluster 2	77	M	T2N0Mx	IB	41.7	1	ad 2.5 w w/BAC	adw,acinar/adm bac		0
AD169	Cluster 1	47	F	T2N0Mx	IB	20	3	ad 2.5 m	adw/pap or BAC,mucinous w/pap	bone, myocardium	21.6
AD170	Cluster 3	61	F	T1N0M0	IA	78.4	1	ad 2.5 w w/pap	BAC & pap,well		60
AD173	Cluster 1	57	F	T2N1Mx	IIB	22.3	d	ad 5.0 m-p	admod,acinar		27
AD179	Cluster 2	85	M	T2N0Mx	IB	24.3	3	ad 5.6 m w/BAC	adw//adw,acinar	lung, bone	24.75
AD187	Cluster 1	69	M	T1N0Mx	IA	86.3	3	ad 1.8 p	adp	lung	120
AD183	Cluster 2	75	F	T1N0Mx	IA	42.2	2	ad 2.0 m BAC	adw//adw,acinar		22.5
AD188	Cluster 1	74	F	T2NxMx	IB	21.6	d	ad 2.7 BAC	adw,acinar		116
AD201	Cluster 1	46	M	T1N2	IIIA	12.3	3	ad 1.5 m		lung, bone	90
AD203	Cluster 3	60	F	T1N0Mx	IA	106.1	1	ad 2.2 m-p			0
AD207	Cluster 3	64	F	T2	IB	66.8	4	ad 3.5 w BAC	ad m		0
AD212	Cluster 3	55	F	T2N0M0	IB	59	1	ad 3.0 m-p			54
AD213	Cluster 2	69	M	T1Nx	IA	48.8	d	ad 2.5 m			111

AD225	Cluster 1	88	M	T2NxMx	IB	2.6	4	ad 3.5 m			72
AD226	Cluster 1	56	F	T1N0Mx	IA	60.5	1	ad 2.0 m			18
AD228	Cluster 1	60	F	T2N0	IB	41.2	3	ad 3.0 m		brain	75
AD230	Cluster 3	56	M	T1N0	IA	56.7	1	ad 2.5 p	adp		60
AD232	Cluster 3	73	M	T1Nx	IA	56.3	a	ad 2.4 w BAC	adm (BAC cluster)		25
AD236	Cluster 1	53	F	T2N0Mx	IB	14.2	3	ad 5.5 m-p		lung, brain	40
AD239	Cluster 3	60	M	T2N0M0	IB	58.5	1	ad 2.9 m w/BAC	BAC		40
AD240	Cluster 1	77	F	T1N0M0	IA	43.5	1	ad 2.0 m-w			5
AD243	Cluster 2	64	F	T1N0M0	IA	50.1	1	ad 1.5 w w/BAC	adw resemblance to BAC		30
AD247	Cluster 3	49	M	T1N0	IA	71.1	1	ad 2.0 m			32
AD249	Cluster 1	67	M	T1Nx	IA	31	4	ad 1.2 m			45
AD250	Cluster 3	61	F	T1Nx	IA	91	2	ad 2.0 w w/BAC	adm	lung	10
AD252	Cluster 2	66	F	T1N0	IA	16.5	3	ad 1.4		LN, CSF, brain	50
AD255	Cluster 3	79	M	T2N0	IB	44.8	1	ad 3.5 m			50
AD258	Cluster 1	67	M	T2Nx	IB	12.3	3	ad 4.5 p		bone	54
AD259	Cluster 2	58	M	T3N0	IIB	20.5	d	ad 5.0			45
AD260	Cluster 1	61	M	T2Nx	IB	21	d	ad 3.0 m	adm some BACpattern		50
AD261	Cluster 2	66	F	T1N0	IA	57.6	1	ad 2.7 w w/BAC			75
AD262	Cluster 2	63	F	T4N1Mx	IIIB	16.6	4	ad 2.0 m-p			10
AD266	Cluster 2	65	F	T1N0	IA	41.9	3	ad 2.5 w w/BAC	adm	lung, bone, liver	0
AD267	Cluster 3	61	M	T2N0M0	IB	56	1	ad 2.8 m-p			120
AD268	Cluster 3	50	F	T2N0M0	IB	50.1	1	ad 3.5 p			10
AD276	Cluster 2	68	M	T2N2	IIIA	4.5	3	ad 2.1 m-p		pleura, brain	140
AD277	Cluster 3	72	F	T1Nx	IA	8.2	3	ad 3.0 m		liver, ?bone	27
AD283	Cluster 3	78	M	T1N0	IA	47.2	3	ad 2.5 m w/pap		lung, LN, bone, groin	20
AD287	Cluster 3	36	F	T4Nx	IIIB	7.4	d	ad 4.0 p	adp		10
AD296	Cluster 1	63	M	T1N1	IIA	9.3	3	ad 2.4 m-p w/pap		liver	88
AD299	Cluster 1	78	F	T1N0M0	IA	37.9	3	ad 2.2 m-p		lung	50
AD301	Cluster 1	59	F	T2N0M0	IB	7.8	3	ad 4.0 p		brain	40

