有志者，事竟成。

-出自<<后汉书>>

Where there is a will, there is a way.

- Thomas Edison

To my family

Cover picture

Image of yeast ORF microarray hybridization of yeast RNA. The microarray comprises of 62,000 yeast ORFs obtained by amplification of *S. cerevisiae* S288C genomic DNA using ResGen primers. Yeast total RNAs were prepared from yeast cells (strain FY1679) cultured both with and without Rapamycin. Fluorescent Dyes (Cy3 and Cy5) were coupled to these two different RNA probes via amino-allyl dUTP incorporation. This work was performed in EMBO practical course (DNA Microarrays: Applications and Data Analysis) in EMBL, Heidelberg, Germany in 2001.

ISBN 91-7349-432-1

ABSTRACT 4

LIST OF PUBLICATIONS 5

ABBREVIATIONS 6

INTRODUCTION 7

1. Cancer, genes and genome 7
1.1 Cancer and cancer caused genes 7
1.2 Human genome draft sequence and cancer research 12
1.3 Genes on human chromosome 3 with lung cancer and kidney cancer 13

2. CpG islands and methylation 16
2.1 CpG islands, methylation and genes 16
2.1.1 CpG islands and genes 16
2.1.2 Hypomethylation, hypermethylation and genes 19
2.2 Epigenetic mechanism of gene inactivation 24
2.3 *Not*I linking and jumping clones 26

3. Microarrays 30
3.1 cDNA/oligo microarrays with gene expression profiling 31
3.2 PAC/BAC microarrays with allele copy number change 34
3.3 *Not*I microarrays to detect copy number and methylation changes 35

4. Novel methodologies in cancer research 39
4.1 COP and CODE procedures for cloning different sequences 39
4.1.1 COP (cloning of polymorphisms) 39
4.1.2 CODE (cloning of deleted sequences) 41
4.2 CIS procedure for cloning identical sequences 43
4.3 *Not*I microarrays for isolation TSGs 45
4.4 GIT for functional identification of TSGs 46

SCOPE OF THIS THESIS 48

RESULTS AND DISCUSSION 49

1. Cloning identical and different sequences from complex genome (paper I-III) 49

***2. Not*I linking and jumping libraries and *Not*I microarrays (paper IV-VI)** 51

3. GIT and candidate TSGs in 3p21 region (paper VII-IX) 53

ACKNOWLEDGEMENTS 56

REFERENCES 57

PAPER I-IX

ABSTRACT

Development of tumor is a complex process involving multiple steps. New technologies for cloning and identifying genes playing critical role in cancer development are necessary. That is why we have focused our research on development of these approaches.

The new methods include CIS, cloning of identical sequences, COP, cloning of polymorphic sequences and CODE, cloning of deleted sequences. Although these methods are based on the same combination of biochemical techniques, their aims are different. These methods are fully complementary; therefore they may be applied together to analyze a given object. If one aims to clone a disease gene responsible for familial cancer syndrome, these methods may be applied as follows. CIS may be used to identify the sequences identical by descent comparing the DNA obtained from affected family members. COP may be used to find sequences that are different between affected members, and CODE would be useful to compare tumor and normal (control) samples to isolate deleted sequences (putative candidate tumor suppressor genes) and amplified sequences (putative oncogenes). COP and CODE procedures may be applied to analyze the CpG islands thus allowing direct candidate gene identification.

NotI microarrays are the microarrays giving the opportunity to detect copy number and methylation changes. *Not*I microarrays are based on large-scale sequencing of total human *NotI* linking clones, which were previously shown to be tightly associated with CpG islands and genes. We have solved the main problems for genome wide screening created by the size of human genome and numerous repeat sequences by developing a new method for labeling genomic DNA where only sequences surrounding *NotI* sites are labeled, *NotI* representation (NR). A pilot experiment using NR probes demonstrated the power of the method, and we successfully detected Chr 3 *NotI* clones deleted in ACC-LC5 and MCH939.2 cell lines. *NotI* arrays will speed up cancer research very significantly and can replace CGH, LOH and many cytogenetic studies, since the high-density grids with 50.000 *Not*I linking clones were constructed, 22 551 unique *Not*I flanking sequences were generated, covering a total of 16.2 Mb of the human genome.

The candidate tumor suppressor genes (TSGs) cloned by above new methods will be entered into gene inactivation test (GIT), which is a new functional test system for TSGs identification. GIT is based on our hypothesis that TSG must be inactivated in growing tumors in experimental conditions as it happened in nature; this inactivation of a TSG can be achieved by mutation, deletion, methylation etc. To verify our hypothesis, known suppressor genes *RB* and *p53* were built into pETI and pETE vectors that permitted tetracycline/doxycycline regulated expression of the cloned genes in cancer cell lines growing not only in vitro, but in vivo as well. These cell lines are tTA producing cell lines. Wild type but not mutated *RB* and *p53* genes were deleted/inactivated during tumor growth in SCID mice. Furthermore, no inactivation/deletion was observed for *3PK*, *MLH1*, *rhoA* genes even after two passages in SCID mice. The two multiple cancer regions (3p21.3T and 3p21.3C regions) were identified in lung cancer and kidney cancer. The smallest overlapping homozygous deletion (app. 100 kb) in 3p21.3C region includes 8 genes. All these genes were included in functional gene inactivation test. One gene from homozygous deletion 3p21.3T region also was included in GIT. Until now, *RASSF1A*, *RASSF1C*, *Gene21*, *SEMA3B* and *CACNA2D2* were shown to have growth suppression activity *in vitro* and *in vivo*, and were inactivated in the tumors following SCID mice passage. The results suggested that these genes play important role in the lung and kidney pathogenesis.

Key words: Tumor suppressor gene (TSG), Cloning of deleted sequences (CODE), Cloning of polymorphisms (COP), Cloning of identical sequences (CIS), Gene inactivation test (GIT).

LIST OF PUBLICATIONS

I. J. Li, F. Wang, V. Zabarovska, C. Wahlestedt and E.R. Zabarovsky. (2000) Cloning of polymorphisms (COP): enrichment of polymorphic sequences from complex genomes. *Nucleic Acids Res.* **28(2)**, e1.

II. J. Li, F. Wang, V. Kashuba, C. Wahlestedt and E.R. Zabarovsky. (2000) Cloning of deleted sequences (CODE): a new genomic subtraction method for enriching and cloning deleted sequences. *Biotechniques.* **31(3)**, 788-793.

III. V. Zabarovska, **J. Li**, O. Muravenko, L. Fedorova, E. Braga, I. Ernberg, C. Wahlestedt, G. Klein and E.R. Zabarovsky. (2000) CIS - cloning of identical sequences between two complex genomes. *Chromosome Research* **8**, 77-84.

IV. Zabarovsky ER, Gizatullin R, Podowski RM, Zabarovska VV, Xie L, Muravenko OV, Kozyrev S, Petrenko L, Skobeleva N, **Li J**, Protopopov A, Kashuba V, Ernberg I, Winberg G, Wahlestedt C. (2000) *Not*I clones in the analysis of the human genome. *Nucleic Acids Res.* **28(7)**, 1635-1639.

V. Li J, Protopopov A, Wang F, Senchenko V, Petushkov V, Vorontsova O, Petrenko L, Zabarovska V, Muravenko O, Braga E, Kisselev L, Lerman MI, Kashuba V, Klein G, Ernberg I, Wahlestedt C, Zabarovsky ER. (2002) *Not*I subtraction and *Not*I-specific microarrays to detect copy number and methylation changes in whole genomes. *Proc Natl Acad Sci U S A* **99(16)**, 10724-10729.

VI. M.I. Lerman, J.D. Minna, … V.I. Kashuba, A.I. Protopopov, **J. Li**, G. Klein, E.R. Zabarovsky, … B. E. Johnson. (2000) The 630-kb Lung cancer Homozygous Deletion Region on Human Chromosome 3p21.3: Identification and Evaluation of the Resident Candidate Tumor Suppressor Genes. *Cancer Research* **60**, 6116-6133.

VII. J. Li, A.I. Protopopov, R.Z. Gizatullin, C. Kiss, V.I. Kashuba, G. Winberg, G. Klein and E.R. Zabarovsky. (1999) Identification of new tumor suppressor genes based on in vivo functional inactivation of a candidate gene. *FEBS Letters* **451**, 289-294.

VIII. A.I. Protopopov, **J. Li**, G. Winberg, R. Gizatullin, V.I. Kashuba, G. Klein and E.R. Zabarovsky. (2000) Human cell lines engineered for tetracycline-regulated expression of tumor suppressor candidate genes from a frequently affected chromosomal region 3p21. *J Gene Med.* **4**, 397-406.

IX. K. Dreijerink, E. Braga, I. Kuzmin, L. Geil, F.M. Duh, D. Angeloni, B. Zbar, M.I. Lerman, , E.J. Stanbridge, J.D. Minna, A.I. Protopopov, **J. Li,** V.I. Kashuba, G. Klein, and E.R. Zabarovsky. (2001) The candidate tumor suppressor gene, RASSF1A, from human chromosome 3p21.3 is involved in kidney tumorigenesis. *Proc Natl Acad Sci U S A* **98(13)**, 7504-7509.

ABBREVIATIONS

TSG	Tumor suppressor gene
RCC	Renal cell carcinoma
SCLC	Small cell lung cancer
MCH	Microcell hybrid
MMCT	Microcell mediated chromosome transfer
CGH	Comparative genome hybridization
DNMT	DNA methyl transferase
EST	Expressed sequence tag
STS	Sequence tag site
HAT	Histone acetylase
HDAC	Histone deacetylase
MBP	Methyl-CpG binding protein
SCID	Severe combined immunodeficiency
LOH	Loss of heterozygosity
FISH	Fluorescence in situ hybridization
BAC	Bacterial artificial chromosome
PAC	P1 artificial chromosome
NR	*Not*I representation
CODE	Cloning of deleted sequences
COP	Cloning of polymorphisms
CIS	Cloning of identical sequences
RDA	Representational difference analysis
RFLP	Restriction fragment length polymorphism
SNP	Single-nucleotide polymorphism
GMS	Genomic mismatch scanning
COBRA	Combined bisulfite restriction analysis
LUCA	Lung cancer
GIT	Gene inactivation test
tTA	Transcriptional transactivator
CGI	CpG island
LINE	Long interspersed nuclear element
TMA	Tissue microarray
DMH	Different methylation hybridization
IBD	Identical by descent
SAGE	Serial analysis of gene expression
IRP	Island rescue PCR
RLGS	Restriction landmark genomic scanning
Chr	Chromosome

INTRODUCTION

1. Cancer, genes and genome

1.1 Cancer and cancer caused genes

Cancer is a genetic disease and can be classified into inherited group and sporadic group on the basis of the genetic defect. Most cancers are caused by abnormalities in DNA sequences, which disrupt the harmonious checks and balances that regulate normal cellular growth and development. Throughout life, the DNA in human cells is exposed to mutagens and suffers mistakes in replication, resulting progressive, subtle changes in the DNA sequence in each cell. Occasionally, one of these mutations alters the function of a critical gene, providing a growth advantage to the cell in which it has occurred and resulting in the emergence of an expanded clone derived from this cell. In turn additional mutations in the relevant target genes, and consequent waves of clonal expansion, produce tumor cells that invade surrounding tissues and metastasize.

It is now widely accepted that cancer arises via a multiple process, based on variation and selection. The exponential relationship of cancer incidence to age suggests that multiple events are required. In 1954, Armitage and Doll published age/incidence curves for 17 common types of cancer, from which they conclude that carcinogenesis was at least a six or seven stage process. The multistep progression model of colorectal tumorigenesis from adenoma to carcinoma has been well defined (Fig. 1), which indicates that development of sporadic forms of the colorectal cancer requires six steps and different classes of interacting cancer caused genes including tumor suppressor genes (TSGs), oncogenes and mutator genes are involved.

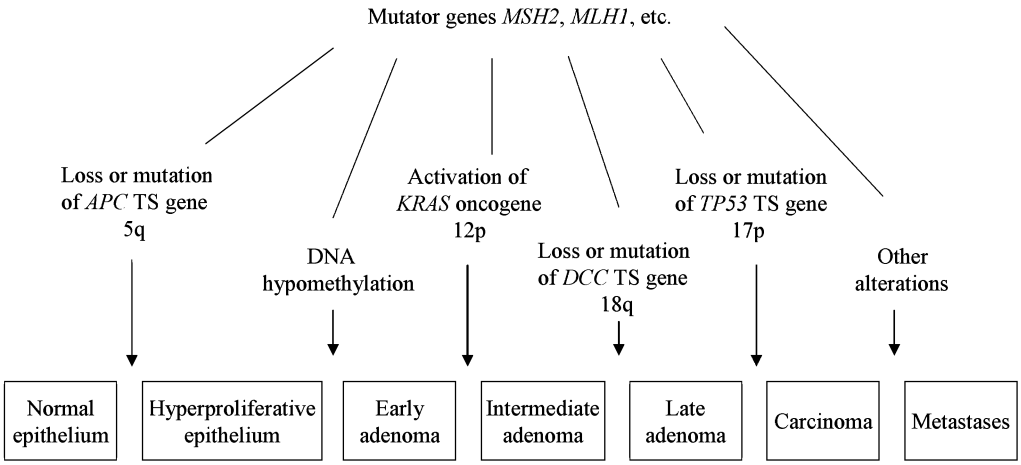

Fig. 1. Fearon and Vogelstein's model for the development of colorectal cancer. It shows the interaction of oncogenes and tumor suppressor genes in colorectal tumorigenesis.

TSGs are wild-type alleles of genes that play negative regulatory roles in cell proliferation, differentiation, and other cellular processes. It is their loss or inactivation that is oncogenic. Germline mutations in TSGs strongly predispose to cancer, and they are also mutated somatically in sporadic forms of cancer. Tumor suppression was first demonstrated in somatic cell hybrids produced by fusion between normal and tumor cells (Sager 1985; Harris 1986; Sager 1986; Klein 1987; Harris 1988). It was evident that hybrids were initially nontumorigenic as determined by their inability to grow in immunocompromised hosts, like the normal parent; subsequently, as the hybrid clones were propagated in culture, chromosomes were lost, and reversion to tumor-forming ability occurred. It was demonstrated unambiguously that chromosome carrying TSGs were lost when suppressed hybrids regained tumor-forming ability. Tumor suppression associated with the transfer of single human chromosomes via microcell fusion (Table 1) also provided the evidence for existence of TSGs in human genome.

Table 1. Tumor suppression associated with the transfer of single human chromosomes via Microcell fusion

	Tumor suppression chromosome			
Tumor cell line	Expected*	Suppressed	Non-suppressed	Reference
Cervical carcinoma (HeLa)	11	11	X	Saxon et al. 1986
Cervical carcinoma (SiHa)	--	11	--	Oshimura et al. 1990
Retinoblastoma	13	13	--	Banerjee et al. 1992
Renal cell carcinoma	3	3	11	Shimizu et al. 1990
Wilms tumor	11p13	11p15	X, 13	Dowdy et al. 1991
Colorectal carcinoma (COKFu)	5, 17, 18	5, 18	11	Tanaka et al. 1991
Colorectal carcinoma (SW480)	5, 17, 18	4, 17, 18	15	Goyette et al. 1992
Endometrical carcinoma	--	1, 6, 9	11	Yamada et al. 1990
Melanoma	6	6	--	Trent et al. 1990
Neuroblastoma (NGP)	1	1p, 17	11	Bader et al. 1991
Neuroblastoma (SK-N-MY)	1	1	11	Oshimura et al. 1989
Fibrosarcoma (HT1080)	1	1, 11	2, 7, 12	Kogoh et al. 1990
Ovarian carcinoma (HEY)	--	3p21.1-21.2	11	Rimessi et al. 1994
Breast carcinoma	11	11	--	Negrini et al 1992
Rhabdomyosarcoma (A204)	11	11	--	Oshimura et al. 1990
Rhabdomyosarcoma (RD)	11	11p, 11q	--	Loh et al. 1992
Bladder carcinoma	13	13	--	Banerjee et al. 1992
Prostate carcinoma	13	13	--	Banerjee et al. 1992
Nasopharyngeal carcinoma (HONE1)	9, 11, 17	11q13, 11q22-23	17, 9(del.p21)	Cheng et al. 2000

(*Expected from cytogenetic and RFLP analyses)

In most cases, transfer of a particular single copy of the normal chromosome is sufficient to induce growth inhibition in vitro or tumor suppression in vivo. The second piece of evidence for existence of TSGs came from studies of hereditary cancers. The discovery of the retinoblastoma (*RB*) gene stems from a prediction of Knudson, based on the statistical of age/incidence curves of familial vs. sporadic retinoblastoma. Knudson suggested (Knudson, 1985) that retinoblastoma arose from two sequential mutational events on both alleles of the same gene; this is two-hit hypothesis. Since the isolation of the *RB* gene in 1986, around 30 TSGs have been identified (Table 2). The suppressor genes already identified are involved in cell cycle control, growth and transcriptional regulation, signal transduction, angiogenesis, and development, indicating that they contribute to a broad array of normal and tumor related functions. It is proposed that TSGs provide a vast resource for anticancer therapy.

Table 2. Tumor Suppressor Genes

Symbol	Accession	Name	Locus	Cancer syndrome	Cancer type (germline and/or somatic mutations)
RB1	P06400	Retinoblastoma gene	13q14.1- q14.2	Familial retinoblastoma	Retinoblastoma, sarcomas, breast cancer, small cell lung cancer
TP53	P04637	Tumor suppressor p53 gene	17p13.1	Li-Fraumeni syndrome	Sarcoma, adenocortical carcinoma, glioma, other tumor types
APC	P25054	Adenomatous polyposis of the colon gene	5q21-q22	Familial polyposis of the colon	Colorectal, pancreatic cancers, desmoids, hepatoblastoma
BRCA1	P38398	Familial breast/ovarian cancer gene 1	17q21	Hereditary breast/ovarian cancer	Hereditary breast/ovarian cancers
BRCA2	P51587	Familial breast/ovarian cancer gene 2	13q12.3	Hereditary breast/ovarian cancer	Hereditary breast/ovarian cancers
CDH1	P12830	Cadherin 1 gene	16q22.1	Familial gastric carcinoma	Lobular breast cancer
CDKN2A	P42771	Cyclin-dependent kinase inhibitor 2A (p16) gene	9p21	Cutaneous malignant melanoma 2	Melanoma, other tumor type
CYLD	CAB93533	Familial cylindromatosis gene	16q12-q13	Familial cylindromatosis	Cylindromas
EP300	Q09472	300-kD E1A-binding protein gene	22q13	N/A	Colorectal, breast, pancreatic cancers
EXT1	Q16394	Multiple extoses type 1 gene	8q24.11-q24.13	Multiple exostoses type 1	Exostoses, osteosarcoma
EXT2	Q93063	Multiple extoses type 2 gene	11p12-p11	Multiple exostoses type 1	Exostoses, osteosarcoma
CDKN1C	P49918	Cyclin-dependent kinase inhibitor 1C gene	11p15.5	Beckwith-Wedernmann syndrome	Wilms' tumor, rhabdomyosarcoma
STK11	Q15831	Serine/threonine kinase 11 gene	19p13.3	Peutz-Jeghers syndrome	Jejunal harmartomas, ovarian tumors, testicular and pancreatic cancers
MAP2K4	P45985	Mitogen-activated protein kinase kinase 4	17p11.2	N/A	Pancreatic, breast, colon cancers
MEN1	O00255	Multiple endocrine neoplasia type 1 gene	11q13	Multiple endocrine neoplasia type 1	Parathyroid/pituitary adenoma, islet cell carcinoma, carcinoid tumors
MLH1	P40692	*E. coli* MutL homologue gene	3p21.3	Familial non-polyposis colorectal cancer	Colorectal, endomentrial, ovarian cancer
MSH2	P43246	*E. coli* MutS homologue 2 gene	2p22-p21	Familial non-polyposis colorectal cancer	Colorectal, endomentrial, ovarian cancer
NF1	P21359	Neurofibromatosis type 1 gene	17q11.2	Neurofibromatosis type 1	Neurofibroma, glioma
NF2	P35240	Neurofibromatosis type 2 gene	22q12.2	Neurofibromatosis type 2	Meningioma, acoustic neuroma
PRKAR1A	P10644	Protein kinase A type 1-α regulatory subunit gene	17q23-q24	Carney complex	Myxoma, endocrine tumors
PTCH	Q13635	Homologue of *Drosophila* patched gene	9q22.3	Nevoid basal cell carcinoma syndrome	Basal cell carcinoma, medulloblastoma
PTEN	O00633	Phosphatase and tensin homologue gene	10q23.3	Cowdens syndrome	Harmartomas, glioma, prostate and endometrial cancers
SDHD	O14521	Succinate dehyrdogenase cytochrome B small subunit gene	11q23	Familial paraganglioma	Paraganglioma
MADH4	NP_005350	Homologue of *Drosophila mothers against decapentaplegic* 4 gene	18q21.1	Juvenile polyposis	Gastrointestinal polyps, colorectal and pancreatic cancers
SMARCB1	Q12824	Swi/Snf5 matrix-associated actin-dependent regulator of chromatin gene	22q11	Rhabdoid predisposition syndrome	Malignant rhabdoid tumors
TSC1	Q92574	Tuberous sclerosis 1 gene	9q34	Tuberous sclerosis 1	Hamartomas, renal cell carcinoma
TSC2	P49815	Tuberous sclerosis 2 gene	16p13.3	Tuberous sclerosis 2	Hamartomas, renal cell carcinoma
VHL	P40337	Von Hipple-Lindau syndrome gene	3p26-p25	Von Hipple-Lindau syndrome	Renal cell carcinoma, hemangioma, phaeochromocytoma
WT1	P19544	Wilms tumor 1 gene	11p13	Familial Wilms tumor	Wilms tumor

The proto-oncogenes encode proteins, which are components of the cell signaling pathways. In the normal cell, the expression of these proto-oncogenes is tightly controlled and they are transcribed at the appropriate stages of growth and development of cells. Mutations in these genes act dominantly and lead to gain in function accelerating cell division. Oncogenes were initially discovered as retroviral transmitted tumor-causing agents. The realization that such retroviral oncogenes constitute specifically altered versions of cellular genes: proto-oncogenes was a landmark discovery in cancer research. Moreover, the studies on oncogene functions have been instrumental in delineating many of the paradigms of cellular signal transduction.

In contrast to the original studies in animals, oncogenic activation through retroviral transmission does not appear to be a major factor in human tumorigenesis (Munger 2002). Frequently cellular proto-oncogenes are activated by gain-of-function mutations which produce dominant phenotypes by a variety of mechanisms including point mutation, chromosomal translocation resulting in the production of a fusion protein, rearrangements, deletion of part of the protein, aberrantly expression due to amplification, increased promoter activity, or protein stabilization and loss of appropriate control. A direct method, DNA transfection assay, was used to identify those sequences in tumor cells, which were responsible for uncontrolled cell proliferation. Chromosome translocation and amplification in tumor cells can also be used for identification of cellular oncogenes. Proto-oncogenes usually function as growth factors, growth factor receptors, signal transducers, nuclear proto-oncogenes and transcription factors (Table 3).

Table 3. Function of cell-Derived Oncogene Products

Function		Genes
Growth Factors	*sis*	PDGF B-china growth factor
	int-2	FGF-related growth factor
	hst (*KS3*)	FGF-related growth factor
	FGF-5	FGF-related growth factor
Receptor and Nonreceptor	*src*	Membrane-associated nonreceptor protein-tyrosine kinase
Tyrosine Kinases	*yes*	Membrane-associated nonreceptor protein-tyrosine kinase
	fgr	Membrane-associated nonreceptor protein-tyrosine kinase
	lck	Membrane-associated nonreceptor protein-tyrosine kinase
	fps/fes	Nonreceptor protein-tyrosine kinase
	abl/bcr-abl	Nonreceptor protein-tyrosine kinase
	ros	Membrane associated receptor-like protein-tyrosine kinase
	erbB	Truncated EGF receptor protein-tyrosine kinase
	neu	Receptor-like protein-tyrosine kinase
	fms	Mutant CSF-1 receptor protein-tyrosine kinase
	met	Soluble truncated receptor-like protein-tyrosine kinase
	trk	Soluble truncated receptor-like protein-tyrosine kinase
	kit (W	Truncated stem cell receptor protein-tyrosine kinase
	sea	Membrane associated truncatedreceptor-like protein-tyrosine
	ret	Truncated receptor-like protein-tyrosine kinase
Membrane-Associated G Proteins	H-*ras*	Memberane-associated GTP-binding/GTPase
	K-*ras*	Memberane-associated GTP-binding/GTPase
	N-*ras*	Memberane-associated GTP-binding/GTPase
	gsp	Mutant activated form of Gs α
	gip	Mutant activated form of Gi α
Cytoplasmic Protein-Serine	*raf/mil*	Ctoplasmic protein-serine kinase
Kinases	*pim-1*	Ctoplasmic protein-serine kinase
	mos	Ctoplasmic protein-serine kinase (cytostatic factor)
Cytoplasmic Regulators Nuclear	*crk*	SH-2/3 protein that binds to phosphotyrosine-containing protein
Transcription Factors	*myc*	Sequence-specific DNA-binding protein

	myb	Sequence-specific DNA-binding protein
	lyl-1	Sequence-specific DNA-binding protein
	P53	Mutant form may sequester wild-type p53 growth suppressor
	fos	Combines with c-jun product to form AP-1 transcription factor
	jun	Sequence-specific DNA-binding protein; part of AP-1
	erbA	Dominant negative mutant thyroxin (T3) receptor
	rel	Dominant negative mutant NF-κB-related protein
	ets	Sequence-specific DNA-binding protein
	pbx	Chimeric E2A-homeobox transcription factor
Receptors Lacking Protein Kinase Activity	*mas*	Angiotensin receptor

(The table is selected and obviously incomplete. These oncogenes were originally detected as retroviral oncongenes or tumor oncogenes)

A single oncogene is insufficient for transformation; collaboration between oncogenes could result in fully transformation. It was demonstrated that disruption of the intracellular pathways regulated by large-T (LT), oncogenic *ras* and telomerase suffices to create a human tumor cell (Hahn et al., 1999, Elenbaas et al., 2001). These minimal changes involved the inactivation of the p53 and RB pathways achieved by LT, telomere maintenance conferred by *hTERT*, and acquisition of a constitutive mitogenic signal provided by oncogenic H-*ras*.

Interaction between TSGs and proto-oncogenes was found in cells. For example, suppressor gene *NF1* that is GTPase activating protein (GAP) converts active RAS p21 back to inactive GDP bound form. This interaction maintains homeostasis in normal cells, while in cancer, the homeostasis is impaired, and mutations of both proto-oncogenes and TSGs are part of the malignant process. TSGs and oncogenes are directly involved in the controls of cell cycle, which involves a series of events resulting in DNA duplication and cell division. For example, the one effect of the known TSG *p53* activation is a block in the cell-division cycle by stimulating the expression of a cyclin-dependent kinase inhibitor, $p21^{WAF1/CIP1}$ (Vogelstein et al., 2000). The cell cycle is divided into four distinct phases. The first gap phase (G1), DNA replication (S) and second gap phase (G2) together make up interphase; this is followed by mitosis (M). The cell cycle progression from one stage to the next is carefully controlled by the sequential activation and degradation of the cyclins, activation of cyclin partners, the cyclin-dependent kinases (CDKs) and their inhibitor proteins known as cyclin-dependent kinase inhibitors (CDKIs) (Fig. 2). In normal cells this process is carefully controlled but in tumor cells, mutations in the genes associated with the cell cycle result in progression of cells with damaged DNA through the cycle.

Mutation or deletion of mutator genes involved in DNA repair, e.g. *MSH2* on human chromosome 2p21 (homolog of the bacterial *MutS* and the yeast *MSH2* mismatch repair genes) or functionally similar *MLH1* gene on chromosome 3p21, leads to specific multi-cancer syndromes. Genomic instability at short repeated sequences that is termed replication error positive (RER positive) may reflect up to thousand-fold increase in mutation rates. With typical mutation rates of 10^{-6} per gene per cell and perhaps six specific mutations for turning a normal cell into a malignant cancer cell, the probability of this happening to any one of the 10^{14} cells in a person is 10^{14} X $(10^{-6})^6$, or $1:10^{22}$. Mutation rate will be 10^{-3} when mismatch repair genes are mutated, then the probability of malignant cell appearance in a person is 10^{14} X $(10^{-3})^5$, or 1:10. Mutations in mutator genes result in a higher than normal mutation rate, allowing the accumulation of mutations in other genes such as *p53* or *APC*, finally lead to high probability of tumor development.

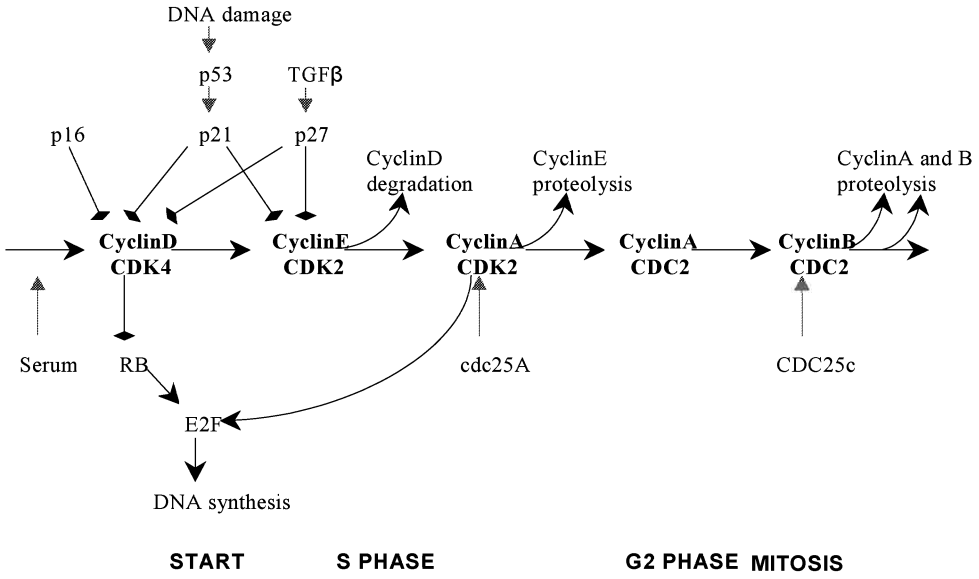

Fig. 2 Interaction of the cyclones/CDKs and three of the CDKIs in the different cell cycle stages. Steps involving activation are shown in red, and those that are inhibitory are shown in black.

The genes involved in apoptosis, angiogenesis and tumor metastasis are also important for tumor development. This thesis mainly focuses on the identification of new TSGs associated with the tumorigenesis of kidney cancer, lung cancer and other cancers.

1.2 Human genome draft sequence and cancer research

In February of 2001, the Human Genome Project and Celera Genomics (www.celera.com) reported the first draft of the human genome sequence, covering 85 to 90% of the entire genome. Analysis of the genome sequence revealed 26,000 to 40.000 protein-encoding transcripts. Only 1.1% of genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided location of 2.1 million single-nucleotide polymorphisms (SNPs). A random pair of human haploid genomes differed at a rate of 1bp per 1250 on average and less than 1% of all SNPs resulted in variation in proteins.

The availability of an ultra-high density SNP map opens the possibility of studying by association genetic factor important in complex genetic traits like cancer in the human, taking advantage of the fact that genetic markers in close proximity to mutant genes may be in linkage disequilibrium (LD) to them (Johnson and Todd 2000; Risch 2000). Association studies can be done with a genome-wide approach or with a candidate gene approach. Genotyping of SNPs will likely be a major part of every genetic association study, and the appropriate genotyping method is critical to the success of the study. SNPs will be used as molecular markers associated with a biologic phenomenon such as enhanced susceptibility to disease.

To cancer researchers, the availability of the human genome sequence provides the opportunity to discover oncogenes and TSGs. The hallmark of a cancer genome is that it is replete with DNA alterations that perturb the normal function of proto-oncogenes and TSGs. A powerful approach to pinpoint gene alterations in cancer genomes has been initiated under

the Cancer Genome Anatomy Project (CGAP), that is a network of cancer researchers deciphering the genetic changes that occur during cancer formation and progression (http://cgap.nci.nih.gov). The project focuses on automated sequencing of cDNA libraries from 13 precancerous and 117 cancers samples. A better approach to understanding the cancer genome might be the sequencing of the entire normal and paired cancer genomes. This project would reveal all the differences (especially in intergenic and intragenic regulatory sequences) within the tumor genome, rather than the coding sequence only, on which most analyses concentrate.

Comparative genomics can be used to understand genome structures, infer gene function, and identify conserved regulatory sequences. It was shown that novel genes and potential regulatory sequences have been identified by small-scale comparative genomics of orthologous chromosomal regions between the human and the mouse (Loots et al., 2000; Onyango et al., 2000). It will be beneficial to compare entire human genome with mouse genome, *Saccharomyce cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster* genomes, which have been almost entirely sequenced. Most significant is the attempt to generate chromosome rearrangements in the mouse or transgenic mouse as models for human cancers, using cre-loxP technology (Zheng et al., 2001).

One of the aims behind the strategy of human genomic sequencing is to provide an inventory of all the genes and regulatory sequences required to build an organism. Annotation of the human genome is certain to produce benefits not only for understanding basic biology, but also for identifying the molecular basis of disease and for accelerating the rate of drug discovery and development. The repositories of genes and their regulatory sequences represent the starting point of the new challenge of post-sequence functional genomics, which is to understand how these components interact and function (Lockhart and Winzeler 2000). DNA microarrays and tissue microarrays (TMAs) provide a powerful approach to identify large number of new candidate genes, and rapidly validate their clinical impact in large series of human tumors. These technologies will soon lead to better molecular understanding of tumors, and accelerate the identification of new prognostic markers or therapeutic targets.

1.3 Genes on human chromosome 3 with lung cancer and kidney cancer

Aberrations on the short arm of human chromosome 3 occur frequently in renal cell carcinoma (RCC), von Hipple Lindau disease, lung cancer and other malignancies (Kok et al., 1987, Kovacs et al., 1988). Different data suggest that 3p may carry multiple TSGs.

Microcell mediated chromosome transfer (MMCT) studies showed tumor suppression activity of chromosome 3 in RCC (Shimizu et al., 1990; Sanchez et al., 1994; Ohmura et al. 1995), lung adenocarcinoma (Satoh et al., 1993), ovarian carcinoma (Rimessi et al., 1994) and nasopharyngeal carcinoma (Cheng et al., 1998, 2000). Part of chromosome 3 showed the tumor suppression as well. Sanchez et al. (1994) reported that 3p14-p12 region showed suppression in non-papillary renal cell carcinoma. Killary et al. (1992) reported that a 2 Mb chromosome 3 fragment HA(3)BB9F from 3p21.3 spanning the three SCLC and one breast cancer homozygous deletions suppressed tumorigenicity of mouse fibrocarcinoma A9 in athymic nude mice. Later, an 80 kb P1/294 clone located inside the same region containing gene *SEMA3F*, *GNAT1*, *G17* and *GNAT2* showed suppression of A9 tumor growth in mice (Todd et al., 1996). Recently, it was shown that genes located on 3p14.2-p21, 3p12-21.1 and 3p21.3-p22 suppressed telomerase activity in RCC and breast cancer was reported (Tanaka et al., 1998; Cuthbert et al., 1999).

Deletion mapping using microsatellite markers and the detection of homozygous deletions represented until now the most powerful method to localize potential TSGs. Loss of heterozygosity (LOH) in 3p frequently happened in different human tumors especially RCC

and lung cancer. But reports were different, for example, on the extent of 3p losses in different tumors, with some paper reporting large terminal deletions, and others claiming interstitial deletions (Kok et al., 1997; Braga et al., 1999, 2002; Wistuba et al., 2000; Girard et al., 2000). One of the reasons could be that most of these studies did not use precise microdissection of tumor from normal cells that was unavoidable source of errors in LOH studies of solid tumors. In this regard, Wistuba et al. have performed high-resolution LOH studies on 97-lung cancer and 54 preneoplastic/preinvasive microdissected respiratory epithelial samples using 28 chr 3 markers. Allelic losses of 3p were detected in 96% of the lung cancers and in 78% of the preneoplastic/preinvasive lesions. The allele losses were often multiple and discontinuous, with areas of LOH interspersed with areas of retention of heterozygosity. Analysis of all of the data indicated multiple regions of localized 3p allele losses. The 3p21.3 region, 3p14.2 (*FHIT*/*FRA3B*) and 3p21 (*ROBO1*, also called *DUTT1*) regions were common. Seven homozygous deletions (Table 4) were found along 3p. The candidate genes located in these seven homozygous deleted regions and information confirming their role in tumor suppression have been discussed in detail elsewhere (Zabarovsky et al., 2002).

Chromosome 3-specific *Not*I linking and jumping libraries were constructed in our group to map and clone TSGs whose inactivation play important roles in the development of kidney and lung cancer. It will be described in part 2.3. Several candidate genes isolated from chromosome 3 will be described below in part 4.3.

Table 4. Chromosome 3p regions most frequently affected in lung, kidney cancers.

Chromosome 3p region	Homozygous deletion	TSG activity	Genes	Mutations/ intragenic homozygous deletions	Methyla-tion in tumors	Growth suppression in vitro	Tumor suppression in vivo	Controlled suppression effect
3p24-p26	Yes (nasopharynx)	Yes	*VHL*	Yes/Yes	Yes	Yes	Yes	Yes, tet-system
			RARβ	No/No	Yes	Yes	Yes	Not done
3p21.3T (AP20)	Yes (lung, kidney, etc.)		*DLEC1*	No/No	No	Yes (not in all cell line)		
3p21.3	Yes (mesothelioma)	Yes	*CTNNB1*	Rare/Yes				
			Genes in CER1					
			HD-PTP	No/No				
3p21.3C (LUCA)	Yes (lung, breast, etc.) (NCI-H1450, NCI-H740, GLC20, HCC1500)	Yes	*RBM6*	No				Not done
			RBM5	No/No		Yes, but moderate in HT-1080		Not done
			SEMA3F	No/No	No	More weak than *SEMA3B*	Yes	Not done
			SEMA3B	Rare/No	Yes	Yes	Yes	Yes, tet-system
			HYAL1	Rare/No	Yes	Yes	Yes	Yes, tet-system
			HYAL2	No/Yes				
			Fus1	Rare/Yes	No	Yes	Yes	Yes, ecdysone regulated
			RASSF1A	Yes/No	Yes	Yes	Yes	Yes, tet-system
			Blu	Rare/No	Yes	Yes		
			CACNA2D2	No/No	Yes	Yes	Yes	Yes, tet-system
3p21.1-p21.2	Yes (breast)	Yes	*ARP1*	Yes/No				
			BAP1	No/Yes		Yes		
			DRR1	Yes, but not in expressed gene/No		Yes		
3p14	Yes (lung, renal, etc.)	Yes	*FOXP1*					
			FHIT	No/Yes	Yes	Yes (not in all cell lines)	Yes, but mutant FHIT has the same effect.	Not done
3p12-p13	Yes (lung, breast) (U2020, HCC38, NCI-H219X)	Yes	*DUTT1*	No/Yes	Yes	No	No	Not done

2. CpG islands and methylation

2.1 CpG islands, methylation and genes

2.1.1 CpG islands and genes

The mammalian genome can be conveniently divided into two fractions with respect to DNA methylation (Cooper et al., 1983; Bird et al., 1985). In the major fraction the dinucleotide CpG (about 98% of the total) occurs on average every 50 to 100 bp and is heavily methylated. In the minor fraction (about 2%), CpG occurs approximately every 10 bp and is nonmethylated. The minor fraction is distributed through the genome in 45,000 short regions of 1 kb, known as CpG islands. These colocalize with the 5' end of genes (Bird 1987). In humans, about 60% of genes are associated with CpG islands, including all housekeeping genes so far analysed and about 40% of tissue-restricted genes. In most cases, the island contains the promoter and one or more exons of an associated gene. Various computer programs attempt to identify CpG islands on the basis of primary sequence alone without testing for the absence of cytosine methylation. The definition proposed by Gardiner-Garden and Frommer (1987) was embodied in a computer program to search the draft human genome sequence for CpG islands, using both the full sequence and the sequence masked to eliminate repeat sequences. The number of regions satisfying the definition of a CpG island was 50,267 in the entire sequence and 28,890 in the repeat-masked sequence. The difference reflects the fact that some repeat elements (notably Alu) are GC-rich. Although some of these repeat elements may function as control regions, it seems unlikely that most of the apparent CpG islands in repeat sequences are functional. The predicted number of 28,890 CpG islands is reasonably close to the previous estimate of about 35,000 (Antequera and Bird 1993). Most of the islands are short, with 60-70% GC content (Table 5). More than 95% of the islands are less than 1,800 bp long, and more than 75% are less than 850 bp. The longest CpG island (on chromosome 10) is 36,619 bp long, and 322 are longer than 3,000 bp. Some of the larger islands contain ribosomal pseudogenes, although RNA genes and pseudogenes account for only a small proportion of all islands (<0.5%). The small islands are consistent with their previous hypothesized function, but the role of these larger islands is uncertain (International Human Genome Sequencing Consortium, 2001).

Table 5. Number of CpG islands by GC content

GC content of island	Number of islands	Percentage of islands	Nucleotides in islands	Percentage of nucleotides in islands
>80%	22	0.08	5,916	0.03
70-80%	5,884	20	3,111,965	16
60-70%	18,779	65	13,110,924	66
50-60%	4,205	15	3,589,742	18
Total	28,890	100	19,818,547	100

Potential CpG islands were identified by searching the draft genome sequence one base at a time, scoring each dinucleotide (+17 for GC, -1 for others) and identifying maximally scoring segments. Each segment was then evaluated to determine GC content (≥ 50%), length (>200) and ratio of observed proportion of GC dinucleotides to the expected proportion on the basis of the GC content of the segment (>0.60), using a modification of a program developed by G. Michlem.

The density of CpG islands varies substantially among some of the chromosomes. Most chromosomes have 5-15 islands per Mb, with a mean of 10.5 island islands per Mb. However, chromosome Y has an unusually low 2.9 islands per Mb, and chromosome 16, 17 and 22 have 19-22 islands per Mb. The extreme outlier is chromosome 19, with 43 islands per Mb. Similar trends are seen when considering the percentage of bases contained in CpG islands. The relative density of CpG islands correlates reasonably well with estimates of relative gene density on these chromosomes. It was shown that CpG islands are predominantly found in the early replication (R band) regions of the human genome. Conversely, late replicating (G band) DNA is sparsely populated with islands. The highest concentration of CpG island is in a subset of R bands, most of which are known as T bands. (Craig and Bickmore 1994; Aissani and Bernardi 1991). Chicken CpG islands are highly concentrated on the microchromosomes, whereas macrochromosome 1-6 are comparatively gene-poor (McQueen et al., 1996). Unlike in other vertebrate genomes that have been examined (human and Chicken), extreme clustering of CpG islands was not seen in the mouse genome. Despite the more even distribution of CpG islands in mouse at a gross chromosomal level, at finer resolution concentration of CpG islands are seen to correspond to the R-band early replicating regions of the genome (Cross et al., 1997).

As CpG islands are effective markers for identifying genes, cloned CpG islands would be useful reagents for mapping genes. Furthermore, each CpG island is present in genomic DNA in equimolar quantities unlike transcripts represented by cDNA libraries. Thus, the isolation and cloning of CpG islands would provide access to a large number of genes independent of developmental stage or tissue of expression. Cross et al. (1994) has developed a procedure (Fig. 3) for bulk isolation of CpG islands from human genomic DNA by construction an affinity matrix that contains the methyl-CpG binding domain (MBD) from the rat chromosomal protein MeCP2, attached to a solid support. A column containing the matrix fractionates DNA according to its degree of CpG methylation, strongly retaining those sequences that are highly methylated. The strategy can be divided into four steps.

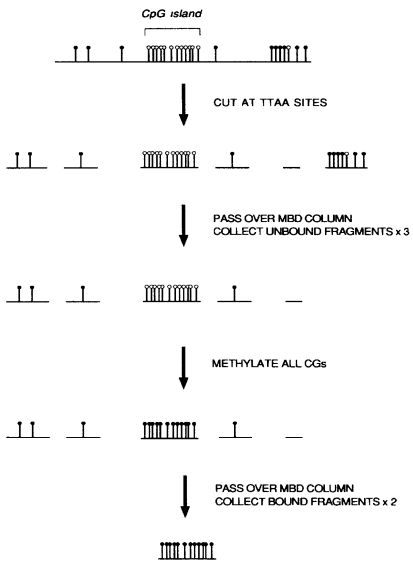

Fig. 3 Flow diagram illustrating the strategy for the purification of MseI fragments containing CpG islands. Vertical lines indicate the position of CpG. Open and solid circles denote unmethylated and methylated CpGs, respectively

In the first step, total genomic DNA is fragmented with *Mse*I. This enzyme was chosen because its recognition site (TTAA) is found relatively rarely within CpG islands (once per 1,000 bp) but frequently in bulk DNA (once per 140 bp). Therefore *Mse*I is expected to give predominantly intact CpG islands, plus small fragments from bulk DNA. The second step involves removal of *Mse*I fragments that contain clusters of methylated CpGs, by selecting only fragments that bind weakly to the column. This "stripping" procedure is required to remove highly methylated *MseI* fragments that would contaminate the final purified CpG island fraction. During the third step the bacterial methyltransferase, M.*SssI* methylate all non-methylated CpGs in the fractions that have been stripped. On most fragments, nearly all CpGs are methylated already and therefore their affinity for the MBD column is not changed. CpG island fragments, on the other hand, are converted from weak binding to strong-binding molecules. By selecting the fragments that elute at high salt, step four should yield a fraction that is highly enriched for CpG islands. CpG islands libraries constructed by this strategy have been established for humans (Cross et al., 1994), mouse (Cross et al., 1997), and chickens (Mcqueen et al., 1996) and these libraries have been used to examine the organization of genes and genomes (Cross and Bird 1995; Mcqueen et al., 1996; Cross et al., 1997).