AD302	Cluster 3	65	F	T2N3Mx	IIIB	57.8	3	ad 3.7 w BAC	adm w/BAC	lung	0
AD304	Cluster 2	71	F	T2N0	IB	8.2	3	ad 5.0 p		lung, liver, spleen	35
AD308	Cluster 3	62	M	T2N0	IB	79	2	ad 4.0 m		brain	66
AD309	Cluster 1	77	F	T2N0	IB	37.6	3	ad 3.4 w	adw	lung	0
AD311	Cluster 1	63	F	T2N0	IB	50.5	1	ad 5.0 m	ok 50%		13
AD313	Cluster 1	74	F	T1N0	IA	25.3	3	ad 1.5 m-p	adp	LN	90
AD317	Cluster 3	41	F	T2Nx	IB	99.1	1	ad 3.5 m pap			7
AD318	Cluster 1	54	M	T2N0M0	IB	83	1	ad 4.0 muc	adm		100
AD323	Cluster 1	56	F	T2N1	IIB	6.8	d	ad 4.0 p			39
AD327	Cluster 1	50	F	T2N0	IB	81.9	1	ad 6.5 m			27
AD330	Cluster 3	50	F	T1N1	IIA	7.3	3	ad 2.4 m		brain	40
AD331	Cluster 3	59	M	T1N0M0	IA	52.9	1	ad 2.0 m			45
AD332	Cluster 1	52	M	TxN0	I	6	3	ad m		pleura, liver, colon, ?adrenal, ?pancreas	75
AD335	Cluster 2	40	F	T3N0	IIB	46.9	1	ad 4.5 m			20
AD336	Cluster 3	71	M	T2N0Mx	IB	21.1	4	ad 1.7 m			0
AD338	Cluster 3	55	F	T2NxMx	IB	75.4	1	ad 5.0 w BAC	(1) ad w/BAC or (2)BAC		15
AD346	Cluster 2	65	F	T1N0	IA	17.3	1	ad 2.5 m			50
AD347	Cluster 1	65	F	T2N0Mx	IB	0.5	1	ad 3.5 m BAC	adm		20
AD351	Cluster 2	43	F	T2N1	IIA	24.3	3	ad 5.5 m		lung, LN	0
AD353	Cluster 3	69	M	T2N0Mx	IB	13.7	1	ad 3.5 m BAC	adw w/bac		30
AD356	Cluster 2	72	M	T2N0	IB	49.2	1	ad 4.0 w BAC			50
AD361	Cluster 1	54	F	T2N	IB	6.4	4	ad 4.5 p			0
AD362	Cluster 3	56	M	T2N0	IB	71.5	d	ad 6.5 BAC	BAC muc		40
AD366	Cluster 2	71	M	T2N2	IIIA	9.4	3	ad 6.2 m-p w/pap		lung	23
AD367	Cluster 1	55	F	T2N0	IB	76.1	2	ad 6.5 m-p		brain	25
AD368	Cluster 3	33	F	T2N0	IB	62.6	1	ad 6.0 m-p w/muc			32
AD374	Cluster 1	51	M	T2N0	IB	8.8	3	ad 11.0 p		lung, pleura, pericardium, diaphragm	100
AD375	Cluster 2	47	F	T2N0	IB	23.4	d	ad 7.2 p	adm		13
AD379	Cluster 2	65	M	T2N1	IIB	35.4	2	ad 5.5 w/clear		lung, adrenal, brain	80