In positional cloning project the task is usually to detect genes within clones containing between 35 kb (cosmids) and 300 kb (BACs) of genomic DNA. Several different methods, such as exon trapping (Krizman et al., 1997) and direct cDNA selection (Lovett et al., 1991; Parimoo et al., 1991), have been devised to allow the detection of putative gene sequences. Complementary approaches aimed at isolating largely intact CpG islands from large genomic clones by exploiting their sequence characteristics to identify the gene sequences have been developed. The first is a PCR-based method, island rescue PCR (IRP), which is dependent on an Alu sequence being present close to a CpG island (Valdes et al., 1994). This has been used for YAC clones, as the Alu-specific primers do not bind yeast DNA. In the second method, segregation of partly melted molecules (SPM) fragments from CpG islands is selected by the retention of partly melted DNA fragments in a denaturing gradient gel (Shiraishi et al., 1998). In both these methods only parts of CpG islands would be recovered. Kato and Sasaki (1998) have reported a method to quickly identify and localize CpG islands in large genomic fragments by partial digestion with *Hpa*II and *Hha*I. This method is based on the presence of multiple H*pa*II and H*ha*I sites in CpG islands, at a frequency 30 times higher than in the rest of the genome. The steps include complete digestion of DNA with a rare-cutting restriction endonuclease (to produce large fragments with defined ends), partial digestion with H*pa*II and *Hha*I, and subsequent Southern hybridization with an end probe.

The fact that nonmethylated sites for rare-cutting restriction enzymes such as *Not*I (GCGGCCGC, methylation sensitive enzyme), *Eag*I (CGGCCG), *Sac*II (CCGCGG), *Bss*HII (GCGCGC), *Sma*I (CCCGGG), *Nae*I (GCCGGC) and *Nar*I (GGCGCC) are concentrated in CpG islands (Brown and Bird 1986) were helpful for cloning CpG islands and searches for genes. Libraries of *Not*I flanking sequences within CpG islands have been constructed. Different methods for constructing *Not*I linking library have been proposed (Arenstorf et al., 1991; Hattori et al., 1993; Wallace et al., 1989; Ito and Sakaki 1988; Saito et al., 1991). In our group, we have improved the procedure for construction *Not*I linking and jumping libraries. A number of chr-specific *Not*I linking and jumping libraries of human chromosome 3 have been constructed (Zabarovsky et al., 1990, 1991, 1994; Kashuba et al., 1999). Genes, including new genes have been identified using *Not*I clones and mapped to chromosome 3 (Allikmets et al., 1996; Szeles et al., 1996; Protopopov et al., 1996; Kashuba et al., 1997). Finally, *Not*I flanking sequences from the total human genome have been generated and chromosome 3 specific *Not*I clone microarrays are ready (Zabarovsky et al., 2000; Li et al., 2002). *Not*I linking and jumping libraries will be discussed in 2.3.

Methylation of cytosine is the only known endogenous modification of DNA in mammals and occurs by the enzymatic addition (S-adenyl methionine, SAM) of methyl group to the carbon-5 position of cytosine (Doerfler 1983). The majority of 5'-methylcytosine in mammalian DNA is present in 5'-CpG-3' dinucleotides (Riggs and Jones 1983). Non-CpG sequences such as 5'-CpNpG-3' (Clark et al., 1995) or non-symmetrical 5'-CpA-3' and 5'-CpT-3' (Woodcock et al., 1997) may also exhibit methylation, but generally at a much lower frequency. In mouse embryonic stem cells, however, non-CpG methylation comprises 15-20% of total 5'-methylcytosine (Ramsahoye et al., 2000). DNA methylation is present in organisms from bacteria to human. In bacteria, methylation is a part of defence mechanism to reduce the gene transfer between species. Particular mutant strains of bacterial that lack detectable methylation nevertheless survive and proliferate. In contrast to bacteria, deletion of any one of three DNA methyltransferase genes from mice is lethal, suggesting that methylation has additional and indispensable function in mammals (Bestor et al., 1988; Okano et al., 1999). Establishing DNA methylation patterns proceeds through defined phases during development of an organism. In general, germ cells of females are less methylated than those of males, and gamete methylation patterns are erased by a genome wide demethylation near the eight-cell stage of blastocyst formation (Monk et al., 1987; Kafri et al., 1992). During the implantation stage, methylation patterns are established following a wave of de novo methylation (Monk et al., 1987; Kafri et al., 1992). In the adult, the amount and pattern of methylation are tissue and cell type specific and there is evidence for age-related methylation changes of CpG islands in the promoter of genes, including the estrogen receptor gene and *MYOD1* (Issa 2000). Methylation patterns of certain genomic regions appear polymorphic between people and can be inherited, suggesting either the persistence of certain methylation at all stages of development, or encryption of methylation pattern (Silva and White 1988).

Two models by which CpG islands become methylated in cancer have been outlined (Baylin et al., 1998; Jones 1999; Tycko 2000). One proposed mechanism involves the loss of factors that normally protect the CpG island from methylation. Protective factors might be structure proteins (Zardo and Caiafa 1998) or transcription factors (Brandeis et al., 1994). A second model suggests that aberrant CpG island methylation is an active process and causes inappropriate gene silencing. Three DNA methyltransferases, DNMT1, DNMT3A, and DNMT3B, have been identified in mammalian cells (Bestor et al., 1988; Okano et al., 1999). Initial methylation of DNA requires *de novo* methylase activity that is mostly present during early embryonic development (Jahner et al., 1982). DNMT1 uses hemimethylated DNA as a preferential template (Bouchrd and Momparler 1983). The hemimethylated pattern of the parent strand is recognized and then faithfully reproduced on the daughter strand, allowing this feature to be heritable after DNA replication and cell division. Thus DNMT1 is termed a maintenance methylase. *DNMT1* is ubiquitously expressed in somatic tissue (Bestor et al., 1988) and interacts with proliferating cell nuclear antigen (PCNA) at the replication fork (Leohardt et al., 1992; Chuang et al., 1997). DNMT1 also interacts with HDAC2 and DMAP1 (DNMT1 associated protein) to mediate transcriptional repression (Rountree et al., 2000). DNMT3A and DNMT3B appear to function as *de novo* methylases since they can methylate hemimethylated and unmethylated DNA with equal efficiencies (Xie et al., 1999). The expression of DNMT3B was significantly increased in tumor suggesting that it may have a role in tumorigenesis (Robertson et al., 1999). A specific mammalian demethylase that uses methylated CpG DNA as a substrate has been identified (Bhattacharya et al., 1999). The role of this interesting enzyme with respect to maintenance of DNA methylation patterns and gene expression remains to be elucidated.

2.1.2 Hypomethylation, hypermethylation and genes

The delicate organization of DNA methylation and chromatin states that regulates the normal cellular homeostasis of gene expression patterns become unrecognizable in the cancer cell. Fig. 4 showed that DNA methylation is at the center of the normal and malignant behavior of

the cell. Altered methylation patterns are known to occur in the DNA of cancer cells. Two patterns have been observed: wide areas of global hypomethylation of a whole genome, and localized areas of hypermethylation at certain specific sites, e.g. the CpG islands and within the gene promoter region (Feinberg and Vogelstein 1983; Baylin et al., 1986). Global hypomethylation will be introduced only briefly, and more discussions will be focused on hypermethylation of CpG islands of TSGs.

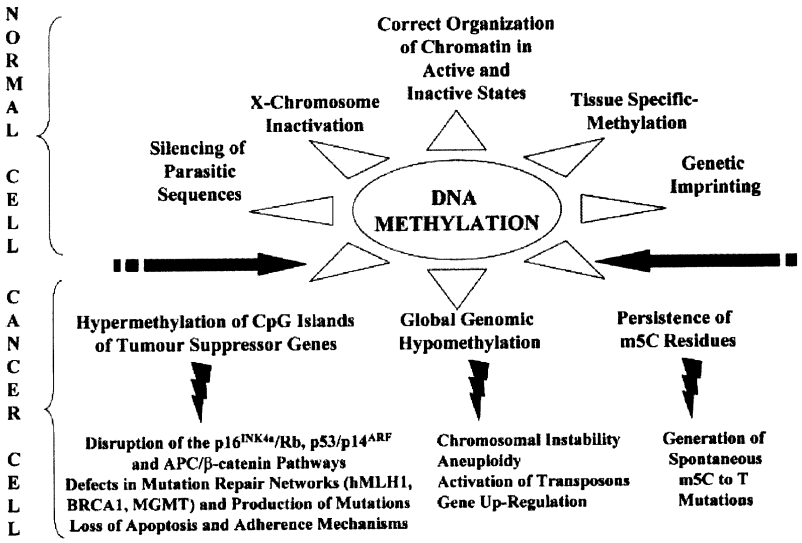

Fig. 4 DNA methylation at the center of the normal and malignant behavior of the cell. (Esteller and Herman 2002)

The amount of 5'-methylcytosine in genomic DNA can be measured directly by HPLC (Gama-Sosa et al., 1983a) or indirectly as an inverse value of the capacity of a DNA sample to accept tritiated methyl groups from a universal methyl donor s-adenosylmethionine (Kim et al., 1994).

These distinct methods have shown that simultaneously with hypermethylation of CpG islands, the genome of the cancer cell undergoes a dramatic global hypomethylation. The malignant cell can have 20-60% less genomic 5MC (5'-methylcytosine) than its normal countpart (Lapeyre and Becker 1979; Lu et al., 1983). The loss of methyl groups occurs mainly in the 'body' (coding region and introns) of genes. The extent of genome wide hypomethylation in tumors parallels closely the degree of malignancy, though this is tumor type dependent. It was reported that in breast, ovarian, cervical, and brain tumors, hypomethylation increases progressively with increasing malignancy grade (Gama-Sosa et al., 1983b; Kim et al., 1994; Qu et al., 1999; Narayan et al., 1998). Thus, hypomethylation may serve as a biological marker with prognostic value. There are several mechanisms for global DNA hypomethylation contribution to carcinogenesis. These mechanisms include chromosomal instability, reactivation of transposable elements, and loss of imprinting. Undermethylation of DNA might favor mitotic recombination leading to loss of heterozygosity, as well as promoting karyotypically detectable rearrangements. Additionally, extensive demethylation in centromeric sequences is common in human tumors and may play a role in aneuploidy. Hypomethylation of the malignant cell DNA can also reactivate intragenomic parasitic DNA: loss of methylation has been observed in L1 (long interspersed nuclear elements, LINES) and Alu (recombinogenic sequences) repeats in cancer cells (Yoder et al., 1997; Thayer et al., 1993; Alves et al., 1996). These and other previously silent

transposons may now be transcribed and may even move to other genomic regions where they can disrupt normal cellular genes. Finally, the loss of methyl groups can affect imprinted genes. The best-studied case affects the H19/IGF-2 locus in chromosome 11p15 (Feinberg 1999), where the disturbance of methylation may cause overexpression of an anti-apoptotic growth factor (*IGF-2*) and loss of a transformation-suppressing RNA (*H19*) in certain childhood tumors.

Holliday and Pugh (1975) proposed that if hypomethylation leads to inappropriate activation of genes important for neoplastic growth, then hypomethylation could provide a selective advantage for the tumor cells. Such cells could then clonally evolve and would appear as a prominent population in the tumor. Hypomethylation within the body of a number of genes has been found in primary cancers (Feinberg and Vogelstein 1983), including known oncogenes such as CMYC and HRAS (Del Senno et al., 1989; Vachtenheim et al., 1994). Although reduced levels of methylation of genes including CMYC in human tumors have been reported, it is not been possible to show convincingly that this is indeed responsible for increased levels of gene expression rather than merely a secondary characteristic observed in cancer cells (Sharrard et al., 1992).

Several lines of evidence suggest that DNA hypomethylation and chromosome instability may result from insufficient dietary folate. Folate provides carbon units for a number of biochemical processes, including production of SAM, a universal methyl donor that also supplies the methyl group to cytosine in DNA as mentioned above. There is strong epidemiological evidence that sufficient dietary folate is important to reduce the risk of certain cancer (Kim 1999). At present, reduction of folate supply is the only known cellular mechanism leading to genome hypomethylation in cancer.

CpG sites have been shown to act as hotspots for mutations and have been estimated to contribute to 30% of all point mutations in the germline (Cooper and Youssoufian 1988). In addition, CpG sites in the coding regions of TSGs are strong hotspots for acquired somatic mutations leading to cancer (Rideout et al., 1990; Greenblatt et al., 1994). 5-Methylcytosine can undergo spontaneous deamination to form thymine at a rate much higher than the deamination of cytosine to uracil (Shen et al., 1994). As thymine is a normal component of human DNA, this mutation may not be correctly recognized by the DNA repair mechanism. If the deamination of 5-methylcytosine is unrepaired, it will result in a C to T transition mutation. This phenomenon was used to explain the high incidence of CpG to TpG transition mutations observed in the *p53* tumor suppressor gene (Rideout et al., 1990). However, this phenomenon may be more complex and may involve additional events (Schmutte and Jones 1998). For example, the carcinogen benzopyrene, preferentially formed adducts at the methylated CpG sites of the *p53* gene and these sites are hot spots for mutation (Denissenko et al., 1997). The relative importance of spontaneous deamination of 5-methylcytosine and carcinogen adduct formation at CpG sites in producing mutations remains to be clarified.

CpG islands surround the transcription start regions of almost half of the genes in the human genome and are normally unmethylated. Hypermethylation of CpG islands in cancer are associated with transcriptional silencing of the genes in which this change occurs. The mechanism of transcriptional silencing by methylation of CpG islands will be discussed in 2.2 section. Knudson proposed that both alleles of TSG have to be inactivated in tumors. Initially, point mutations and chromosomal deletions were considered to be the major events involved in the inactivation of TSGs. The discovery that many TSGs can also be inactivated by aberrant methylation of the CpG islands in their promoter region clearly indicates that epigenetic events also play a very important role in tumorigenesis (Jones and Laird 1999). For some genes, methylation provides a similar selective advantage as mutations or deletions; hypermethylation should be considered as one of the inactivating mechanism in Knudson's model of TSGs (Baylin et al., 1998). A model for the biallelic inactivation of TSGs by aberrant DNA methylation alone or in combination with mutations or deletions is shown in

Fig. 5. This model is supported by the reports that some cancer-related genes were found to be inactivated by biallelic methylation of CpG sequences (Batova et al., 1997; Veigl et al., 1998).

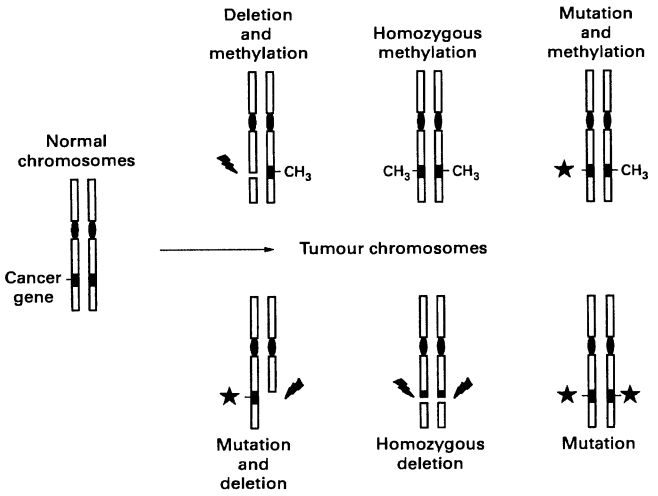

Fig. 5 The combination of genetic and epigenetic events in cancer provides a mechanism for complete inactivation of both allelic locations. (Costello and Christoph 2001)

The candidate gene approach tests for aberrant methylation known cancer genes, particularly specific alleles that do not harbor genetic alterations. This lucrative approach has uncovered methylation related gene silencing that can account for most types of malignant behavior exhibited by human cancer cells (Table 6). Genes involved in cellular pathways ($p16^{INK4a}$/Rb/cdk4, p53/p14ARF/MDM2, APC/β-catenin/E-cadherin), cell cycle regulation, DNA repair, drug resistance and detoxification, differentiation, apoptosis, angiogenesis, metastasis, and invasion are inappropriately silenced by methylation. Similar gene silencing events are recapitulated in chemically and genetically induced mouse models of human cancer (Belinsky 1998; Akama et al., 1997). Thus, there is considerable evidence that CpG island methylation contributes directly to malignancy (Robertson and Wolffe 2000; Baylin et al., 1998; Jones 1999).

Table 6. Aberrantly methylated genes in cancer

Fuction	Genes	*References (examples)*
Apoptosis	Death associated protein kinase (*DAP kinase*, 9q34)	Katzenellenbogen et al. 1999; Aggerholm and Hokland 2000
	Caspase 8 (*CASP8*, 2q33-34)	Teitz et al. 2000
	Target of methylation induced silencing (*TMS1*, 16p11.2-12.1)	Conway et al. 2000; McConnell and Vertino 2000
Agiogenesis	Thrombospondin-1 (*THBS1*, 15q15)	Li et al. 1999
Cell cycle	Retinoblastoma (*RB*, 13q14)	Stirzaker et al. 1997; Greger et al. 1989
	p14ARF (9p21)	Esteller et al. 2000; Zhang et al. 2000
	Cyclin dependent kinase 2A (*CDKN2A*, 9p21)	Merlo et al. 1995; Costello et al. 1996
	Cyclin dependent kinase 2B (*CDKN2B*, 9p21)	Dodge et al. 1998; Herman et al. 1996
	p27/*KIP1* (12p13)	Worm et al. 2000
	p73 (*TP73*, 1p36)	Corn et al. 1999
	14-3-3σ (stratifin, *SFN*, 1p)	Iwata et al. 2000; Suzuki et al. 2000

Differentiation	Myogenic differentiation antigen-1 (*MYOD*, 11p15.4)	Jones et al. 1990
	Paired box gene 6 (*PAX6*, 11p13)	Salem et al. 2000
	Retinoic acid receptor (*RARβ2*, 3p24)	Virmani et al. 2000; Arapshian et al.
	Wilms tumor (*WT1*, 11p13)	Malik et al. 2000
DNA repair	*hMLH1* (3p23-p21.3)	Deng et al. 1999; Simpkins et al. 1999
	O-6-methylguanine-DNA methyltransferase (*MGMT*, 10q26)	Harris et al. 1994; Qian et al. 1995
Metastasis/invasion	E-cadherin (*CDH1*, 16q22.1)	Saito et al. 1998; Yoshiura et al. 1995
	Tissue inhibitor of metalloproteinase 3 (*TIMP-3*)	Bachman et al. 1999
	mts-1	Tulchinsky et al. 1995
	Maspin (protease inhibitor 5, *PI5*, 18q21.3)	Domann et al. 2000
Drug resistance/	Glutathione S-transferase π (*GSTP1*, 11q13)	Lee et al. 1994; Esteller et al. 1998
Detoxification	Multi-drug resistance 1 (*MDR1*, 7q21.1)	Kantharidis et al. 1997
Signal transduction	Adenomatous polyposis of colon (*APC*, 5q21-22)	Tsuchiya et al. 2000
	PTEN (10q23.3)	Salvesen et al. 2001; Cairns et al. 1997
	Androgen receptor (*AR*, Xq11-12)	Jarrard et al. 1998
	Oestrogen receptor (*ESR1*, 6q25.1)	Li et al. 1998; Ottaviano et al. 1994
	Ras association domain family member 1 (*RASSF1A*, 3p21.3)	Dammann et al. 2000
	Serine/threonine protein kinase 11 (*STK11* or *LKB1*, 19p13.3)	Esteller et al. 2000
Transcription/	Von Hippel-Lindau syndrome (*VHL*, 3p26-25)	Kuzmin et al. 1999; Herman et al. 1994
transcription	Hypermethylated in cancer (*HIC-1*, 17p13.3)	Wales et al. 1995; Fujii et al. 1998
	Breast cancer, type 1 (*BRCA1*, 17q21)	Rice et al. 1998; Rice et al. 2000
Other	CD44 antigen (*CD44*, 11pter-p13)	Lou et al. 1999
	Cyclo-oxygenase 2 (*COX2*, 1q25.2-25.3)	Toyota et al. 2000
	Calcium channel, voltage dependent, T type, alpha-1G subunit (*CACNA1G*, 17q22)	Toyota et al. 1999
	Calcitonin (*CALCA*, 11p15.2-15.1)	Nelkin et al. 1991; Baylin et al. 1986
	Fragile histidine triad gene (*FHIT*, 3p14.2)	Tanaka et al. 1998
	Telomerase reverse transcriptase (*TERT*, 5p15.33)	Dessain et al. 2000; Devereux et al. 1999
	Transmemberane protein containing epidermal	Liang et al. 2000
	Chondroitin sulphate proteoglycan 2 (*CSPG2*, 5q12-14)	Adany et al. 1990
	Mammary-derived growth inhibitor	Huynh et al. 1996
	MAGE-1	Weber et al. 1994; Coral et al. 1999

Treatment of cells with 5-azacytidine and its deoxy version 5-aza-2'-deoxycytidine results in a progressive loss of DNA methylation with each round of cell division (Juttermann et al., 1994). Such treatment revealed that a large number of genes could be reactivated; however, there appeared to be some specificity to this effect. It was later realized that cell lines contained altered patterns of 5-methylcytosine distributions relative to primary cells. This is presumably caused by adaptation to tissue culture conditions, and exactly this 'abnormal' or culture-associated methylation was preferentially removed by 5-azacytidine treatment (Jones 1985; Antequera et al., 1990). While both compounds are relatively poor chemotherapeutic agents they have been extremely valuable in studying the role of DNA methylation in gene expression.

2.2 Epigenetic mechanism of gene inactivation

Eukaryotic chromosomes are made up of active regions, in which chromatin structure is 'open' and accessible to DNA binding proteins, and silent regions, where packed chromatin renders the DNA inaccessible. The chromatin in mammalian cells consists of a series of nucleosomes arranged in a compact configuration. The nucleosome consists of 146 bp DNA wrapped around a protein octamer containing two molecules each of histone H2A, H2B, H3 and H4. Histone acetylation precedes transcription and results in decondensation of the chromatin to permit binding of transcription factors to DNA. Histone acetyltransferase and histone deacetylase play an important role in this process. Histone acetylation is associated with active transcription (Wolffe 1996). A strong correlation between DNA hypermethylation, transcriptional silence and tightly compacted chromatin has been established in many different systems (Kass et al., 1997).

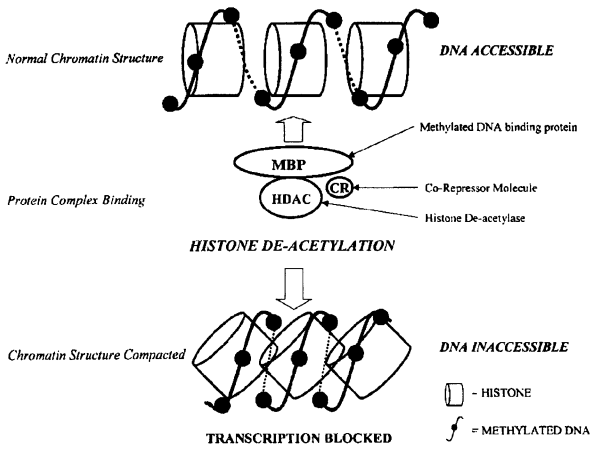

Fig. 6 Transcriptional repression resulting from alteration of chromatin structure. Methylated DNA binds to a protein complex consisting of a methyl binding protein (MBP), which has a methyl-binding domain and a transcriptional repression domain, a corepressor molecules (CR), and a histone deacetylases (HADC). After binding of this complex to the methylated DNA, the histones around which the DNA is wrapped become deacetylated, resulting in a compression and compaction of the chromatin structure. This makes the DNA inaccessible, and thus functional transcription is no longer possible.

The transcriptional silencing process was initially thought to be due to the physical effect of the methyl group protruding from the DNA and interfering with the mechanics of transcription (Razin and Riggs 1980). At present the inhibitory mechanism is thought to occur through the binding of specific proteins to the methylated DNA sequences. These proteins, such as MeCP2, belong to a family of proteins that contain a methyl-CpG binding domain (MBD) that recognizes and binds preferentially to methylated CpG groups irrespective of gene sequence. The protein also contains a transcriptional repression domain (TRD), which forms a complex with a variety of corepressor molecules (e.g., mSin3A) and histone deacetylase protein (e.g., HDAC1, HDAC2). When this complex binds to methylated DNA, the histone proteins around the DNA strands are wrapped to form chromatin and become deacetylated. This causes changes in the chromatin structure, making it more condensed and DNA less accessible, preventing active transcription. Importantly, the histone deacetylase inhibitor, trichostatin (TSA), can activate the transcription of certain genes (Cameron et al., 1999). This provided the first direct connection between two transcriptional silencing pathways. It is known now that methyl-binding proteins (MBD1, MBD2, MBD3, MBD4 and MeCP2) are responsible for recruiting histone deacetylases (HDACs) and other chromatin factors. It appears that methylated inactive genes contain underacetylated histone whereas unmethylated active genes

preferentially associate with highly acetylated histones. The interaction of histone acetylation and DNA methylation resulting in transcriptional repression by alteration of chromatin structure is shown in Fig. 6.

A variety of methods are used to evaluate the methylation status of genomes and genes. Methods for detection of methylation patterns and genome-wide changes include HPLC (Kuo et al., 1980), methyl acceptance assay (Baladhi and Wagner 1993), restriction landmark genomic scanning (RLGS), methylation-sensitive-representational difference analysis (MS-RDA) (Ushijima et al., 1997), methylation-sensitive arbitrarily primed PCR (MS-AP-PCR) (Liang et al., 1998), methyl-CpG binding domain column/segregation of partly melted molecule (MBD/SPM) and powerful specific microarrays such as CGI (CpG islands) microarrays (Huang et al., 1999), our *Not*I microarrays which will be introduced in 3.3. RLGS is a highly reproducible two-dimensional gel electrophoresis of genomic DNA that allows the assessment of over 2000 loci simultaneously (Hatada et al., 1991; Okazaki et al., 1995). This technique is based on the digestion of the genome with a methylation-sensitive restriction enzyme and has been used for various purposes including genetic mapping, identification of novel imprinted genes, genomic amplifications, regions of hypomethylation, regions of hypermethylation, candidate TSGs, and measuring the degree of CpG island hypermethylation in cancer (Costello et al., 2000; Hayashizaki et al., 1993; Miwa et al., 1995; Okazaki et al., 1996; Okuizumi et al., 1995; Plass et al., 1996; Shibata et al., 1995; Smiraglia and Plass 2002). MS-RDA is also based on the digestion of genomic DNA with a methylation-sensitive restriction enzyme and PCR amplification of the entire digestion product with a universal adaptor. The difference in the methylation status is converted into the difference in the presence or absence of particular DNA fragments, which can be easily identified by a genomic subtraction technique, RDA. MS-AP-PCR method is based on the digestion of genome with a methylation sensitive restriction enzyme and AP-PCR.

Methods for detection of methylation changes in single genes include Southern hybridization (Lindsay and Bird 1987), restriction enzyme PCR (Kane et al., 1997), Ms-PCR (Herman et al., 1996), Ms-SSCP (Maekawa et al., 1999), methylation-sensitive single nucleotide primer extension (MS-SNuPE) (Gonzalgo and Jones 1997), combined bisulphite restriction analysis (COBRA) (Xiong and Laird 1997) and bisulfite genomic sequencing (Clark et al., 1994; Frommer et al., 1992). In Southern hybridization method, the genomic DNA is cleaved with methylation-sensitive and insensitive endonucleases specific for the same sequence, such as *Hpa*II and *Msp*I, and then followed with Southern blot. In restriction enzyme PCR method, digested DNA is then amplified using primers flanking the target region. One of the limitations of Southern blot is the large amount of DNA required (5-10 μg per sample) that is not always possible especially in the case of tumor samples. Another problem (also valid for restriction enzyme PCR) is incomplete enzymatic digestion that can lead to results that are difficult to interpret. Other methods are based on bisulfite treatment that results in deamination of all the cytosines to uracil, which replicates as thymine during PCR amplification. During this treatment the 5-MC (5'-methylcytosine) is unchanged (Hayatsu et al., 1970), so the sequence will be different between methylated and unmethylated DNA after bisulfite treatment. Bisulfite DNA sequencing provides an excellent tool for the detection of methylation. In Ms-PCR, primers are designed specifically to anneal to sequences that contain either methylated (C) or modified (T) sites. Careful selection of primers is very important because it is possible to obtain false-positives with both methylated and unmethylated primer pairs, making it difficult to interpret the results. Incomplete bisulfite modification of genomic DNA can also yield false results. It has been reported that C adjacent to methylated CpG sites can be resistant to bisulfite treatment (Harrison et al., 1998). The protocols for COBRA and bisulfite sequencing refer to **paper IX** in this thesis.

2.3 *Not*I linking and jumping clones

*Not*I linking and jumping libraries offer a powerful tool for long-range genome mapping and isolation of genes. *Not*I clones possess a number of properties that make them as useful markers in different strategies for genome mapping:

They contain CpG islands, which have been shown to be conserved in the genome and can be used for comparative genome mapping in different species (Hino et al., 1993). Over 20% of CpG islands containing genes possess *Not*I site(s) in their sequences, and 65% have *Xma*III site(s). Partially sequenced *Not*I linking clones serve as markers for expressed sequences on specific human chromosomes, tagging approximately 10 to 20% of all transcribed sequences. Approximately 90% of the *Not*I linking clones tag expressed sequences in humans, as determined by Northern blot hybridization (Allikmets et al., 1994). *Not*I linking clones contain on average 65% C+G and are mainly located at the R-bands of chromosomes, marking CpG islands and housekeeping genes. Analysis of nucleosome formation potential (NFP) with 1 kb flanking sequences of 142 chromosome 3-specific *Not*I clones demonstrated that regions flanking *Not*I sites are less likely to form nucleosomes efficiently as their NFP values are below -1 and therefore resemble promoter regions in this feature.

They contain recognition sites for rare-cutting enzymes useful for physical mapping and they are likely to contain polymorphic sequences (if the insert is bigger than 6 kb; Bechman and Weber 1992). A *Not*I restriction map can be used as a framework to join the information obtained from other mapping methods, such as cosmid and YAC clones, radiation hybrid maps, in situ hybridization, etc. *Not*I linking clones can be used for joining physical and genetic mapping efforts. They also can be useful for comparative studies.

It has been demonstrated that short stretches of DNA surrounding *Not*I sites can be very useful not only for production of STSs but also for localization of known and identification of new genes.

Several different methods have been proposed for construction *Not*I linking and jumping libraries (Arenstorf et al., 1991; Hattori et al., 1993; Wallace et al., 1989; Ito and Sakaki 1988; Saito et al., 1991). These techniques for construction *Not*I linking and jumping libraries have several drawbacks. A new procedure for construction linking and jumping libraries was described in our group (Fig. 7). The essential features of this procedure include:

Special λSK diphasmid vectors were generated (Zabarovsky et al., 1993). Then, a partial filling-in reaction is used to eliminate cloning of artificial clones and to obviate the need for a selectable marker. No selection marker (like supF) is used.

As a result of the new procedure, jumping libraries that are nearly representative can be obtained using only 1-2 μg of vectors arms while the volumes of ligation and packaging reactions are significantly reduced, in comparison to other protocols, which require large amounts of vector arms (40-150 μg) and specially prepared packaging extracts (Poustka and Lehrach 1998; Collins 1988).

In 1993, representative *Not*I linking and jumping libraries specific for human chromosome 3 and for the total human genome were constructed in our group (Zabarovsky et al., 1993). Chromosome 3 *Not*I linking libraries were constructed to facilitate physical mapping and identification of tumor suppressor genes. Since the same scheme (Fig. 7) is employed to construct *Not*I linking and jumping libraries, the two types of libraries contain the same DNA fragments but are differently arranged.

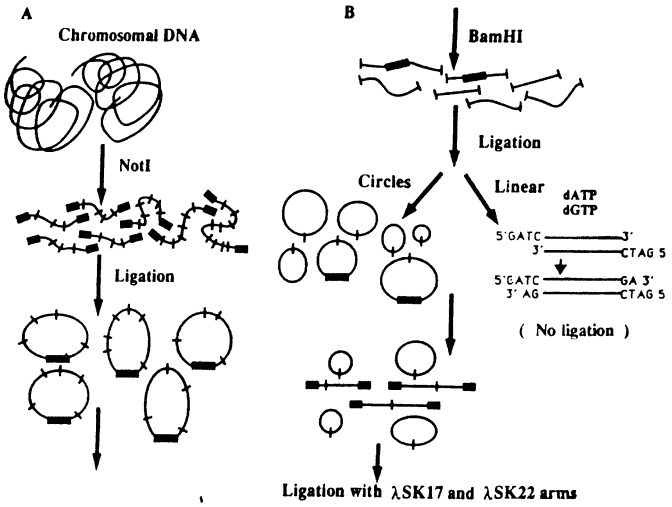

Fig. 7 Flow diagram of the cloning procedure. Black bars, NotI sites; vertical slashes, BamHI sites. A, B construction of the jumping library. B construction of the linking library. In this case, digestion of the genomic DNA with BamHI is the first step.

*Not*I linking clones contain pairs of sequences flanking a single *Not*I restriction site, while *Not*I jumping clones contain DNA sequences spanning between neighboring *Not*I recognition site (Fig. 8). Information about 500-600 bp surrounding each *Not*I sites can easily be obtained by sequencing with direct and reverse primers in the vector of both type of libraries. Subsequently the linear order of the *Not*I clones (*Not*I contig) can be established using this shotgun sequencing strategy. Finally, *Not*I restriction long-range maps (650 kb resolution) can be confirmed by pulsed field gel electrophoresis with ordered *Not*I jumping/linking clones from large defined regions of chromosome. The *Not*I map can be used for verification and joining other maps as it can be easily checked.

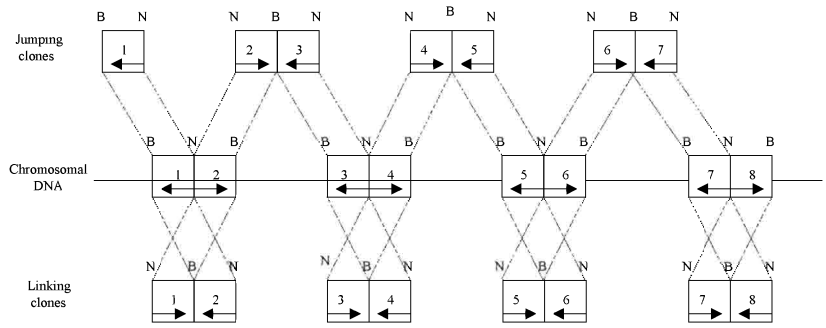

Fig. 8 The general scheme for arrangement of NotI flanking sequences in genomic DNA, linking and jumping clones. N, NotI; B, BamHI. Arrows indicate the sequences flanking NotI sites.

More than 1000 *Not*I linking clones isolated from human chromosome 3-specific libraries were partially sequenced. Of these clones, 162 were unique chromosome 3-specific clones (Kashuba et al., 1999). At the beginning, *Not*I linking clones were rapidly mapped to different regions using differential hybridization to somatic hybrids and Alu-PCR (Zabarovsky et al., 1994). Then the clones were precisely mapped using a combination of fluorescence in situ hybridization (FISH) and hybridization to somatic cell or radiation hybrids. In many cases, chromosome jumping was successful used to resolve ambiguous mapping. *Not*I restriction maps (8 Mb) for interesting regions were constructed. A search of the EMBL nucleotide

database with these sequences revealed homologies (90-100%) to more than 150 different genes or expressed sequence tags (ESTs). Many of these homologies were used to map new genes to chromosome 3.

High-density grids with 50,000 *Not*I linking clones derived from six total human *Not*I linking and three *Not*I jumping libraries were constructed. Altogether, these libraries contained nearly 100 times the total estimated number of *Not*I sites in the human genome. After sequencing 22,551 unique human *Not*I flanking sequences (16.2 Mb) were generated. Analysis of sequences demonstrated that about 50% of these clones displayed significant similarity to protein and cDNA sequences. Among these unique sequences, 10,993 (48.7%) were novel sequences, not present in the EMBL or EST (Expressed Sequence Tag) databases (similarity ≥ 90% over 50bp).

Analysis of CG content for the first 350 bp is shown in Table 7. Comparing these data with Lander et al. (2001), two main features are apparent: the fraction of sequences with > 80% CG content is seven times higher in the *Not*I collection (142 versus 22 sequences). Another striking finding is that even *Not*I flanking sequences with a CG content < 50% have a very high ratio of observed versus expected frequency of CG dinucleotides (0.71). This suggests that essentially all *Not*I flanking sequences generated in the study are located in CpG islands and, therefore, the computational method misses at least 8.7% of CpG islands associated with *Not*I sites (Kutsenko et al., 2002).

Table 7. GC content of *Not*I flanks

GC content (%) of *Not*I flanks	Number of *Not*I flanks	Percentage of *Not*I flanks	Ration between observed and expected CG pairs
Total	22,551	100	0.77
>80	142	0.6	0.96
70-80	4005	17.8	0.87
60-70	9629	42.7	0.78
50-60	6813	30.2	0.70
40-50	1751	7.8	0.70
<40	211	0.9	0.75

*Not*I sequences were used to verify the assembled human genome sequences. The draft human genome sequence (Lander et al., 2001; Venter et al., 2001) contains a significant portion of the *Not*I sequence collection (Fig. 9). With stringent criteria, 55.7% of the *Not*I flanking sequences were present in a public assembly of the human genome (December 2001, identity ≥ 90%). Inclusion of the Celera sequences identified an additional 1.5% of *Not*I flanks. All public databases (EMBL + HTGS + EST) matched 89.2% of the *Not*I flanking sequences (July 2002) and search stringency is important here, this number increased to 91.1% at identity ≥ 78% and went down to 84.1% at identity ≥ 95%. The public draft sequence contained 19,552 *Not*I sites (39,104 *Not*I flanking sequences).

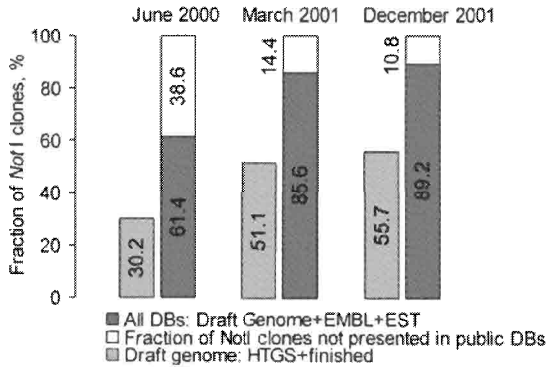

Fig. 9 Fraction of the NotI flanking clones present in EMBL, EST and HTGS databases (similarity ≥90% over 50 bp).

The data showed that the draft human genome sequence has a strong bias against *Not*I flanking sequences, as a significant number of the human *Not*I sequences were not detected. Several explanations can be offered to account for the low representative of *Not*I flanking sequences in the draft human genome sequences. Our previous studies (Kashuba et al., 1995, 1999; Wei et al., 1996) demonstrated that large-insert vectors containing DNA fragments from regions most frequently deleted in tumors were unstable and sensitive to deletions and rearrangements. Thus one explanation is that the cloning of some *Not*I site-containing regions may be selected against in experiments with large-insert cloning vectors. An alternative explanation is that some *Not*I sites were incorrectly fused in the assembly process (some different *Not*I flanking sequences can have 100% identity over long DNA stretches; Kashuba et al., 2002). Furthermore, our experience demonstrates that sometimes it is very difficult to read *Not*I flanking sequences because of the extremely high CG content. During human genome assembly such sequences would be eliminated as possessing low quality data. Thus *Not*I clones will be helpful as probes to close existing gaps in the draft human genome sequence and in estimating the completeness of the human genome sequence due to the independent approach used.

*Not*I sequences can be applied as a tool for gene identification since *Not*I sites are located in the 5' end of genes and the first methionine found in *Not*I sequences corresponds in most case to the first methionine of cDNAs (Zabarovsky et al., 2000). We have demonstrated that nearly all *Not*I clones contained genes (Kashuba et al., 1999; Allikmets et al., 1994). *Not*I flanking sequence comparisons with full-length human cDNA coding sequences from Unigene and expressed sequences from EST database are shown in Fig. 10A and 10B. The number of sequences matching 5' and 3' end of ESTs is higher than the total number of *Not*I sequences that are likely to be expressed (for example, 11.3% + 33.9% = 45.2% > 37.1% for 90% similarity, see Fig. 10B). This is because the same *Not*I sequence can match 5' as well as 3' end of ESTs. These data further support a previous suggestion that many of the matching '3' end of EST' sequences are actually situated in the 5' end of genes that contain *Not*I sits in the first exons (Kashuba et al., 1999; Zabarovsky et al., 2000). *Not*I sequences were compared to ESTs from other organisms; several hundred additional ESTs (661 for identity ≥ 78%) were similar to *Not*I flanking sequences. These *Not*I clones most likely represented human genes evolutionarily related to the genes from other organisms.

One important application of *Not*I clones in cancer research is to isolate cancer associated genes in tumors. The complexity of the *Not*I linking libraries is at least 100 times lower than the complexity of the whole human genome; it is approximately equal in complexity to yeast genome. Genomic subtraction using linking libraries instead of genomic DNA allowed us to clone deleted *Not*I surrounding sequences due to the decreased complexity of the system.

*Not*I-CODE was developed just for this purpose (refer to **paper V**). *Not*I microarrays based on the human *Not*I linking clones were used to detect copy number and methylation changes in tumors (refer to **paper V**) and it will be discussed in 3.3.

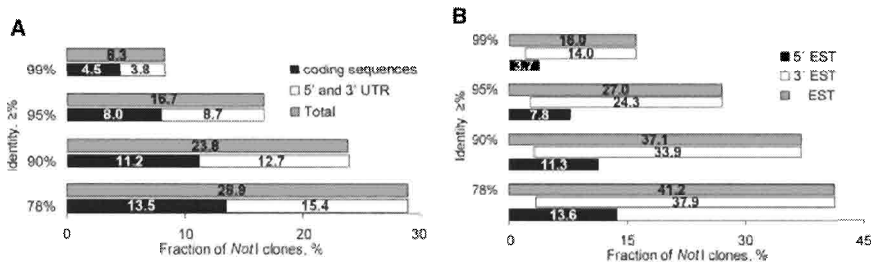

Fig. 10 Similarity between NotI flanking sequences and human Unigene mRNA sequences (A), and NotI flanking sequences and EST sequences (B).

3. Microarrays

DNA microarrays are ordered, high-density arranged nucleic acid spots (Lockhart and Winzeler 2000). A 'target' is tethered to an immobile surface (glass, silicon, nylon and nitrocellulose membranes, gels and beads have been used) and then exposed to the 'probe', consisting of the free nucleic acid sample that is being analyzed, and probes are radioactive or fluorescent labelling. In practice, the amount of target material per spot ranges from approximately 50 ng (for Nylon macroarrays) to 1 ng or less for glass microarrays (Schena et al., 1995); this corresponds roughly to 50 X 10^9 and 10^9 individual target molecules, respectively (for 1 kb double stranded DNA targets). Oligonucleotide chips are said to contain 10^6 to10^7 identical oligonucleotides (Lockhart et al., 1996). The methods of microarray fabrication of targets include various inkjet and microjet deposition or spotting technologies and processes using split needle (Telecom), needle-ring (GMS/Affymetrix) or ink-jet needle (Packard), in situ or on-chip photolithographic oligonucleotide synthesis processes, and electronic DNA probe addressing processes. The array hybridization conditions are very different from the familiar Northern or Southern saturation hybaidizations, in which the probe is a single molecular species, present at high concentration and in huge excess over its target (typically 1,000 to 1 in Southern). In contrast to it, in microarray hybridizations only 0.1%-1% of the target molecules have actually bound a labelled probe molecule. Table 8 shows the comparison of sensitivity for different arrays. That is why microarray experiments have raised a wide range of computational requirements, including image processing (Chen et al., 1997), instrumentation and robotics (Cheung et al., 1999), database design (Ermolaeva et al., 1998; Aach et al., 2000), data storage and retrieval (Ringwald et al., 2000), microarray design based on available ESTs (Miller et al., 1997), and data analysis (Bassett et al., 1999). Furthermore, microarray data need to be interpreted in the context of other biological knowledge, involving various types of 'post-genomic' informatics (Kanehisa 2000), including gene networks (Somogyi and Sniegoski 1996), gene pathways (Kanehisa and Goto 2000), and gene ontologies (Ashburner et al., 2000).

Table 8. Comparison of Sensitivity for Macroarrays, Microarrays, and Oligonucleotide Chips.

Parameter	Nylon macroarrays	Nylon microarrays (Enzymatic det.)	Nylon microarrays (Radioactive det.)	Glass microarrays	Oligonucleotide chips
Targets	cDNA clones (bacterial colonies or PCR products)	cDNA clones (PCR products)	cDNA clones (PCR products)	cDNA clones (PCR products)	20-mer oligos synthesized in situ
Support and format	50–2,000 spots on a 8 × 12 cm^2 Nylon membrane	9,600 spots on a 2.7 × 1.8 cm^2 Nylon membrane	200 spots on a 5 × 4 mm^2 Nylon membrane	50–6,400 spots on a 1.8 × 1.8 cm^2 glass slide	64,000 oligonucleotides on a 1.28 × 1.28 cm^2 glass chip
Sample for complex probe	25 µg of total RNA (approx. 0.5 µg mRNA)	1 µg of mRNA	0.1 µg of total RNA, i.e. 2 ng of mRNA	2 µg of mRNA	10 µg of mRNA
Labelling	^{32}P or ^{33}P	Enzymatic	^{33}P	Fluorescence	Fluorescence
Hybridisation volume	10–40 ml	10 µl	100 µl	2–10 µl	200 µl
Temperature and duration of hybridisation	65°C during 20 hours	68°C during 12 hours	65°C during 48 hours	65°C during 18 hours	40°C during 15 hours
Image acquisition	Imaging plate device	Flatbed scanner or digital camera	Imaging plate device	Confocal scanning	Confocal scanning
Detection limit (relative to mRNA abundance)	1/20,000	Approx. 1/20,000	1/10,000	1/100,000	1/300,000
Sensitivity* (minimum number of molecules in sample for detection)	25 × 10^6 molecules	60 × 10^6 molecules	0.2 × 10^6 molecules	20 × 10^6 molecules	30 × 10^6 molecules

* Sensitivity is expressed as the minimum number of molecules of a given sequence species that must be present in the starting sample (RNA) to measure a signal after array hybridization. Calculations assume that mRNA represents 2% of total RNA for a typical mammalian cell and that the average size of an mRNA molecule is 1.7 kb; 1 pg of mRNA then corresponds approximately to 106 molecules.

Recently, protein microarray and tissue microarray (TMA) technologies were developed. Protein microarray-based assays have huge potential for diagnostic and proteomic applications (Bussow et al., 2001). Tissue microarray (TMA) technology is designed to efficiently test the clinical relevance of candidate genes. TMAs are histologic slides containing samples from hundreds of individual tumor specimens. The TMA slides can be used for large-scale, massively parallel in situ analysis of DNA by fluorescence in situ hybridization (FISH), RNA by mRNA in situ hybridization and protein by immunohistochemistry. In DNA microarrays, one tumor is analyzed for the expression of thousands of genes or DNA copy number changes and TMA are used for the analysis of one gene in hundreds of different tumors. (Bubendorf 2001).