AD382	Cluster 2	51	F	T2N2Mx	IIIA	30.1	3	ad 5.0 p		brain	31

[1]These clusters are obtained from hierarchical cluster analysis of the 84 samples and 21 survival marker genes we selected.
[2]Patient status at last followup or death (1= alive; 2=alive with recurrence; 3= dead with recurrence; 4= dead without evidence of recurrence; d= dead, disease status unknown)
[3,4]diameter (cm) subtype (BAC = bronchioloalveolar carcinoma). type (ad = adenocarcinoma) differentiation (p, m-p, m, m-w, w) /w= with
[5]Smoking: patient smoking history (self-reported) in pack/year

Appendix Table S15 List of 10 derived lung adenocarcinoma prognosis marker gene signatures selected by SVM class-differentiation systems

Gene Name (EST number)	Number of signatures which included this gene	Gene rank in each signature (Number of selected gene in each signaure)									
		1 (51)	2 (54)	3 (42)	4 (34)	5 (46)	6 (54)	7 (57)	8 (50)	9 (53)	10 (47)
ADFP(X97324)	10	1	46	35	28	19	22	15	18	3	13
CXCL3(X53800)	10	2	37	24	7	23	3	14	4	6	19
PLD1(U38545)	10	5	7	2	31	41	17	8	9	11	3
SLC2A1(K03195)	10	6	3	12	3	3	8	13	12	2	11
SPRR1B (M19888)	10	7	10	29	11	10	7	9	10	5	12
GALNT4 (Y08564)	10	8	23	25	27	11	32	25	14	1	28
LDHB(X13794)	10	10	11	1	1	15	16	11	8	15	1
FXYD3(U28249)	10	11	6	7	29	14	52	18	42	22	5
REG1A(J05412)	10	13	8	9	23	9	15	16	45	14	6
CHRNA2 (U62431)	10	14	24	26	30	8	46	28	40	27	27
SERPINE1 (J03764)	10	18	30	16	22	12	2	1	31	4	15
FUT3(U27326)	10	19	14	19	21	2	28	10	15	30	21
PRKACB (M34181)	10	20	5	5	15	6	1	3	1	33	4
TUBA4A(X06956)	10	21	1	14	25	13	53	49	29	26	14
VEGF(M27281)	10	22	33	8	26	30	14	26	19	23	32
RPS3(X55715)	10	25	2	10	2	5	39	55	13	17	36
ANXA8(X16662)	10	28	32	18	12	21	20	4	22	18	26
VDR(J03258)	10	32	39	33	6	4	30	2	11	16	37
CXCR7(U67784)	10	33	47	30	24	43	41	37	27	39	29
POLD3(D26018)	10	35	25	15	18	1	11	50	2	31	8
BSG(X64364)	10	36	38	39	17	33	48	27	3	20	33
CYP24(L13286)	9	23	13	34	20	22	23		41	19	25
HLA-G (HG273-HT273)	9	30	27	11		25	19	34	32	24	31
WNT10B (U81787)	9	39	35	28		39	36	29	25	38	41
GARS(U09510)	9	41	26	31		31	26	19	46	44	20
SPRR2A(M21302)	9		21	13	34	40	21	21	34	47	18
NULL (HG2175-HT2245)	9		49	37	5	44	34	56	35	53	16
CD58(Y00636)	8	16	12	3	14	17	6		44		34
KRT14(J00124)	8			20	9	34	25	12	23	12	22
E48(X82693)	7	9	15			20	33	22	5	48	
FADD(X84709)	7	12		6		35	51	17		8	9
STX1A(L37792)	7	15	18	22		46	5		6		24
ENO2(X51956)	7	24	4			32	38	32		45	47
SPRR2A(L05188)	7	29	41			7	45	44	48	28	
FEZ2(U69140)	7	38		23			42	30	26	9	17
KRT18(X12876)	7	43	42	41		26	44	6	43		
ALDH2(X05409)	7		19		10	45	4	20	21		23
UCN(U43177)	6	4		36	13	18	9				10
SCYB5(L37036)	6	31	16		33	42	31			29	
AIP-1(U23435)	6	37		42		28	18			32	7