3.1 cDNA/oligo microarrays with gene expression profiling

The complete human genome sequence does not explain how a cell or organism may respond to normal and abnormal biological processes. The main challenge in the postgenomic era is to develop systematic approaches for identifying the biological function of all genes. By examine the mRNA content of a cell; researchers can determine which genes are being activated in response to a stimulus. cDNA/oligo microarray technology for gene expression analysis represents a major advance in this regard.

Traditional methods in molecular biology generally work on a "one gene: one experiment" basis, which means that the throughput is very limited and the "whole picture" of gene function is hard to obtain. To study each of the 60,000 to 80,000 genes individually under each biological circumstance is not practical. Recently, novel high-throughput techniques have emerged, such as differential display (Liang and Pardee 1992), serial analysis of gene expression (SAGE) (Velculescu et al., 1995), and microarrays (also known as gene or cDNA chips) (Schena et al., 1995). Of these high-throughput techniques, cDNA microarrays which were developed at Stanford University (Schena et al., 1995; Shalon et al., 1996; DeRisi et al., 1996; Schena et al., 1996; Spellman et al., 1998; Iyer et al., 1999) are much more efficient.

They are capable of profiling gene expression patterns of tens of thousands of genes in the entire genome in a single experiment. cDNA microarray schema is shown in Fig. 11.

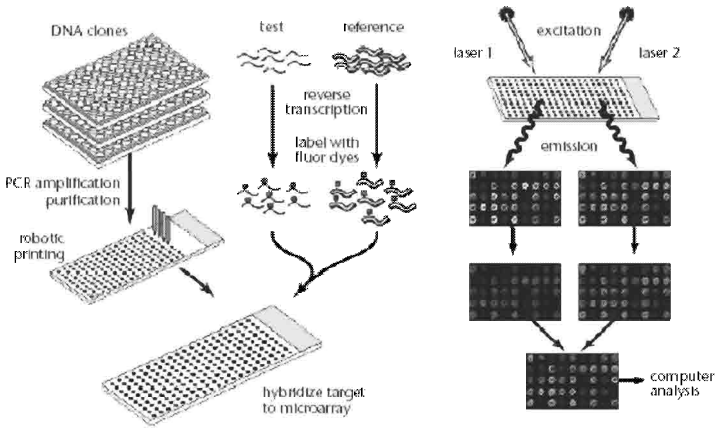

Fig. 11 cDNA microarray schema. Templates for genes of interest are obtained and amplified by PCR. Following purification and quality control, aliquots (~5 nl) are printed on coated glass microscope slides using a computer-controlled, high-speed robot.

In an array experiment, many gene-specific polynucleotides derived from the 3' end of RNA transcripts by PCR are individually spotted on a single matrix (spot sizes range between 80 and 150 μm in diameter, and arrays that contain up to 80,000 spots can be obtained). Spotting is carried out by a robot (Omnigrid from GeneMachines, 417 Arrayer from Affymetrix), which deposits a nanoliter of PCR product onto the matrix in serial order. Nylon filter arrays largely have been replaced by glass-based arrays, typically microscope slides, which have the advantage of two-color fluorescent labelling with low inherent background fluorescence. DNA adherence to the slide is enhanced by treatment with polylysine or other cross-linking chemical coating. Spotted DNA is cross linked to the matrix by ultraviolet irradiation and denatured by exposure to either heat or alkali. This matrix is then simultaneously probed with fluorescently tagged cDNA representations of total RNA pools from test and reference samples. To compare the relative abundance of each of these gene sequences in two pools, the two samples are first labelled using different fluorescent dyes (fluorophores Cyanin 3-green and Cyanin 5-red typically are used, for example, test labelled with Cy3 and reference labelled with Cy5). There are different approaches for probe labelling. Direct enzymatic incorporation of Cy3/Cy5 labelled nucleotides without amplification (SuperscriptII/LifeTech) affects the efficiency of incorporation because of dye molecular structure and Cy5 is generally not as efficient as Cy3, which generates artificial signal bias. Amersham Pharmacia Biotech recently developed a new reverse transcriptase, CyScript™, to enhance the incorporation of CyDye-labeled nucleotides with even efficiency for Cy3 and Cy5 (CyScript™ Kit, Cat. # RPN6202). Several indirect probe labelling including mRNA amplification (Eberwine), amino-allyl (Atlas/Clontech), 3D label (3DNA/Genisphere) and Ag X Ab reaction (MicomaxTSA/NEN) are widely used in the laboratories. Labelled probes are then mixed and hybridized to the arrayed DNA spots. The hybridization procedures affect reproducibility, the hybridization conditions such as probe concentration, ionic strength, and temperature, largely depend on the length of the DNA fragments present on the array and need to be optimized for a given experiment. The high-quality hybridization translates to high specificity in combination with maximum signal and minimum background. After the hybridization, slide is scanned by laser (GenePix 4000B from Axon, 428 ArrayScanner from Affymetrix). Laser causes excitation of fluorescent labelled cDNA probes. The emission is measured using a

scanning confocal laser microscope and data are analyzed by appropriate software. The absence of the fluorescence of the specific spot means that complementary mRNA is not present in the sample. If the fluorescence is present, the intensity of the signal is dependent on the level of a particular mRNA in the examined sample. If particular mRNA from test sample is in abundance, the spot with a complementary probe will be green; if the concentration of the particular mRNA is higher in reference sample, the spot will be red. If both samples contain the same amount of a given mRNA, the spot will be yellow. The generation of such vast volumes of data from expression profiling of tens of thousands genes requires specialized methods to catalogue, group, analyze, and interpret the biological data. In general, most of the analysis tools today use computational methods to group (cluster) genes or experiments with similar profiles of changes in expression level (hierarchical clustering, K-mean clustering, divide clustering, and self-organizing map) (Sherlock 2000). Hierarchical clustering (Eisen et al., 1998) is the most commonly used tool in gene expression analysis. An interesting application of this approach is the clustering of tumors to find new possible tumor subclasses. In a recent paper by Alizadeh et al. (2000), diffuse large B-cell lymphoma (DLBCL) was studied using 96 samples of normal and malignant lymphocytes. Applying a hierarchical clustering algorithm to these samples they showed that there is diversity in gene expression among the tumors of DLBCL patients. They identified two molecularly distinct forms of DLBCL, which had gene expression patterns indicative of different stages of B-cell differentiation. Interestingly, these two groups correlated well with patient survival rates, thus confirming that the clusters are meaningful.

cDNA microarray technology is a new and efficient approach to extract data of biomedical relevance for a wide range of applications. In cancer research, it will provide high-throughput and valuable insights into differences in an individual's tumor as compared with mRNA expression in normal cells for identification of tumor-specific molecular markers and molecular tumor classification. Microarray technology is a powerful tool for identifying novel molecular drug targets and for elucidating mechanisms of drug action. Furthermore, microarrays can monitor the global profile of gene expression in response to specific pharmacologic agents, providing information on drug efficacy and toxicity.

Oligonucleotide-based microarrays can be manufactured by depositing oligonucleotide solutions onto glass surfaces or directly synthesizing oligonucleotides on the glass surfaces (Lipshutz et al., 1999). This method, historically called DNA chips, was developed at Affymetrix™, Inc, which sells its products under the GeneChip™ trademark. In the most successful implementation, the oligonucleotides (20 to 25 oligos) are synthesised in situ by using photochemical reactions and a mask technology similar to that routinely used for the manufacture of microprocessors (Fodor et al., 1991; Lipshutz et al., 1999). More recently, peptide nucleic acids (PNAs) chips are a new development and are said to have a higher affinity for the target and so give more accurate results. Currently produced oligo-chip at Affymetrix (http://www.affymetrix.com) have 64,000 "features" on a 1.28 cm^2 chip, each 50 X 50 μm^2 square element containing one to ten million identical oligonucleotides. Over past few years, a number of commercial companies have introduced a range of array products (Hyseq Inc./HyChip™/up to 8,000 oligos, Incyte Pharmaceuticals Inc./GEM/1,000-10,000 oligo and German Cancer Institute/1,000 PNA prototype) covering a wide range of applications. Oligo microarrays could be used for the expression profiling, moreover, oligonucleotide-based microarrays have more controlled specificity of hybridization comparing to cDNA microarrays, which makes oligo microarrays particularly useful for the analysis of single nucleotide polymorphisms (LaForge et al., 2000) or mutation analysis and resequencing (Hacia 1999; Drobyshev et al., 1997).

3.2 PAC/BAC microarrays with allele copy number change

Accumulations of genetic changes in somatic cells induce phenotypic transformations leading to cancer. Among these genetic changes, gene amplification and deletion are most frequently observed in several kinds of cancers. Amplification of an oncogene and/or deletion of a tumor suppressor gene, together with dysfunction of a gene by point mutation, are the main causes of cancer. Comparative genomic hybridization (CGH) is a modified in situ hybridization technique, which allows genome-wide detection and mapping of DNA sequence copy differences between two genomes in a single experiment (Kallioniemi et al., 1992). CGH has been utilized to identify DNA copy number abnormalities in various kind of cancers and several reports have shown its usefulness in screening of the genes involved in carcinogenesis, and also in the identification of prognostic factors in cancer (Hauptmann et al., 2002; Singh et al., 2001; Larramendy et al., 2000; Yen et al., 2001). In CGH analysis, two differentially labelled genomic DNA (test and reference) are co-hybridized to normal metaphase spreads. Chromosomal locations of copy number changes in the DNA segments of the study genome are revealed by a variable fluorescence intensity ratio along each target chromosome. However, the resolution of the cytogenetic method has restricted its use (detection limit for deletions is ~10 Mb and for amplification is >2 Mb). Microarray with BAC/PAC clones is an obvious solution to all the limitations of conventional CGH. In BAC/PAC microarrays, the chromosomal regions (fully or partially covered) are artificially created on glass slides, by array large-insert genomic clones such as BACs. These clones are then used as the targets for hybridization. In experiments, it is important to include some clones from chromosome X, as well as clones from other chromosomes as reference clones while preparing BAC microarrays to cover certain region. BAC/PAC microarrays are reliable and give resolution less than the length of the clone. The measurement precision using BACs for CGH has standard deviation <10% and low-level copy number changes affecting a single clone in a large array can be detected with high statistical confidence (Pinkel et al., 1998; Albertson et al., 2000). BAC microarrays are potentially useful for research and clinical applications in medical genetics and cancer. Usually, 10-15 μg of clean BAC DNA without bacterial genomic DNA contamination is sonicated into 1-20 kb fragments and spotted on slides. Preparing and spotting of BAC DNA is problematic, however, because BACs are single-copy vectors, the yield of DNA from BAC cultures is low compared to that from plasmid-bearing cultures and spotting high-molecular weight DNA at sufficient concentration to obtain a good ratio of signal to noise in the hybridizations may be difficult. To overcome these problems, ligation-mediated PCR (Klein et al., 1999) and BAC-derived inter-Alu PCR were used to generate representation of BAC DNAs for spotting on arrays. For example, sufficient spotting solution (0.8 μg/μl DNA in 20% DMSO) from 1 ng BAC DNA was produced by ligation-mediated PCR to make tens of thousands of arrays (Snijders et al., 2001). The ratios using arrays comprised of BAC representations are essentially identical to ratios from the same BACs. It was reported that modified DNA printing on nature glass surface, that is slightly negatively charged, would improve specific signal-to-noise ratios. Current microarray technologies almost exclusively focus on activating the surfaces, a strategy that invariably introduces hybridization background (Cai et al., 2002). A bifunctional crosslinker (3-glycidoxypropyltrimethoxysilane) compound with an epoxide group that can covalently attach to DNA at a slightly alkaline pH and a reactive group that is specific only on glass surfaces was used to modify DNA so that the extent of modification (modified bases per kilobase pair) was optimal and the hybridization specificity retained. Genomic imbalances are detected by BAC microarrays with high resolution, allowing copy number changes to be associated with individual loci and genomic markers (Solinas-Toldo et al., 1997). In addition to screening for imbalances affecting known tumor suppressor or oncogenes (Bruder et al.,

2001), high-density arrays covering the whole genome can be used to screen for yet unknown regions harbouring genes of pathogenic relevance.

cDNA microarrays was also used for genome-wide analysis of DNA copy-number changes (Pollack et al., 1999; Heiskanen et al., 2000). cDNA microarrays provide an opportunity to analyse in parallel the changes in DNA copy number and expression levels of thousands of genes in the same tumor sample. The results obtained from comparison of breast cancer cell lines and primary tumors with normal human mammary epithelial cells have shown that most of highly expressed genes were not amplified and not all amplified genes were highly expressed. cDNA microarrays containing 10,000 genes or more are routinely employed for gene expression analyses (Iyer et al., 1999), but no resource currently exists for full-genome coverage with large genomic clones. However, reliable diagnosis of a genomic imbalance by cDNA chips relies on the integration of signal ratios from separate cDNAs, whose genes are localized within the respective region.

3.3 *Not*I microarrays to detect copy number and methylation changes

*Not*I clones contain genomic fragments in CpG islands and therefore, *Not*I microarrays also can be used for detection of copy number changes in a given genomic DNA. Unlike BACs, *Not*I clones cannot cover the whole human genome continually. The resolution of *Not*I microarrays (~200-500 kb) for detection of copy number changes is a little bit less than BAC microarrays, but *Not*I microarrays could directly detect the copy number changes of particular genes. Aberrant DNA methylation of CpG sites is among the earliest and most frequent alterations in cancer. Methylation changes are frequently the earliest events in tumor development and can be detected one year before tumor formation (Palmisano et al., 2000). Several studies suggest that aberrant methylation occurs in a tumor type-specific manner (discussed in 2.1.2). It is necessary to develop high throughput assays for methylation detection for large-scale analysis. *Not*I microarrays comprising of *Not*I-flanking sequences on slide can be used for genome-wide screening for deleted, amplified, and methylated *Not*I sites. *Not*I microarrays have the features of both BAC microarrays and cDNA microarrays (Table 9) since almost all *Not*I clones contain genes. At the same time, both BAC and cDNA microarrays cannot be used for detecting methylation. The applications of *Not*I microarrays include i) comparison of normal and malignant cells at genomic and/or RNA level; ii) comparison of primary tumors and metastases; iii) analysis of families suffering from hereditary diseases including cancers.

Table 9. Efficiency of the array method to detect the particular features

Microarray feature	cDNA	CGH (BAC, P1, PACs)	Representation	SNP	RST (*Not*I microarrays)
Homozygous Deletions	Low	Yes	Yes/No	No	Yes
Hemizygous Deletions	Low	Yes	No	No	Yes
LOH	No	No	Yes	Yes	Yes
Amplification	Low/Medium	Yes	Yes	No	Yes
Methylation	No	No	No	No	Yes
Number of available markers	More than 40,000	10,000-30,000	1,500 (polymorphic *Bgl*II fragments per genome	1,300 (can be increased)	10,000-20,000
Connection to genes	Direct	Indirect	No (indirect)	No (indirect)	Direct

The main application of *Not*I microarrays is genome-wide detection of methylation/deletions in tumors for tumor class prediction and discovery, isolation of TSGs. The fundamental problems of genome-wide screening using *Not*I clones are the size and complexity of the human genome, the number of repeat sequences and the comparatively small size of the inserts in *Not*I clones (average 6-8 kb). To solve these problems, a specific procedure, NR (*Not*I representation) labeling (**paper V**), was developed to amplify only regions surrounding *Not*I sites. Other DNA fragments were not amplified. Therefore, only 0.1-0.5% of total DNA is labeled. Sequences surrounding *Not*I sites contain 10-fold fewer repetitive sequences than the human genome on average (Kusenko et al., 2002), and these microarrays are not as sensitive as other methods to the background hybridization caused by repeats. Ribosomal rRNA genes were virtually absent from these *Not*I flanking sequences. The NRs can be efficiently used for genomic subtraction, and any enzyme can be used in this procedure for preparing restriction enzyme representations (RRs). By selection two or three restriction enzymes cutting mainly in CpG islands, this procedure will result in differential cloning of almost all CpG-island-containing DNA fragments. The same RRs can be used for genome screening of corresponding microarrays.

Genomic subtraction of NRs by CODE procedure (**paper II**), *Not*I-CODE procedure, which was used before hybridization is helpful to solve the ''contamination'' problem in experiments. It was shown that methylated or deleted *Not*I sites could be detected with the help of *Not*I-CODE in two RCC biopsies with 30-40% of normal cells contamination. Using *Not*I microarray, it is possible to discriminate between deleted and methylated sequences. To achieve this aim, NR should be produced using DNA that is unmethylated. This can be done by different approaches: limited PCR amplification after the first digestion with restriction enzyme(s), enzymatic demethylation, etc.

The creation and use of microarrays at the genomic level may provide information unavailable at the level of mRNA/cDNA (for example, methylation or silencing of specific alleles, hemizygous deletions, epigenetic factors, genetic predisposition, working with old samples, etc.). Furthermore, *Not*I microarrays are more sensitive than cDNA microarrays in several ways, because genomic markers are normalized naturally and differences in their copy numbers cannot reach 10^4 times that is valid for the expression of many genes. Many genes are expressed at level below 50-100 copies per cell and probably cannot be properly analyzed with cDNA microarrays at all. Another strong advantage of *Not*I microarrays compared with cDNA microarrays is that there is no standard for comparing expression profiles (Brazma et al., 2001). RNA is not a stable molecule and differences in RNA content between normal and cancer cells depends on many different factors e.g. physiological conditions. With the expression microarrays, it is sometimes difficult to identify the first events and first genetic lesions that lead to the development of cancer.

It is important to mention that *Not*I microarrays were not designed to replace expression microarrays; on the contrary, they should be used as a complementary approach that can yield additional information.

With the increasing understanding of the role of DNA methylation in carcinogenesis, several new methodologies have been developed to facilitate genome-wide screening for changes in DNA methylation (Palmisano et al., 2000; Costello et al., 2000; Rountree et al., 2001; Adorjan et al., 2002; Galm et al., 2002). An array-based method, called differential methylation hybridization (DMH or CpG islands microarrays, CGI) (Fig. 12) was developed to be used for a genome-wide screening of hypermethylated CpG islands in tumor cells (Huang et al., 1999).

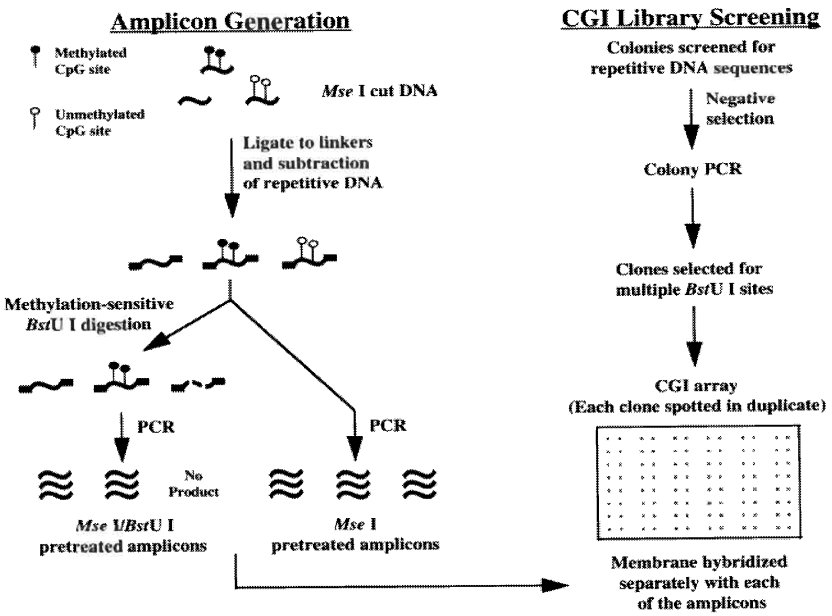

Fig. 12 Schematic flowchart for differential methylation hybridization. The diagram illustrates the preparation of amplicons used as hybridization probes and selection of CpG island genomic clones gridded on high-density arrays.

In this approach, many CpG islands loci (Fig. 3) derived from a genomic library, CGI (Cross et al., 1994) were pre-screened with ^{32}P-labeled Cot-1 DNA. Clones negative or weakly positive for Cot-1 hybridization signals were picked. Among these clones, the clones containing multiple *BstU*I sites were spotted on CGI arrays. Genomic DNA from samples was first treated with the four-base pairs cutter *Mse*I. Its recognition sequence, TTAA, rarely occurs within GC-rich regions, leaving most CpG islands intact. Moreover, the digested fragments were expected to match the *Mse*I-digested inserts originally used in the construction of the CGI genomic library (Cross et al., 1994). After this, digested DNA was ligated to linkers. Repetitive sequences such as the *Alu*I and *Kpn*I families were removed from the digests using a Cot-1 subtractive hybridization approach (Craig et al., 1997). Half of the subtracted DNA was treated further with methylation-sensitive endonuclease *BstU*I, whose recognition sequence, CGCG, occurs frequently within CpG islands but rarely in bulk DNA. This endonuclease was selected for the methylation analysis because more than 80% of the CGI inserts contain *BstU*I sites (Cross et al., 1994). Both *BstU*I-digested and undigested control DNAs was used as templates for linker-PCR. Genomic fragments containing unmethylated *BstU*I sites were cut and could not be amplified in the treated samples, whereas the same fragments were amplified in the undigested, control samples. The PCR products designated as '*MseI*-pre-treated amplicons' or '*Mse*I-*BstU*I-treated amplicons' were used as probes for screening hypermethylated sequences. *Mse*I-*Bstu*I-treated amplicons derived from the 17-paired tissues of breast tumors and normal controls were labelled with Cy3 and Cy5, and then hybridized to CGI arrays. 1% (83/7776 loci) of examined CpG islands were hypermethylated in this patient group (Yan et al., 2001). Shi et al. (2002) reported the use of expressed CpG island sequence tags (ECISTs) for dual analysis of DNA methylation and gene expression in breast cancer cell line and results showed that ECISTs and CGI arrays are

effective markers for identifying novel genes, the expression of which is silenced via CpG island hypermethylation.

Compared to *Not*I microarrays, CGI microarrays have several limitations. They are not suitable to study copy number changes; unlike the *Not*I microarrays, any incomplete digestion will produce an artificial positive signal; whole human genome DNA was used for labelling, etc (Table 10).

Table 10. Comparison of *Not*I and CGI microarrays

Feature	*Not*I microarrays	CGI-microarrays
Uncompleted restriction digest	No effect	Artificial result
Specificity of labeling	0.1-0.5% of the total human DNA	100% total human DNA
Repeats	10% compared to the average in human genome	Approximately the same as in average
rRNA genes	No	Yes
Homozygous deletions	Yes	No
Hemizygous deletions	Yes	No
Hemizygous methylations	Yes	No
Oligo microarrays	Yes	???
Homozygous methylation in cancer cells	Yes	Yes
Quality of clones	All sequenced, all contain genes	Partly sequenced, many reiterated clones and repeat sequences like LINE etc.
Number of available clones	> 5,000	Unknown

There are several other high-throughput approaches for detection aberrant methylation. David Muddiman presented a novel and potentially high-throughput approach to the characterization of alterations in DNA based on methylation of cytosine residues of CpG islands through the application of electrospray ionization-Fourier transform ion cyclotron resonance-mass spectrometry (ESI-FTICR-MS) (Hannis and Muddiman 1999). In Muddiman's approach, methylation-specific PCR incorporates bisulfite conversion of unmethylated cytosines to uracils in the template DNA with subsequent conversion of uracil to thymine during in vitro amplification, and the 15-Da difference between cytosine and thymine is detected by ESI-FTICR-MS. In this approach, locations of methylation can be precisely located and correlated with patterns of repeats, and individuals can be typed for number of trinucleotide repeats at a particular locus. Only few genes can be tested simultaneously.

Restriction landmark genomic scanning method indeed allows analyzing thousands of *Not*I sites per experiment (Costello et al., 2000). However, this approach is rather technically challenging.

Tompa et al. (2002) described a simple technique for genome-wide mapping of DNA methylation patterns. Fragmentation by a methyl-sensitive restriction endonuclease is followed by size fractionation and hybridization to microarrays. It was demonstrated the utility of this method by characterizing methylation patterns in *Arabidopsis* methylation mutants.

Methylation-specific oligonucleotide (MSO) microarray which combines bisulfite DNA assay and oligonucleotide microarray has been developed for detecting changes of DNA

methylation in cancer (Gitan et al., 2001; Adorjan et al., 2002). The method uses bisulfite-modified DNA as a template for PCR amplification. The amplified product, therefore, may contain a pool of DNA fragments with altered nucleotide sequences due to differential methylation status. A test sample is hybridized to a set of oligonucleotide (19-23 nucleotides in length) arrays that discriminate methylated and unmethylated cytosine at specific nucleotide positions, and quantitative differences in hybridization are determined by fluorescence analysis. The cross-hybridization between imperfect-match probes and targets resulted in careful selection the optimal sequence composition for each oligonucleotide probe and limited the wide use of this method. Another alternative, also based on the principle of oligonucleotide microarray, can achieve a comprehensive analysis of DNA methylation. In this case, oligonucleotide targets are designed such that their 3' termini end just before interrogating CpG sites; these targets are arrayed on glass slides by attaching to the surface via their 5' ends. Unlabeled bisulfite-modified probes are prepared. After target-probe annealing, in situ polymerase reaction is performed to allow extension to only one base, which is either Cy5 (red)-ddTTP or Cy3 (green)-ddCTP, away from an oligonucleotide primer. The extended molecule is identified on the basis of the incorporation of different fluorescence dyes for either converted or unconverted nucleotides, that is, unmethylated or methylated cytosine residues, at the interrogating CpG sites. Quantitative analysis of methylation is determined by two-color fluorescence analysis.

The general limitation of these MSO microarray techniques is that each tested gene must be amplified by gene-specific primers. Therefore, only limited number of genes can be tested in one experiment.

4. Novel methodologies in cancer research

4.1 COP and CODE procedures for cloning different sequences

The genes responsible for many genetic diseases were successfully cloned during last years. This success was mainly related to construction of the detailed genetic and physical maps of human genome, to improved computer programs and to the human genome sequencing effort. However, now it seems clear that already cloned disease genes were relatively easy to clone. It became obvious that the development of new methods providing new opportunities and more efficient analysis than the existing approaches is essential for cloning new disease genes responsible for complex genetic disorders (psychiatric disorders, obesity, carcinogenesis, etc.). New methods, cloning of polymorphic sequences (COP) and cloning of deleted sequences (CODE), were developed for cloning different sequences in our laboratory.

4.1.1 COP (cloning of polymorphisms)

Single-nucleotide polymorphism (SNP) is the most widespread form of DNA polymorphism in human genome (Gu et al., 1998; Brookes 1999).

We developed a simple and robust procedure for cloning SNPs, which represent actually restriction fragment length polymorphisms (RFLP). The procedure was called COP, cloning of polymorphisms (**paper I**). We used DNA isolated from blood of two different persons (A and B). To increase complexity of the DNA representation DNAs A and B were digested with three enzymes having the same sticky ends, i.e. *BamH*I+*Bgl*II+*Bcl*I and then ligated to the Blsubtr1/2 linker. Following ligation, PCR of DNA B was performed in the presence of the four dNTPs and of biotinylated primer Pbsub. PCR of the DNA A was run using the same primer but without biotin and with dUTP instead of dTTP. After denaturing, first hybridization

was run at 1: 50-100 ratio of DNA B to A, and then the mixture was consequently treated with uracil-DNA glycosylase and mung bean nuclease. The subtracted product was purified with streptavidin-coupled magnetic beads and the cycle was repeated once more (see the figure in **paper I**). The final product was again PCR-amplified and cloned into the plasmid vector. The identified SNPs represent RFLPs, which may be detected using Southern blotting and/or PCR amplification.

Importantly, polymorphic fragments should be PCR-enriched because of their relatively small size. At the same time, digestion with restriction enzymes may generate two fragments in DNA B from single RFLP fragment present in DNA A. This may increase the probability of forming homodimers, partially from the replacement of a short plus strand with a longer plus strand in a duplex originally composed of a long minus strand and a short plus strand. This duplex formed by two long strands will be more stable than the duplex formed by short and long strands. There are two possible ways of the RFLP enrichment. Enzyme digestion of one long DNA A fragment (e.g. 15 kb) may generate a long fragment (A1, e.g. 14.5 kb) and a very short fragment (A2, e.g. 0.5 kb). After the first step of PCR amplification only fragment A2 would be present in amplified DNA. If enzyme digestion generates two short fragments (e.g. 0.5 kb and 0.4 kb) from a medium-size DNA fragment (e.g. 0.9 kb), all three fragments would be amplified after first PCR and would therefore participate in subsequent reactions. It is important to note that the COP procedure allows cloning of short and long polymorphic fragments (**paper I**). In both scenarios, after the first PCR amplification complexity of the DNA would become considerably lower and the part of polymorphic (mainly shorter) fragments would dramatically increase.

Three enzymes (*BamH*I, *Bgl*II and *Bcl*I) were used for digestion of the human DNA. These enzymes generate fragments of about 1.5 kb average sizes. Alternatively, *Sau3*A, which generates restriction fragments of about 250-300 bp, may be used.

If we assume that two homologous chromosomes originating from different parents have on average one SNP per 1 kb of DNA (Gu et al., 1998; Brookes 1999), then at least 10,000 polymorphic fragments could be cloned using the COP procedure. In this calculation we assume that the size of the haploid human genome is $3x10^9$ bp and roughly 1.5 kb of human DNA contain one hydrolyzable site for one of the enzymes used in the study, i.e. *BamH*I, *Bcl*I or *Bgl*II. Using *Sau3*A (one site per 250 bp), 40,000 fragments could be cloned. Using combinations of restriction enzymes with 4-bp recognition site (4-bp cutters), including enzymes recognizing multiple and non-palindromic sequences, virtually all SNPs could be identified and cloned by this procedure. Using only two 4-bp cutters containing CG pair(s) in the recognition site could allow cloning of approximately 80,000 SNPs located mostly in gene-rich regions. This effort would be as productive as SNP generation by EST cloning and sequencing program, but much cheaper. It should be noted that the SNPs of this type might be located not only in expressed but also in promoter/enhancer regions. These SNPs are functionally very important but cannot be detected by ESTs sequencing approach.

The COP procedure resembles the 'RFLP subtraction' (Rosenberg et al., 1994) because both result in cloning RFLP different for two DNAs by subtractive procedure (Rosenberg et al., 1994; Corrette-Bennett et al., 1998). However, these methods are entirely distinct in biochemical techniques applied for cloning RFLP and in the obtained results. The RFLP subtraction is a considerably more complicated and laborious method (Table 11). In addition to the multiple (3-4 cycles) 'classical' subtractive hybridization steps it uses gel purification, a reassociation step to remove poorly hybridizing DNA, subtraction based on representational difference analysis (Lisitsyn et al. 1993), and multiple combinations of linkers and PCR primers. The most efficient enrichment steps in the COP procedure (with UDG and mung bean nuclease) are not used in the RFLP subtraction. The 'RFLP subtraction' method results in cloning RFLP segments that is present in one DNA preparation (after agarose gel purification) and absent in the other. The COP procedure yields DNA fragments that are heterozygous in one DNA sample but homozygous in the other.

Representational difference analysis (Lisitsyn et al., 1993) may also be used to clone polymorphic sequences but the main obstacle here is even not its complexity but its low productivity as it yields only few polymorphic sequences.

The efficiency of the COP procedure, if necessary, may be improved by adding one more cycle of purification. The COP procedure appears to possess all features essential for the ideal method to detect new SNPs (Kwok and Chen 1998): it is highly efficient, simple, inexpensive, based on safe reagents, and possible to perform on standard laboratory equipment.

Table 11. Comparison of COP and RFLP subtraction procedure

Step	COP	RFLP subtraction
Gel purification	No	Yes
Number of oligonucleotides	4	10
First PCR amplification	Yes	Yes
dUTP, UDG treatment	Yes	No
Mung bean nuclease treatment	Yes	No
Subtractive hybridization (cycles)	2 cycles	3-4 cycles at step 6 and 5-6 altogether including step 8-9
Second PCR amplification	Yes	Yes
Removing poorly hybridizing DNA (HindIII digestion, avidin-biotin subtraction)	No	Yes
Removing poorly hybridizing DNA (subtractive hybridization, RDA)	No	Yes
Third PCR amplification	No	Yes

4.1.2 CODE (cloning of deleted sequences)

Although subtractive methods represent potentially powerful tools for the identification of deleted sequences, they have not been extensively applied for isolation of TSGs, probably because of the great complexity of the human genome. However, this approach produced rewarding results in less complex genomes, like those of yeast and bacteria. Various successful approaches for subtraction at cDNA level were suggested and used (Diatchenko et al., 1996, 1999), but, among all genomic subtraction methods, only a modified version called representational difference analysis (RDA) (Lisitsyn et al., 1993) has demonstrated reproducible results.

One of the main ideas of this approach was to use for genomic subtraction only a subset of genomic sequences, e.g. all *Bam*HI fragments shorter than 1 kb. Since complexity of the genome is greatly reduced, the expected results look promising. Another important point in RDA is that this method uses not only subtractive but also PCR kinetic enrichment to purify restriction endonuclease fragments present in only one of the two populations of DNA fragments (Lisitsyn et al., 1993).

For preparation of driver and tester amplicons DNAs are digested with endonucleases *Hind*III, *Bgl*II, etc. and ligated to adaptors. After ligation DNAs are PCR-amplified and digested with *Bam*HI to remove adaptors. After gel purification from adaptors, tester DNA is ligated to new dephosphorylated adaptors, then mixed with driver DNA amplicon (1:80), denatured and hybridized. After filling-in with Taq-polymerase at 72°C, the fragments in DNA mixture are PCR-amplified with the same oligonucleotide to which the tester DNA was ligated. The second oligonucleotide in the adaptor would not ligate to the tester DNA as it is unphosphorylated at the 5' end and cannot serve as a primer for PCR amplification. Only tester homohybrids are successfully exponentially amplified. Heterohybrids show only linear amplification. At the next step DNA mixture is sequentially digested with mung bean nuclease (to destroy ssDNA) and *Bam*HI, ligated to the third adaptor and PCR-amplified. Then the cycle of enrichment is repeated. After each cycle the mixture becomes more and more

enriched in DNA molecules forming homohybrids A; finally these molecules are cloned and analyzed. Three to four cycles are usually applied for the RDA and 1-3 different DNA fragments may be cloned.

However, while having many advantages, RDA still possesses some limitations. The technique is complicated and prone to minor impurities. The size of the differential product is usually between 250-350 bp, inconvenient for many possible applications. It is important to note that there are no publications describing RDA method using the enzymes containing CG pair in the recognition site and thus associated with CpG islands, like *Sal*I, *Xma*III, *Not*I. Thus, differentially cloned DNA fragments have no relation with genes. Non-functional regions are often deleted in the human genome, and cloning of such segments will yield no valuable information.

Another drawback is a relatively low productivity of the RDA: only few probes can be generated per one experiment (Lisitsyn et al., 1993, 1994, 1995). The authors suggested that this limitation could be obviated by diminishing the number of rounds of hybridization/amplification or increasing the complexity of representation (Lisitsyn et al., 1993). Increasing of complexity, however, would result in confronting new challenges as the RDA failed when the complexity of amplicons was not sufficiently low (Lisitsyn et al., 1993).

We attempted to develop a new approach for subtractive hybridization in order to overcome the above-mentioned limitations. The key aims were the following: first, to simplify the procedure; second, to avoid PCR kinetic enrichment steps (e.g. exponential versus linear amplification) which result in cloning very small DNA fragments.

In the previous section we have described a new procedure for cloning of polymorphic sequences (COP) (**paper I**). Since COP proved to be a rather simple, efficient and robust procedure we decided to modify this scheme to clone homozygously deleted sequences. This task is considerably more difficult than cloning of hemizygously deleted polymorphic DNA fragments (Rosenberg et al., 1994; Diatchenko et al., 1999), as shown by the obvious fact that homozygous deletions never exceed a few Mb while hemizygous deletions may be considerably larger than 100 Mb.

The general scheme of the cloning of deleted sequences is similar to CIS and COP procedure (**paper II**). Again, to increase complexity of representation, the tester and the driver DNAs are digested with *Bam*HI+*Bgl*II+*Bcl*I and then, special linkers (Blsubtr1/2) were ligated to the DNA fragments. The driver DNA is amplified with dUTP and unmodified primers and tester DNA is amplified with biotinylated primers and normal dNTP. The products of DNA amplification (at average 1-2 kb), after denaturing and hybridization (1:100 ratio of tester: driver DNA), are digested with UDG to destroy all drivers DNA and then with mung bean nuclease to digest all non-perfect hybrids. The resulting tester homohybrids are purified and concentrated with streptavidin beads. The cycle is repeated three times. Finally, the product is PCR-amplified and cloned.

We believe that the CODE is a simple, effective and robust procedure able to successfully isolate deleted genomic sequences. In contrast to the RDA, many different probes may be generated in one experiment and the procedure is easy to perform. The CODE protocols do not exploit the enrichment basing on the difference between exponential and linear amplification. This keeps biases and artefacts generated by PCR to a minimum, and subtractive enrichment becomes the critical step. The size of the subtractive products was 300-700 bp, but it may be significantly increased (up to 1-2 kb). We recently performed the COP and the CIS procedures using long-distance PCR with decreased number of amplification cycles. As a result, both procedures yielded mainly 1-2.5 kb DNA fragments. The same modification may be applied to the CODE procedure.

Another important difference between CODE and other genomic subtraction methods (e.g. RDA and 'RFLP subtraction' method) is that CODE allows to clone polymorphic fragments of various size hemizygously deleted in tumor DNAs. RDA and 'RFLP subtraction' methods

yield probes detecting hemizygous loss of smaller fragment in the driver DNA. Thus all probes cloned using these methods showed two alleles in tester DNA: a large (e.g. 7 kb) and a small (e.g. 0.6 kb) the small allele was always present in the tester but absent in the driver DNA. In the case of the CODE method, polymorphic fragments with similar length can be differentially cloned. As it was demonstrated previously, the COP procedure (and conceivably the CODE) allowed differential cloning of similar fragments (e.g. 2 kb and 2.3 kb) and detected the loss of a larger polymorphic fragment.

The general scheme of the CODE procedure is similar to the COP (**paper I**). However, several important differences exist. First of all, the approaches used for differential cloning in these two methods are different. The most important in CODE procedure is differential cloning by subtractive hybridization. In the COP procedure, more important is enrichment of small-size polymorphic fragments by PCR amplification.

If we assume that homozygous deletion in a tumor is 0.7 Mb and the average size of the DNA fragments after simultaneous digestion with *Bam*HI, *Bcl*I and *Bgl*II is 1.5 kb, then only approximately 470 DNA fragments will be located in the deleted region. These fragments represent approximately 0.01 % of all DNA fragments available in a diploid human genome (5-6 Gb). At the same time we can expect that from $4x10^6$ DNA fragments approximately 20.000 or 0.5% are polymorphic. We assume that two equivalent chromosomes have on average one SNP per 1 kb (Brookes 1999).

It means that in the CODE procedure we need to increase at least 50 times the selectivity of the method, because homodeletions are much less frequent than hemideletions (0.01% versus 0.5%). Therefore running the differential cloning one should screen all available molecules.

That is why in the CODE procedure we use more subtractive steps (four) and less cycle for PCR amplification (step 1: 20-25 cycles and step 2: 15 cycles) then in the COP procedure (two subtractive steps and 30 cycles for both steps of PCR amplification).

A cDNA subtraction method using PCR amplification with dUTP and UDG treatment has been recently reported (Sugai et al., 1998). Our CODE procedure is considerably different from this cDNA subtraction method. Importantly, the subtraction efficiency of the method (Sugai et al., 1998) was very low (more than 90% of the clones were present in both mRNA preparations), despite the lower complexity of mRNA compared to genomic DNA.

4.2 CIS procedure for cloning identical sequences

Searching for the regions that are identical by descent (RIBD) should be very important to identify the genes involved in genetic diseases including cancer. Identities by decent means that the genome segments are identical in two individuals as inherited from their common ancestor. The RIBD in individuals affected with a disease are likely to contain the disease gene. We assume that a predisposing mutation in the disease gene is rare and appeared many generations ago in the person from whom all the affected individuals inherited the disease gene. Search for the RIBD using microsatellite markers is very laborious and expensive. The higher the number of generations separating two individuals from their common ancestor the more markers should be tested to identify the RIBD (from hundreds of markers for close relatives to many thousands for distantly related individuals).

A new method called genomic mismatch scanning (GMS) was suggested to solve this problem (Nelson et al., 1993). This method is rather powerful, since in one experiment two large genomes are compared and the RIBD are possible to identify. Initially it was shown that this method is applicable to yeast genomes (Nelson et al., 1993), but application of this approach to mammalian genome met several obstacles. The major problem was high complexity of the human genome including presence of numerous and nearly identical repeat sequences. Only

recently it was demonstrated (Mirzayans et al., 1997; Cheung and Nelson 1998; Cheung et al., 1998; McAllister et al., 1998) that GMS efficiently searches for the RIBD sequences and may be applied for the complex mammalian genomes to localize the hereditary disease genes.

The principle of GMS (Nelson et al., 1993; Cheung et al., 1998) is the following. Two DNAs from two individuals (having common ancestors) are separately digested with *Pst*I to yield fragments with protruding 3' ends. The 3' extensions serve to protect these ends from digestion by exonuclease III (ExoIII) at later steps. One of the DNAs is then methylated at all GATC sites with *E. coli* Dam methylase (DAM+). The other DNA remains unmethylated. These two DNAs are then mixed in equal amounts, denatured and reannealed. Digestion of the reannealed DNAs with *Dpn*I and *Mbo*I at fully methylated and at unmethylated GATC sites respectively results in cleavage of the homohybrids to yield smaller duplexes with either blunt (*Dpn*I) or 5'- protruding ends (*Mbo*I). Heterohybrids and imperfect homohybrids with mismatches in GATC sites are resistant toward *Dpn*I and *Mbo*I and are not affected by this treatment.

Discrimination between perfect, mismatch-free heterohybrids and those with mismatches is done by three *E. coli* mismatch repair proteins MutH, MutL and MutS (MutHLS). Single-base-pair differences and small insertions/deletions (up to 4 bp) are efficiently detected by the mismatch-repair proteins (Ellis et al., 1994). They introduce a single-strand nick on the unmethylated strand at GATC sites specifically in the mismatch-containing duplexes. Only perfect duplexes escape nicking at this step. All DNA molecules except mismatch-free are further degraded with ExoIII, a 3' to 5' exonuclease specific for double-stranded DNA (dsDNA). During this treatment double-stranded nicked molecules are either converted into single-stranded DNA fragments (ssDNA) or nicks are considerably enlarged and large single-stranded regions appear in dsDNA molecules.

After treatment with ExoIII all DNA molecules with ssDNA regions are removed by adsorption to benzoylated naphthoylated DEAE cellulose (BNDC) column. This column efficiently absorbs ssDNAs and dsDNA molecules containing about 100-bp-long ssDNA regions. Thus the full-length unaltered heterohybrids are purified from the other DNA fragments. The GMS DNA is then labeled using Alu- repeat-specific primers and hybridized to the microarrays (e.g. DNA from YAC clones covering the total human genome). Using GMS for different pairs of individuals the minimal RIBD present in all affected persons may be identified.

The described method has some drawbacks. MutHLS enzyme is not commercially available, it does not recognize all mismatches and has different sensitivity to different mismatches (e.g. sensitivity to a mismatch for the G-T pair is higher than for the G-G pair, etc. (Nelson et al. 1993; Au et al. 1992). The scheme is not working for heterohybrid molecules containing for instance 8-80 bp mismatches, since MutHLS proteins do not cleave them and BNDC column do not remove them. Not all *Pst*I fragments contain an Alu repeat. Taken together, these drawbacks increase background hybridization, which lowers the efficiency of the procedure.

Since high potential of the GMS approach was evident, we decided to improve this strategy and to develop a modified procedure to clone sequences identical for two DNAs. We called it CIS, cloning of identical sequences (**paper III**).

As mentioned above, using the MutHLS enzyme generates certain inherent limitations. For instance, not all *Pst*I fragments contain the GATC site. This introduces problems both for methylation with Dam methylase and for nicking with MutHLS (GATC is the recognition site for both these enzymes). Moreover, it is difficult to estimate the completeness of methylation for the GATC sites in genomic DNA. Another problem stems from the capacity of MutHLS to digest hemimethylated GATC sequences: activity of this endonuclease does not depend on the presence of mismatched base pairs (Welsh et al., 1987).

It is important to note that MutHLS is not a robust enzyme. Its activity depends on the distance between a mismatch and GATC as well as on the size of a DNA fragment (Au et al., 1992).

The CIS procedure avoids these problems. All DNA fragments contain at least four recognition sites for *Mvn*I, all cytosines are methylated, we do not use BNDC columns, etc. The scheme of the CIS procedure is shown in **paper III**. DNAs A and B are digested with *Bam*HI and ligated to special linkers (Blsubtr1/2) containing two recognition sites for *Mvn*I. Consequently, all molecules contain at least four sites for *Mvn*I. DNA A is amplified by polymerase chain reaction (PCR) in the presence of dUTP and m5dCTP, all cytosines should be methylated. DNA B is amplified by PCR in the presence of dCTP and biotinylated primers. Then, these two DNAs are mixed in equal amounts, denatured, annealed, and digested with *Mvn*I. This enzyme digests only dsDNA molecules without methylcytosine; therefore it would digest all homohybrids B, which contain at least four sites for *Mvn*I.

At the next step, DNA mixture is treated with mung bean nuclease to destroy all imperfect hybrids and ssDNA. Therefore, after this treatment only perfect homohybrids A and perfect (without any mismatches) heterohybrids remain in the mixture. Then DNA mixture is treated with UDG (uracil-DNA glycosylase). This enzyme removes uracil base from the DNA and thus destroys all DNA A. Finally, we end up with ssDNA B which is identical for the persons A and B. Using magnetic beads this DNA is concentrated and purified and then PCR-amplified using a specific primer.

Several features make the CIS procedure more advantageous than GMS but further experiments are needed to prove that this scheme is useful for identification of hereditary disease genes. To test this, we are currently applying the CIS procedure to a panel of DNA pools from families suffering from familial nasopharyngeal carcinoma (NPC) in order to identify potential NPC susceptibility genes. Comparison of the two great grandchildren (with different grandparents) resulted in detection of 0.3%-0.4% of hybridizing PAC clones (RPCI1, UK HGMP, Resource Centre) confirming high selectivity of the CIS procedure. Only CIS fragments larger than 1 kb were used for labeling. Larger DNA fragments were selected by agarose gel electrophoresis. This additional experimental step was added to diminish the background hybridization since the majority of the human DNA fragments originated from different sources and, if longer than 1 kb, should have at least one mismatch. In contrast, many DNA fragments shorter than 500 bp should be completely identical. However, the experiments showed that this additional selection was not essential, since the CIS procedure is by itself sufficiently selective.