NULL(U92014)	6	42	17			27		39		36	42
NULL(L43579)	6	47	54				35	5	24	37	
CEBPA(U34070)	6			17		24	12		47	25	30
KIAA0138 (D50928)	5	34	29		19		37				2
TFF1(X52003)	5	40	34			37	24			43	
KRT19(Y00503)	5	49			4		40	54	20		
RPS26(X69654)	4	17	28				49			21	
S100A2(Y07755)	4	26	51					40		34	
GS3686 (AB000115)	4	46	36						49	41	
EMP1(Y07909)	4		9			38	27		38		
HPCAL1(D16227)	4		43		8			33	36		
LCN2(S75256)	4			38				41	37		44
PEX7(U88871)	4			4		29				40	43
EFNB2(U81262)	3	44							30		40
ALDH8(U37519)	3	45	52						17		
EPS8(U12535)	3		20				50			51	
NDRG1(D87953)	3		22					48			46
CSTB(U46692)	3		40					45		10	
PSPH(Y10275)	3		44	27				23			
CYBA(M21186)	3						29	7	7		
CNN3(S80562)	3							57	39	49	
VIPR1(X77777)	3			40					50		35
NULL(U49020)	2	51							16		
ALDH7(U10868)	2		45				10				
AXL (HG162-HT3165)	2		53							35	
TYRO3(U02566)	2			32		36					
P2RX5(U49395)	2				32	16					
GRO1(X54489)	2								28	42	
ERBB3(M34309)	2							51		7	
BM-002(Z70222)	2				16		13				
LAMB3(U17760)	2			21							39
INHA(X04445)	2							38		46	
TAX1BP2 (U25801)	1	3									
IGHM(V00563)	1	27									
SPRR2A(X53065)	1	48									
NP(K02574)	1	50									
P63(X69910)	1		31								
AP3B1(U91931)	1		48								
C6(X72177)	1		50								
HFL1(M65292)	1										38
PRKCN (HG2707-HT2803)	1							24			
SHB(X75342)	1									13	
EIF5A(S72024)	1								33		
FCGR3B(J04162)	1							47			
GRIN1 (HG4188-HT4458)	1						47				

SLC2A3(M20681)	1										45
CA9(X66839)	1							42			
FLJ20746 (U61836)	1						43				
PPBP(M54995)	1							52			
TUBA4A (HG2259-HT2348)	1						54				
EMS1(M98343)	1							53			
IGF2(M17863)	1							36			
CHAT (HG4051-HT4321)	1							31			
LAMC2(U31201)	1									50	
BMP2(M22489)	1							43			
KIAA0111 (D21853)	1									52	
TNFAIP6 (M31165)	1							35			
NULL (HG415-HT415)	1							46			

LIST OF PUBLICATIONS

1. Synergistic therapeutic actions of herbal ingredients and their mechanisms from molecular interaction and network perspectives X. H. Ma, C.J. Zheng, L.Y. Han, B. Xie, J. Jia, Z.W. Cao, Y.X. Li and Y. Z. Chen. *Drug Discov Today.* 14:579-588(2009).