4.3 *Not*I microarrays for isolation TSGs

Cancer development with both alleles inactivation of TSGs by genetic or epigenetic changes is widely accepted now. Positional cloning process is the common approach for cloning candidate TSGs (Fig. 13).

The way to isolation of TSGs is following two-hit-model, namely, finding genes with abnormalities on two alleles in tumors. There is no single technology at present that could be used for genome-wide detection of all the types of abnormality including large deletions, rearrangements, base substitutions, small insertions and deletions, amplifications, and epigenetic changes such as methylation, that are present in cancer cells; new technology will be required. In order to isolate TSGs, *Not*I microarrays were developed with more advantages than others arrays in genome-wide detection of deletion or methylation of tumor cells (discussed in 3.3). Except *Not*I microarrays, no any other methods could be used to detect genetic and epigenetic changes (deletion and methylation) in tumors at same time.

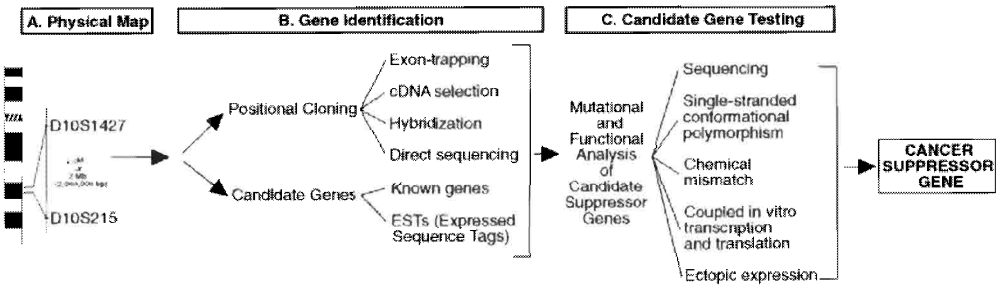

Fig. 13 Positional cloning strategies for gene identification.

We have constructed chromosome 3-specific *Not*I microarrays comprised 150 unique *Not*I clones representing 61 known genes and 49 unknown expressed sequence tags alone human chromosome 3 (**paper V**) and total human genome *Not*I microarrays with the 22,551 sequenced, unique *Not*I clones are under construction (Kutsenko et al., 2002). In the pilot experiments, homozygous deletion and methylation were easily detected (**paper V**).

4.4 GIT for functional identification of TSGs

Designating a candidate gene as a TSG is still challenging. Mutational inactivation analysis of candidate genes in tumor biopsies could be very complicated in case with dominant negative mutations, LOH or heterozygous mutations in genes whose dosage is critical and genes in which one allele is imprinted (Nicolaids et al., 1998; Kohler et al., 1999). In such cases, only functional approaches can help to identify the TSG from candidate genes. Functional evidence of the ability to suppress cellular proliferation cannot always be used for the identification of a tumor suppressor. In many tumors carrying multiple genetic alterations, reconstituting a TSG is not sufficient to reverse completely their malignant phenotype (Zhou et al., 1994). It must be stressed that reconstitution experiments designed to demonstrate direct growth inhibition of tumors are difficult to perform and analyze. It has been shown that a 1.6 cM region in 3p21.3 is deleted during the passage of human chromosome 3 microcell/mouse A9 fibrosarcoma hybrids (Szeles et al., 1997). Based on this, our hypothesis was that TSG must be inactivated in growing tumors in experimental conditions as it happened in nature. This inactivation of a TSG can be achieved in different ways: by deletion, mutation, methylation etc. Thus we decided to develop functional test that will mimic the natural processes. We decided to address this question by asking whether a known suppressor genes, *RB* and *p53*, would show the corresponding behavior and whether this feature can be developed into a functional test for the identification of TSGs. Wild type and mutated RB genes were built into pETI (Elimination Test Integrating) vector that permitted tetracycline regulated expression of the gene in cancer cells growing not only in vitro, but in vivo as well. Wild type but not mutated *RB* and *p53* genes were deleted/inactivated during tumor growth (**paper VII**; Li et al., manuscript). No inactivation/deletion of control genes (mutant *RB, 3PK, MLH1, rhoA*) was observed even after two passages in SCID mice (**paper VII**, **paper VIII**). These results suggest that tests for the identification of TSGs may be based on their functional inactivation in vivo, rather than on growth suppression. From these experiments, we suggested a new functional test for TSG identification – GIT (gene inactivation test). This test is based on the functional inactivation of the analyzed genes in contrast to existing tests based on growth suppression.

When the transferred candidate genes cause suppression of cell growth or dosage effect of candidate genes should be investigated, tight regulation of transgene expression is required. Retrovirus-based vector and tetracycline-regulated vectors inducing the regulation of gene expression are used in GIT.

Tetracycline promoter activating gene (tTA) was inserted into a retroviral vector (pLNCx-CMV) to get new construct named pRV-tTA7. This retrovirus was used to establish tetracycline regulatable mouse A9 and human KRC/Y (RCC) and HeLa cell lines expressing tTA in the first experiment. We have obtained 12 more cell lines (prostate, small cell lung cancer, non-small cell lung cancer, nasopharyngeal, etc) expressing high titers of retrovirus pRV-tTA7 (**paper VIII**). Candidate genes should be tested in different cell lines in GIT because TSGs are tissue-specific.

A tetracycline-based system which is a heterologous system based on prokaryotic mechanisms, is commonly used for inducible gene expression in mammalian cells (Gossen and Bujard 1992). The system uses the repressor protein of an *Escherichia coli* tetracycline-resistance operon (tet-repressor) and the operator sequence to which it binds. The initially described system exploited two plasmids that integrated into the genome. The first critical component contained the gene of interest under the control of a *cis*-regulatory element (tet-promoter), in which seven copies of the 42 bp tet operator sequence located just upstream of the minimal human cytomegalovirus promoter (P_{minCMV}) lacking the strong enhancer normally associated with the CMV promoter. In our system, two tetracycline vectors, pETI and pETE (Elimination Test Episomal) contain this part. The second critical component was a hybrid regulatory protein based on the tet-repressor, tetracycline-controlled transcriptional transactivator (tTA). Fusing the tet-repressor to the activation domain of VP16 from herpes simplex virus created this protein. In our system, hybrid tTA was expressed from retrovirus pRV-tTA7. When tTA bound to the tetracycline operators, the VP16 domain activated transcription from the minimal CMV promoter. The addition of tetracycline or doxycycline prevents tTA from binding to the operator sequences then stops the VP16 domain activating transcription. The level of expression depended on the concentration of tetracycline or doxycycline (Kistner et al., 1996). In successful experiment, virtually no leakage was detected and several hundred-fold activation of expression was achieved (Gossen and Bujard 1992). The tet-system could also control individual gene activity quantitatively and reversibly in a temporal manner in transgenic mice (Kistner et al., 1996), so investigating the dosage effect of candidate genes in GIT is possible.

Our first experiments demonstrated that genes in pETI vectors were not fully repressed, even in the presence of tetracycline. We suggested that this effect could partly be influenced by flanking sequences and that an episomal vector can improve the system. We have constructed such a vector - pETE and have shown that genes cloned in this vector have a significantly decreased leakage but have the same level of expression in the absence of tetracycline. More than 16 candidate genes were introduced in the GIT and the result will be discussed in the last part.

SCOPE OF THIS THESIS

This thesis cover the whole process from deletion mapping to isolation of candidate genes and further to functional identification of tumor suppressor genes by development several novel approaches. This thesis can be divided in three parts, first two parts concern to isolation of candidate TSGs and last part contains experiments to confirm tumor suppression function of these candidate genes.

Part I, described three new methods including COP, CODE and CIS for cloning polymorphic, deleted and identical sequences between complex genomes. In this part, the pilot experiments have shown that these three methods were simple and robust methods in comparison with the existing methods and could achieve their aims successfully. Combination of COP and CIS methods has great potential application in isolation of disease genes responsible for hereditary diseases including cancer. CODE method applied to *Not*I flanking sequences, *Not*I representations (NRs) could directly isolate candidate genes from tumor samples.

Part II, described the applications of *Not*I linking, jumping clones and *Not*I microarrays to isolation of TSGs. All these experiments already started when I joined the group. The 630 kb (LUCA) minimal breast, lung cancer homozygous deletion region on 3p21.3 was initially mapped by *Not*I linking clone NL1-210, then covered by cosmid clones and fully sequenced. Finally 19 critical candidate genes were isolated. The *Not*I clones and *Not*I microarrays from the total human genome will provide us more opportunities to isolate candidate TSGs in different kinds of tumors.

Part III, described the gene inactivation test as a functional test for identification of TSGs. This test is based on the hypothesis that TSG will be functionally inactivated in tumors grown in SCID mice. The known TSGs *RB* and *p53* were tested in GIT. In these experiments, the two wild-type genes were inactivated, while the mutant forms of these two genes functionally maintained. Twelve different tumor cell lines were generated for checking different candidate TSGs in GIT. The candidate TSGs isolated in part I and part II especially those 19 genes in LUCA region were tested in GIT to functionally identify as TSGs. The one of 19 genes, *RASSF1A* gene was shown to have growth suppression both *in vivo* and *in vitro*, and inactivation in GIT. Following the GIT results, we also found hypermethylation of the *RASSF1A* promoter region in near all tested RCC cell lines and aproximately 90% of RCC tumor biopsies and this hypermethylation resulted the silencing of gene expression.

RESULTS AND DISCUSSION

1. Cloning identical and different sequences from complex genome (paper I-III)

Cloning of polymorphisms (COP) is a new procedure for enrichment of single nucleotide polymorphisms (SNPs) that represent restriction fragment length polymorphisms (RFLPs) between two complex genomes (**paper I**). In this paper, DNA A and B from two unrelative individuals were applied to COP procedure. DNA was isolated from ten random clones and eight clones displayed clear polymorphism. The results also showed that PCR could be used to detect polymorphism.

We anticipate that COP will be used predominantly as a method complementary to other SNP detection protocols. For example, sequencing of the whole human genome remains the most practical method to detect the most frequent polymorphisms in the human genome. Similarly, detection of SNPs for specific genes should involve direct sequencing of genes from various individuals. We suggest using COP for isolation of the SNPs from certain regions of the genome. In this case, DNA B would originate from contigs or individual YAC, PAC, or BAC clones from the region of interest. Another obvious applications for this method would be detection of SNPs in CpG islands (Bird 1987) (with CG-containing restriction enzymes) and in different human populations. This method could be used for chromosome landing' (Corrette-Bennett et al. 1998) ('chromosome landing' means isolation of region-specific high-density markers covering a region of interest without physical or genetic mapping) to facilitate positional cloning in organisms for which high-resolution maps are not available.

The important application of COP is very helpful to clone disease genes from the family members affected by a hereditary disease in combination with CIS procedure. COP final DNA with the affected members should not contain the disease gene, since the disease gene locates in the IBD region without any polymorphic sequences at all.

Cloning of deleted sequences (CODE) is a new subtractive hybridization technique and allows cloning homozygous deleted sequences between two complex genomes (**paper II**). In this paper, an ACC-LC5 lung cancer cell line containing a 0.7 Mb homozygous deletion and human lymphocyte DNA was applied to CODE procedure. Similar to COP paper, twenty-four random clones were tested by Southern blot hybridization. DNAs from five clones were deleted in the tumor cell line, three clones were failed to show a specific hybridization signal, and eleven clones were polymorphic because the compared normal lymphocyte DNA is not from the ACC-LC5 source, which was not available. The enrichment was high compared to RDA and other subtractive methods.

The advantages of the CODE procedure prompted us to develop its modification to clone the *Not*I deleted sequences (*Not*I-CODE). This modification was based on our experience in cloning of the differential products (Zabarovsky et al., 1993, 1994). We decided to use for subtraction only the regions surrounding the *Not*I sites. The main principles for this subtraction were the same as in the CODE procedure but genomic DNA was digested with *Not*I+*Bam*HI and other linkers were used to allow PCR amplification of only *Not*I-site containing fragments. Other DNA fragments were not amplified. Only two cycles of subtraction were used. To validate this approach, we again compared ACC-LC5 and normal human DNA. The final product was labelled and hybridized to the grids of *Not*I linking clones. Two *Not*I clones known to be homozygously deleted were identified. In addition, polymorphic and hemizygously deleted *Not*I clones were found. This means that the CODE procedure may be used with restriction enzymes, which contain CG in their recognition site. As these enzymes are associated with CpG islands and genes (Bird 1987; Allikmets et al.,

1994), the CODE procedure will result in direct cloning of the deleted genes. Obviously, the CODE procedure may be used to clone amplified sequences, but in this case tester and driver DNAs change their places in the experimental scheme.

Cloning of identical sequences (CIS) procedure was designed to clone identical sequences between two complex genomes (**paper III**). CIS procedure is very useful in cloning disease genes responsible for familial cancer syndrome by cloning IBD region, which should contain disease gene. In this paper, DNA from mouse-human microcell hybrid cell line MCH903.1 containing complete human chromosome 3 and rat MCH429.11 cell line containing a part human chromosome 3q from the same chromosome 3 as in MCH903.1 were applied to CIS procedure. The result showed that the original MCH429.11 and the DNA obtained using the CIS procedure had identical FISH (fluorescence in situ hybridization) patterns with human metaphase chromosomes. In the control experiment, MCH903.1 and MCH939.2 containing chromosome 3 del(p21-p22) from different individual were compared using the CIS procedure. Labeled DNA from MCH939.2 in reverse painting to normal human metaphase revealed, as expected, hybridization to chromosome 3 except 3p21-p22 regions. In contrast, the FISH revealed only background fluorescence with the CIS probe. This experiment has shown that CIS procedure discriminates between identical and nearly identical DNA fragments, since DNA from two homologous human chromosomes of different origin is 99.9% identical.

In the CIS procedure, the crucial step is efficiency of recognization and cleavage of mismatches in the heterohybrids by enzyme. In GMS method MutHLS are used in this step, still several drawbacks exist as we discussed before. We found that T4 endonuclease VII, which recognizes a broad spectrum of DNA substrates ranging from branched DNAs to single base mismatches (Youil et al. 1995; Babon et al. 1995) has great potential application in CIS procedure. We modified CIS procedure by treatment with T4 endonuclease VII before mung bean nuclease digestion. In order to optimize the conditions of modified-CIS procedure, we applied modified-CIS procedure to DNAs from the family affected by hereditary colon cancer, in which *MLH1* gene is responsible for the disease. BAC/PAC arrays which partially cover 8 Mb region around *MLH1* were generated on nylon filters. Modified-CIS probes from two pairs of affected members were hybridized on these minimal BAC/PAC arrays and as expected two BACs on the arrays containing *MLH1* gene gave hybridization signals. The combination of modified-CIS and COP procedure applied on family members affected by a hereditary syndrome should significantly improve the efficiency of screening for localization of hereditary disease genes (Fig. 14). We are currently applying modified-CIS procedure and COP procedure together to a panel of DNA pool from families suffering from familial nasopharyngeal carcinoma (NPC) in order to identify potential NPC susceptibility genes.

Although these methods (COP, CODE and CIS) are based on the same combination of biochemical techniques their aims are different. These methods are fully complementary; therefore they may be applied together to analyze a given object. If one aims to clone a disease gene responsible for familial cancer syndrome, these methods may be applied as follows. CIS may be used to identify the sequences identical by descent comparing the DNA obtained from affected family members. COP may be used to find sequences that are different between affected members, and CODE would be useful to compare tumor and normal (control) samples to isolate deleted sequences (putative candidate tumor suppressor genes) and amplified sequences (putative oncogenes). COP and CODE procedures may be applied to analyze the CpG islands thus allowing direct candidate gene identification.

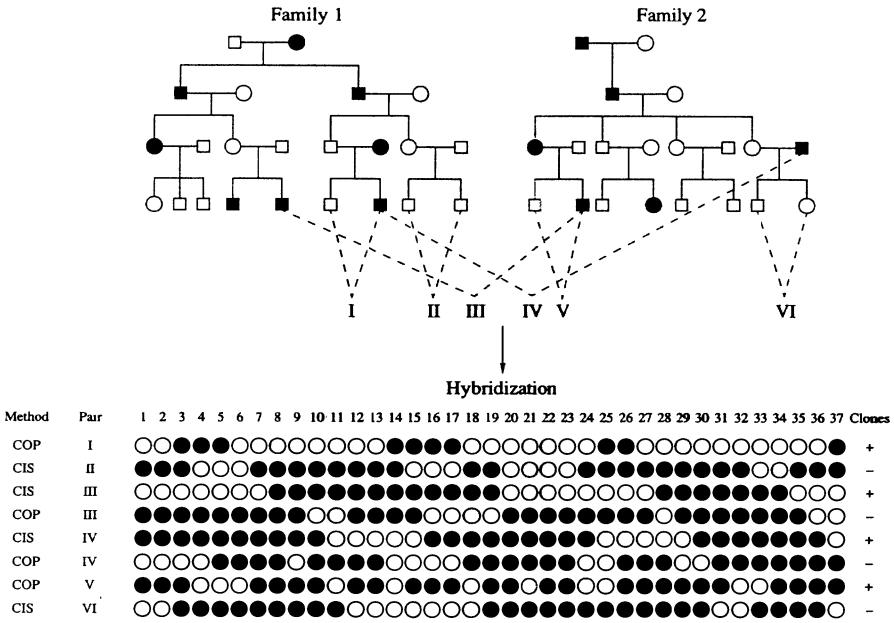

Fig. 14 Application of COP and CIS procedure for localization of a disease gene. Probes obtained by applying CIS and COP methods to pairs of the individuals from the disease-affected family members were hybridized with microarrayed clones covering the whole human genome (shown as circles at the bottom). Clones showing positive hybridization signal are in black. Different combinations of the CIS and COP methods allow detection of the genome regions either probably containing or probably not containing the disease gene (+ or – on the right). In this particular case, clones 16 and 17 have the hightest chance to contain the disease gene.

2. *Not*I linking and jumping libraries and *Not*I microarrays (paper IV-VI)

As described in 2.3, *Not*I linking and jumping clones, as CpG island markers, can be successfully used for joining physical and genetic mapping efforts.

Simplified new procedures for *Not*I jumping and linking library construction were developed and seven chromosome 3-specific *Not*I linking libraries were made. Direct statistical connection between CpG islands, *Not*I site and expressed sequences in the human genome has been experimentally confirmed. Among 150 unique *Not*I linking clones homology (90-100%) to more than 100 different genes or ESTs was found (Kashuba et al., 1999). These genes represent different classes including oncogenes and structural proteins. Several methods were used to localize *Not*I linking clones, PCR, Southern hybridization and two- or three- color FISH. The constructed *Not*I map in several chromosome 3 regions (Kashuba et al., 1995) was shown to be more precise than the physical chromosome 3 map which was based on YAC clones and radiation hybrids (Gemmill et al., 1995).

Our and other studies (Yamakawa et al., 1993; Alimov et al., 2000; Braga et al., 2002; **paper VI**) identified two most frequently rearranged regions in human chromosome 3, namely 3p21.3T (telomeric) or AP20 region and 3p21.3C (centromeric) or LUCA region.

AP20 region was found to be heavily methylated in all studied eleven RCC cell lines and homozygously deleted in several cancers (Zabarovsky et al., unpublished data). This suggests that inactivation by hypermethylation or deletion of TSG(s) in this region may play critical

role in the cancergenesis. Further functional and methylation analysis of candidate genes is necessary since no gene isolated by incorrect map showing any evidence related to cancergenesis. Right now these genes are under investigation in GIT.

LUCA region was firstly discovered by *Not*I linking clone NL1-210 that contains human MAPKAP kinase 3 (*3PK*). Based on this clone, 600 kb cosmid clone contig was generated and 19 genes were found within this overlap region. This region was further subdivided by a nesting 200 kb breast cancer homozygous deletion into two gene sets: 8 genes lying in the proximal ~120 kb segment and 11 genes lying in the distal ~250 kb segment (**paper VI**). The LUCA region was completely sequenced and the International Lung Tumor Suppressor Gene Consortium identified several candidate genes as a result of the major effort. These 19 genes were analyzed by loss of expression and tested for mutations in lung cancer to identify candidate TSGs within this region. Four genes (*CACNA2D2, SEMA3B, BLU* and *HYAL1*) showed loss-of-expression or reduced mRNA levels in non-small cell lung cancer or small cell lung cancer cell lines. Six genes including *BLU, Gene21, FUS1, HYAL1, FUS2*, and *SEMA3B* were found to have two or more amino acid sequence-altering mutations. However, none of the 19 genes tested for mutation showed a frequent mutation rate (>10%) in lung cancer sample. This suggested that the putative lung cancer TSG in this location may either be inactivated by tumor-acquired promoter hypermethylation or belong to the novel class of haploinsufficient genes that predispose to cancer in a hemizygous (+/-) state but do not show a second mutation in the remaining wild-type allele in the tumor. Recently, TSG *RASSF1A* gene in this region was found to be inactivated by hypermethylation at very high percentage in lung, kidney, prostate, nasopharyngeal, bladder and other cancers (refer to 1.3.4). *FUS1* gene was found to function as a TSG in haploisufficient manner (refer to 1.3.4). These results implied that inactivation of the putative TSG(s) in LUCA region might be caused not only by deletion but also by its inactivation through hypermethylation, haploinsufficience and other unknown mechanisms. Further functional testing of the critical genes in this region by gene transfer and gene knock out strategies should permit the identification of the putative lung cancer TSG(s). Now almost all genes are testing in our GIT system.

Numerous linking libraries with different restriction enzymes were constructed in attempt to generate representative *Not*I linking libraries, covering the whole human genome. High-density grids containing 50,000 clones originating from six representative *Not*I linking libraries were constructed (**paper IV**).

It was noted that our 22,551 unique *Not*I clones, covering the whole human genome and containing ~20% of all genes (40-50% of them are not present in ESTs microarrays) are already available. Construction of *Not*I microarrays for genome-wide screening is reasonable. *Not*I microarrays will be used for two purposes. The first aim is very similar to the ESTs microarrays, namely expression-profile experiments. The advantage here is that our arrays will contain different and unique sets of genes and many of them are not included in ESTs microarrays which means *Not*I microarrays are complementary to ESTs microarrays. The second aim is to use *Not*I microarrays for testing tumor genomic DNA in deletion, methylation and amplification studies. Such microarrays will speed up cancer research very significantly and can replace CGH, LOH and other cytogenetic studies.

In the pilot experiments, the chromosome 3-specific *Not*I microarrays were successfully used to detect the *Not*I clones deleted in small lung cancer cell line ACC-LC5 and mouse-human microcell human hybrid line MCH939.2 (**paper V**). *Not*I representations (NRs) were used to label sample DNAs with either radioactivity or fluorescent dyes to dramatically reduce the complexity of human genome. Two *Not*I clones NLJ-003 and NL1-401 from the homozygously deleted region of ACC-LC5 were easily detected.