2. Genetic and Expression Profiles of Target and Bypass Genes Correlate with Sensitivity to Anticancer Tyrosine Kinase Inhibitor Drugs. J. Jia, S.Y. Yang, Y.Q. Wei,and Y. Z. Chen (submitted, 2010)

3. Genome-Scale Search of Tumor-Specific Antigens by Collective Analysis of Mutations, Expressions and T-Cell Recognition. J. Jia, Cui. J. , X. H. Liu, J. H. Han, S. Y. Yang, Y. Q. Wei, and Y. Z. Chen. *Mol. Immunol.* 46:1824-1829(2009)

4. Update of KDBI: Kinetic Data of Bio-molecular Interaction Database. P. Kumar, Z.L. Ji, B.C. Han, Z. Shi, J. Jia, Y.P, Wang, Y.T. Zhang, L. Liang, and Y. Z. Chen. *Nucleic Acids Res.* 37(Database issue): D636-41(2009).

5. Comparative analysis of machine learning methods in ligand-based virtual screening of large compound libraries. X.H. Ma, J. Jia, F. Zhu, Y. Xue, Z.R. Li and Y.Z. Chen. *Comb. Chem. High Throughput Screen.* 12(4):344-357(2009).

6. Mechanisms of drug combinations from interaction and network perspectives J. Jia, F. Zhu, X.H. Ma, Z.W. Cao, Y.X. Li and Y.Z. Chen. *Nat. Rev. Drug Discov.*, 8(2):111-28(2009).

7. Trends in the Exploration of Anticancer Targets and Strategies in Enhancing the Efficacy of Drug Targeting. F. Zhu, C.J. Zheng, L.Y. Han, B. Xie, J. Jia, X. Liu, M.T. Tammi, S.Y. Yang, Y.Q. Wei and Y.Z. Chen. *Curr Mol Pharmacol.* 1(3):213-232(2008).

8. Simulation of the Regulation of EGFR Endocytosis and EGFR-ERK Signaling by Endophilin-Mediated RhoA-EGFR Crosstalk. C.Y. Ung, H. Li, X.H. Ma, J. Jia, B.W. Li, B.C. Low and Y.Z. Chen.*FEBS Lett*. 582:2283-2290 (2008).

9. A support vector machines approach for virtual screening of active compounds of single and multiple mechanisms from large libraries at an improved hit-rate and enrichment factor. L.Y. Han, X.H. Ma, H.H. Lin, J. Jia, F. Zhu, Y. Xue, Z.R. Li, Z.W. Cao, Z.L. Ji, Y.Z. Chen. *J Mol Graph Mod.* 26(8):1276-1286 (2008)

10. Prediction of Antibiotic Resistance Proteins from Sequence Derived Properties Irrespective of Sequence Similarity. H.L. Zhang, H.H. Lin, L. Tao, X.H. Ma, J.L. Dai, J. Jia, Z.W. Cao.*Int J Antimicrob Agents*. 32(3):221-226 (2008)

11. Derivation of Stable Microarray Cancer-differentiating Signatures by a Feature-selection Method Incorporating Consensus Scoring of Multiple Random Sampling and Gene-Ranking Consistency Evaluation. Z.Q. Tang, L.Y. Han, H.H. Lin, J. Cui, J. Jia, B.C. Low, B.W. Li, Y.Z. Chen. *Cancer Res.* 67(20):9996-10003 (2007).

12. Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. L.Y. Han, C.J. Zheng, B. Xie, J. Jia, X.H. Ma, F. Zhu, H.H. Lin, X. Chen, and Y.Z. Chen. *Drug Discov. Today* 12(7-8): 304-313 (2007).