As described in previous part, CODE procedure could improve the subtractive enrichment and could be applied for restriction enzymes containing CG in their recognition sites. The *Not*I-CODE technique reduces complexity of the genome by using only short regions surrounding *Not*I sites (NRs) for subtraction. The hybridization clearly identified the two regions most

frequently deleted in RCC, AP20 region and LUCA region. The impurity problem that occurs with tumor biopsies can be easily solved with *Not*I-CODE procedure and *Not*I microarrays.

A novel human retinoblastoma binding protein 1 homologous gene was isolated by *Not*I-CODE. This new gene was found loss or reduced expression in lung cancer cell lines and RCC cell lines (Li et al. manuscript). The function of this gene involving of tumorigenesis will be further studied.

More than 10 pairs of RCC biopsies are under checking using *Not*I microarrays. Chromosome 3-specific *Not*I microarrays, which attach *Not*I oligos on the slides, are available now from BD Bioscience CLONTECH.

3. GIT and candidate TSGs in 3p21 region (paper VII-IX)

Candidate genes, for example the 19 genes in LUCA region, isolated by the developed new methods should be further studied to confirm to be TSGs. In our group, gene inactivation test (GIT), which is based on the hypothesis that transferred TSG in tumor cell line will be inactivated in tumor (**paper VII** and **paper VIII**).

To prove our hypothesis, wild-type tumor suppressor *RB* gene and deleted mutant *RB* gene were tested in GIT. In the first experiment, wild-type *RB* gene and mutant *RB* gene were cloned in tetracycline vector, pETI and transfected into T711 cell line. Expression of both genes was regulated by tetracycline. The wild-type *RB* gene was deleted or functionally inactivated after the first passage in all 20 tumors tested. In contrast, a non-functional mutant *RB* gene was maintained in all 10 tumors studied. The growth of *RB* transfected tumors was delayed when tetracycline was absent (gene expression switched on), while mutant *RB* transfected tumors were not (**paper VII**). These findings indicated that TSGs might be deleted or inactivated when transfected tumor cells are passaged through SCID mice. In order to strengthen more evidences to prove our hypothesis, GIT of wild-type tumor suppressor *p53* gene and mutant *p53* gene were performed. Both were transfected into tTA producing osteosarcoma cell line Saos2 and T711 cell line. The expression of the genes was tightly controlled by tetracycline. The results were ¨the same as in RB experiment. As expected; wild-type *p53* gene was deleted in 6 tumors and mutant *p53* gene was functionally maintained in 12 tumors even after second passage through SCID mice. Wild-type *p53* gene was inactivated by point mutations in 4 tumors (Li et al. manuscript).

The candidate genes located in LUCA and AP20 regions are investigated by GIT (Table 12, unfinished).

Table 12. Summary table of the candidate genes in GIT

Gene	Cell line	In vitro growth (+/- tet)		In vivo (SCID)	Comments
		Colony	Petri dish	S/NS	
RB	A9	ND	ND	S	RB gene was deleted in all tumors
	KRC/Y	ND	ND	S	
	Saos-2	ND	ND	ND	
Delta RB	A9	ND	ND	NS	Not inactivated
Wp53	Saos-2	ND	ND	No tumor	
	A9	ND	ND	NS	Mutated in tumors
Wp53+Mp53	Saos-2	ND	ND	S	Wp53 deleted, while mp53 not

Mp53	Saos-2	ND	ND	NS	Not deleted
PL6	U2020	ND	NS	NS	Not deleted
101F6	KRC/Y	ND	ND	ND	
	HeLa	ND	ND	ND	
RASSF1A	KRC/Y	S	S	S	Loss of expression in tumors
	U2020	ND	S	ND	
	LNCaP	ND	S	ND	
mRASSF1A	KRC/Y	NS	NS	ND	
	U2020	ND	NS	ND	
	LNCaP	ND	NS	ND	
SEMA3B	U2020	S	S	S	Deleted
	KRC/Y	ND	ND	ND	
FUS2	U2020	ND	NS	NS	Not deleted
FUS1	U2020	ND	NS	NS	Not deleted
Blu	U2020	ND	NS	NS	Not deleted
HYAL1	U2020	ND	NS	No tumor	
	KRC/Y	ND	NS	ND	
HYAL2	U2020	ND	NS	No tumor	
	KRC/Y	ND	ND	ND	
	HeLa	ND	ND	ND	
APRG1b	ACC-LC5	ND	S	ND	
APRG1c	ACC-LC5	ND	NS	ND	
3PK	KRC/Y	ND	NS	NS	Not deleted
	RHEK-1	ND	ND	ND	
TGFBRII	KRC/Y	ND	ND	NS	Not deleted
VHL	KRC/Y	ND	ND	S	Deleted
	HeLa	ND	ND	ND	
RhoA	KRC/Y	ND	ND	NS	Not deleted
MLH1	KRC/Y	ND	ND	NS	Not deleted

RASSF1A, a new RAS effector gene, was located in the LUCA region. The RASSF1A protein interacts with the DNA repair protein, XPA (Dammann et al., 2000). As seen with the mouse Ras effector protein Nore1 and its rat ortholog MAXp1, the human RASSF1A isoform has high homology to the cysteine-rich/phorbolester-binding domain, also known as the protein kinase C conserved region 1. Nore1 interaction with Ras is GTP-dependent and follows receptor activation. *RASSF1A* mRNA was missing in most analyzed RCC an lung cancer cell lines because of methylation of the *RASSF1A* putative promoter region (Dammann et al., 2000; Burbee et al., 2001; Agathanggelou et al., 2001, **paper IX**). In 9 RCC cell lines, hypermethylation of *RASSF1A* was detected by COBRA analysis and confirmed by

sequencing of bisulfite-treated amplified promoter DNA. Hypermethylation of the *RASSF1A* promoter region correlated with transcriptional down-regulation of *RASSF1A* mRNA. *RASSF1A* gene was reexpressed in heavily methylated RCC cell line 786-0 after 5-Aza-CdR treatment. The *RASSF1A* gene was also found to be hypermethylated in 91% of primary clear RCC tumors (39 of 43 tumors). All data suggest that the *RASSF1A* is a renal 3p21.3 TSG. More studies are necessary to elucidate the role of *RASSF1A* in cell biology and in other types of human cancer.

RASSF1C gene is another major alternative transcript of *RASSF1A* with distinct GC-rich promoter. Methylation of *RASSF1C* promoter was not found in RCC cell lines and primary renal tumors. Interestingly, the *RASSF1C* has been reported to be silenced in some ovarian tumor samples and might serve as a novel RAS effector that mediated the apoptotic effects of oncogenic Ras (Vos et al., 2000). Our GIT results also showed that *RASSF1C* also has suppression of tumorigenesis activity in KRC/Y cell line. The function of RASSF1C should be further carefully analyzed.

Inactivation and suppression of other genes such as *HYA22*, *Gene21* and *SEMA3B* in the GIT results are also interesting. These genes are under further investigation.

ACKNOWLEDGEMENTS

This study was carried out at the Microbiology and Tumor Biology Center (MTC), Center for Genomic Bioinformatics (CGB), Karolinska Institute, Stockholm, Sweden. I wish to express my sincere gratitude to all my colleagues and friends who have contributed to the completion of this thesis in different ways. I am grateful to all people who encouraged my scientific career in the past years, but I could not mention all of them here. I especially acknowledge:

Zhenya Zabarovsky, my supervisor, for giving me the opportunity to study at the Karolinska Institute, for his stimulating discussions, for his scientific guidance and enthusiastic encouragement, for his support, understanding and friendship. I could not defend my thesis without your detail helps.

George Klein, for supporting me in the beginning, for forcing to write the regular Friday letters and monthly meeting and discussion, and for all the comments for the papers.

Ingemar Ernberg, for his critical examination that enables me to plan further to defend my thesis, for creating a wonderful study and research atmosphere in the big shared laboratory and in MTC.

Magnus Nordenskjöld, for kindly being my opponent.

Gösta Winberg, for kindly being my mentor.

Michael Lerman, for our collaboration with great pleasure.

Veronika Zabarovska, for your help both on science and life, and for organization of every summer party in your beautiful house.

Vladimir Kashuba, for your help and fantastic discussions in the field of molecular biology.

Alexei Protopopov, for your help in cell culture work and discussions about our collaborated work, especially for your computer assistance.

Rinat Gizatullin, Andre Alimov, Fuli Wang, Olgo Vorontsova, Ali Al-Amin, Li Xie, Klas Haraldson, for your technique help and sharing the experimental equipments and reagents in the group. The visiting scientists **Igor Bazov, Gelena Kilosanidze, Alexey Kutsenko, Alena Malyukova, Olga Muravenko, Eleonora Braga and Vera Sentchenko.**

Endre Kobold, Angelo De Milito, Ami Albin, Bartek Zuber, Karl Ljunberg, Mauro D'Amato, Alberto Cagigi, Daniel Palm, Emma Eriksson and Richard Scuderi, for football enjoyment and winning KI football championship.

Alexandru Feldötö, Greger Blomquist, Margareta Hagelin, Helena Lönnqvist, for technique support**.**

I would like to thank all the following for the help: **Claes Wahlestedt**, **Annika Lindblom, Klas Wiman, Håkan Axelson, Anna Szeles, Hajnalka Kiss, Csaba Kiss, Irina Kholodnyuk, Stefan Imreh, Laszlo Szekely, Elena Kashuba, Ebba Bråkenhielm.**

All Chinese friends: **Ying Yang, Lifu Hu, Fu Chen, Qian Wang, Yintong Xue, Jiezhi Zou, Xiangning Zhang, Liying Chen, Xiongbiao Wang, Yue Chen, Qianhui Qiu, Xiaoming Huang, Hai Huang, Hong Xu, Lei Lu and Di Sun.**

Xiaopeng Li, my oldest brother, for introducing me to Karolinska Institute and always the guidance in my scientific career and my life. **Weidong Li**, my older brother, for supporting me through any ways, for helping me to take care of our parents.

Wenbin Li and **Caicun Tao**, my dear father and mother, for unconditional supporting me, for your endless encouragement, for your deep love.

Xiaomin Guo, my wife, for your love, support, and understanding which are so impotant for me to finish this thesis, and so important for my life.

REFERENCES

Aach J., Rindone W. & Church G.M. (2000) Systematic management and analysis of yeast gene expression data. *Genome Res.*, **10**: 431-445.

Adany R. & Iozzo R.V. (1990) Altered methylation of versican proteoglycan gene in human colon carcinoma. *Biochem Biophys Res Commun.*, **171**: 1402-1413.

Adorjan P., Distler J., Lipscher E., Model F., Muller J., Pelet C., Braun A., Florl A.R., Gutig D., Grabs G., Howe A., Kursar M., Lesche R., Leu E., Lewin A., Maier S., Muller V., Otto T., Scholz C., Schulz W.A., Seifert H.H., Schwope I., Ziebarth H., Berlin K., Piepenbrock C. & Olek A. (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.*, **30**: e21.

Agathanggelou A, Honorio S, Macartney DP, Martinez A, Dallol A, Rader J, Fullwood P, Chauhan A, Walker R, Shaw JA, Hosoe S, Lerman MI, Minna JD, Maher ER and Latif F. (2001). Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. *Oncogene*, **20**, 1509-1518.

Aggerholm A. & Hokland P. (2000) DAP-kinase CpG island methylation in acute myeloid leukemia: methodology versus biology? *Blood*, **95**: 2997-2999.

Aissani B & Bernardi G. (1991) CpG islands: features and distribution in the genomes of vertebrates. *Gene* **106**:173-183.

Akama T.O., Okazaki Y., Ito M., Okuizumi H., Konno H., Muramatsu M., Plass C., Held W.A. & Hayashizaki Y. (1997) Restriction landmark genomic scanning (RLGS-M)-based genome-wide scanning of mouse liver tumors for alterations in DNA methylation status. *Cancer Res.*, **57**: 3294-3299.

Albertson D.G., Ylstra B., Segraves R., Collins C., Dairkee S.H., Kowbel D., Kuo W.L., Gray J.W. & Pinkel D. (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat Genet.*, **25**: 144-146.

Allikmets R., Kashuba V.I., Huebner K., LaForgia S., Kisselev L.L., Klein G., Dean M. & Zabarovsky E.R. (1996) Mapping of 22 NotI linking clones on human chromosome 3 by polymerase chain reaction and somatic cell hybrid panels. *Chromosome Res.*, **4**: 33-37.

Allikmets R.L., Kashuba V.I., Pettersson B., Gizatullin R., Lebedeva T., Kholodnyuk I.D., Bannikov V.M., Petrov N., Zakharyev V.M., Winberg G., et al. (1994) NotI linking clones as a tool for joining physical and genetic maps of the human genome. *Genomics*, **19**: 303-309.

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *J Mol Biol.*, **215**: 403-410.

Alves G., Tatro A. & Fanning T. (1996) Differential methylation of human LINE-1 retrotransposons in malignant cells. *Gene*, **176**: 39-44.

Antequera F., Boyes J. & Bird A. (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell*, **62**: 503-514.

Antequera F. & Bird A. (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci USA*, **90**: 11995-11999.

Arapshian A., Kuppumbatti Y.S. & Mira-y-Lopez R. (2000) Methylation of conserved CpG sites neighboring the beta retinoic acid response element may mediate retinoic acid receptor beta gene silencing in MCF-7 breast cancer cells. *Oncogene*, **19**: 4066-4070.

Arenstorf H.P., Kandpal R.P., Baskaran N., Parimoo S., Tanaka Y., Kitajima S., Yasukochi Y. & Weissman S.M. (1991) Construction and characterization of a NotI-BsuE linking library from the human X chromosome. *Genomics*, **11**: 115-123.

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S.,

Matese J.C., Richardson J.E., Ringwald M., Rubin GM. & Sherlock G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.*, **25**: 25-29.

Au K.G., Welsh K. & Modrich P. (1992) Initiation of methyl-directed mismatch repair. *J Biol Chem.*, **267**: 12142-12148.

Babon J., Youil R. & Cotton R.G.H. (1995) Improved strategy for mutation detection – a modification to the mismatch cleavage method. *Nucleic Acids Res.*, **23**: 5082-5084.

Bachman K.E., Herman J.G., Corn P.G., Merlo A., Costello J.F., Cavenee W.K., Baylin S.B. & Graff J.R. (1999) Methylation-associated silencing of the tissue inhibitor of metalloproteinase-3 gene suggest a suppressor role in kidney, brain, and other human cancers. *Cancer Res.*, **59**: 798-802.

Bader S.A., Fasching C., Brodeur G.M. & Stanbridge E.J. (1991) Dissociation of suppression of tumorigenicity and differentiation in vitro effected by transfer of single human chromosomes into human neuroblastoma cells. *Cell Growth Differ.*, **2**: 245-255.

Balaghi M. & Wagner C. (1993) DNA methylation in folate deficiency: use of CpG methylase. *Biochem Biophys Res Commun.*, **193**: 1184-1190.

Banerjee A., Xu H.J., Hu S.X., Araujo D., Takahashi R., Stanbridge E.J. & Benedict W.F. (1992) Changes in growth and tumorigenicity following reconstitution of retinoblastoma gene function in various human cancer cell types by microcell transfer of chromosome 13. *Cancer Res.*, **52**: 6297-6304.

Bannister A.J., Zegerman P., Partridge J.F., Miska E.A., Thomas J.O., Allshire R.C. & Kouzarides T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature,* **410**: 120-124.

Bassett D.E. Jr., Eisen M.B. & Boguski M.S. (1999) Gene expression informatics--it's all in your mine. *Nat Genet.*, **21**: 51-55.

Batova A., Diccianni M.B., Yu J.C., Nobori T., Link M.P., Pullen J. & Yu A.L. (1997) Frequent and selective methylation of p15 and deletion of both p15 and p16 in T-cell acute lymphoblastic leukemia. *Cancer Res.*, **57**: 832-836.

Baylin S.B., Hoppener J.W., de Bustros A., Steenbergh P.H., Lips C.J. & Nelkin B.D. (1986) DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. *Cancer Res.*, **46**: 2917-2922.

Baylin S.B., Herman J.G., Graff J.R., Vertino P.M. & Issa J.P. (1998) Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv Cancer Res.*, **72**: 141-196.

Beckman J.S. & Weber J.L. (1992) Survey of human and rat microsatellites. *Genomics,* **12**: 627-631.

Belinsky S.A. (1998) Role of the cytosine DNA-methyltransferase and p16INK4a genes in the development of mouse lung tumors. *Exp Lung Res.*, **24**: 463-479.

Berard J., Laboune F., Mukuna M., Masse S., Kothary R. & Bradley W.E. (1996) Lung tumors in mice expressing an antisense RARbeta2 transgene. *FASEB J.*, **10:** 1091-1097.

Bestor T., Laudano A., Mattaliano R. & Ingram V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol.*, **203**: 971-983.

Bestor TH. (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J.*, **11**: 2611-2617.

Bird A. (1987) CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.*, **3**: 342-347.

Bird A.P., Taggart M., Frommer M., Miller Q.J. & Macleod D. (1985) A fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. *Cell,* **40**: 91-99.

Bouchard J. & Momparler R.L. (1983) Incorporation of 5-Aza-2'-deoxycytidine 5'-triphosphate into DNA: interactions with mammalian DNA polymerase and DNA methylase. *Mol Pharmacol.*, **24**: 109-114.

Braga E., Pugacheva E., Bazov I., Ermilova V., Kazubskaya T., Mazurenko N., Kisseljov F., Liu J., Garkavtseva R., Zabarovsky E. & Kisselev L. (1999) Comparative allelotyping of the short arm of human chromosome 3 in epithelial tumors of four different types. *FEBS Lett.,* **454**: 215-219.

Braga E., Sentchenko V., Bazov I., Loginov W., Liu J., Ermilova V., Kazubskaya T., Garkavtseva R., Mazurenko N., Kisseljov F., Lerman M.I., Kisselev L. & Zabarovsky E.R. (2002) Critical tumor-suppressor gene regions on chromosome 3P in major human epithelial malignancies: allelotyping and quantitative real-time PCR. *Int J Cancer,* **100**: 534-541.

Brandeis M., Frank D., Keshet I., Siegfried Z., Mendelsohn M., Nemes A., Temper V., Razin A. & Cedar H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature,* 371: 435-438.

Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J. & Vingron M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet.,* **29**: 365-371.

Brookes A.J. (1999) The essence of SNPs. *Gene,* **234**: 177-186.

Brown W.R.A. & Bird A.P. (1986) Long-range restriction site mapping of mammalian genomic DNA. *Nature,* **322**: 477-481.

Bruder C.E., Hirvela C., Tapia-Paez I., Fransson I., Segraves R., Hamilton G., Zhang X.X., Evans D.G., Wallace A.J., Baser M.E., Zucman-Rossi J., Hergersberg M., Boltshauser E., Papi L., Rouleau G.A., Poptodorov G., Jordanova A., Rask-Andersen H., Kluwe L., Mautner V., Sainio M., Hung G., Mathiesen T., Moller C., Pulst S.M., Harder H., Heiberg A., Honda M., Niimura M., Sahlen S., Blennow E., Albertson D.G., Pinkel D. & Dumanski J.P. (2001) High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum Mol Genet.,* **10**: 271-282.

Burbee D.G., Forgacs E., Zochbauer-Muller S., Shivakumar L., Fong K., Gao B., Randle D., Kondo M., Virmani A., Bader S., Sekido Y., Latif F., Milchgrub S., Toyooka S., Gazdar A.F., Lerman M.I. & Zabarovsky E.R., White M. & Minna J.D. (2001) Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. *J Natl Cancer Inst.*, **93:** 691-699.

Bussow K., Konthur Z., Lueking A., Lehrach H. & Walter G. (2001) Protein array technology. Potential use in medical diagnostics. *Am J Pharmacogenomics,* **1**: 37-43.

Cai W.W., Mao J.H., Chow C.W., Damani S., Balmain A. & Bradley A. (2002) Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays. *Nat Biotechnol,* **20**: 393-396.

Cairns P., Okami K., Halachmi S., Halachmi N., Esteller M., Herman J.G., Jen J., Isaacs W.B., Bova G.S. & Sidransky D. (1997) Frequent inactivation of PTEN/MMAC1 in primary prostate cancer. *Cancer Res.,* **57**: 4997-5000.

Cameron E.E., Bachman K.E., Myohanen S., Herman J.G. & Baylin S.B. (1999) Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet.*, 21: 103-107.

Chen Y., Dougherty E.R. & Bittner M.L. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J Biomed Opt.,* **2**: 364-374.

Cheng Y., Poulos N.E., Lung M.L., Hampton G., Ou B., Lerman M.I. & Stanbridge E.J. (1998) Functional evidence for a nasopharyngeal carcinoma tumor suppressor gene that maps at chromosome 3p21.3. *Proc Natl Acad Sci U S A*, **95**: 3042-3047.

Cheng Y., Stanbridge E.J., Kong H., Bengtsson U., Lerman M.I. & Lung M.L. (2000) A functional investigation of tumor suppressor gene activities in a nasopharyngeal carcinoma cell line HONE1 using a monochromosome transfer approach. *Genes Chrom Cancer,* **28**: 82-91.

Cheung V.G. & Nelson S.F. (1998) Genomic mismatch scanning identifies human genomic DNA shared identical by descent. *Genomics,* **47**: 1-6.

Cheung V.G., Gregg J.P., Gogolin-Ewens K.J., Bandong J., Stanley C.A., Baker L., Higgins M.J., Nowak N.J., Shows T.B., Ewens W.J., Nelson S.F. & Spielman R.S. (1998) Linkage-disequilibrium mapping without genotyping. *Nat Genet.,* **18**: 225-230.

Cheung V.G., Morley M., Aguilar F., Massimi A., Kucherlapati R. & Childs G. (1999) Making and reading microarrays. *Nat Genet.,* **21**: 15-19.

Chuang L.S., Ian H.I., Koh T.W., Ng H.H., Xu G. & Li B.F. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science,* **277**: 1996-2000.

Clark S.J., Harrison J., Paul C.L. & Frommer M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.,* **22**: 2990-2997.

Collins F.S. & Weissman S.M. (1984) Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. *Proc Natl Acad Sci U S A,* **81**: 6812-6816.

Collins F.S. (1988) Chromosome jumping. In ''Genome Analysis: A Practical Approach'' (K.E. Davids, ED), pp. 73-94, IRL Press, Oxford, England.

Conway K.E., McConnell B.B., Bowring C.E., Donald C.D., Warren S.T. & Vertino P.M. (2000) TMS1, a novel proapoptotic caspase recruitment domain protein, is a target of methylation-induced gene silencing in human breast cancers. *Cancer Res.,* **60**: 6236-6242.

Cooper D., Taggart M. & Bird A. (1983) Unmethylated domains in vertebrate DNA. *Nucleic Acids Res.,* **11**: 647-658.

Corn P.G., Kuerbitz S.J., van Noesel M.M., Esteller M., Compitello N., Baylin S.B. & Herman J.G. (1999) Transcriptional silencing of the p73 gene in acute lymphoblastic leukemia and Burkitt's lymphoma is associated with 5' CpG island methylation. *Cancer Res.,* **59**: 3352-3356.

Corrette-Bennett J., Rosenberg M., Przybylska M., Ananiev E. & Straus D. (1998) Positional cloning without a genome map: using 'Targeted RFLP Subtraction' to isolate dense markers tightly linked to the regA locus of Volvox carteri. *Nucleic Acids Res.,* **26**: 1812-1818.

Costello J.F., Berger M.S., Huang H.S. & Cavenee W.K. (1996) Silencing of p16/CDKN2 expression in human gliomas by methylation and chromatin condensation. *Cancer Res.,* **56**: 2405-2410.

Costello J.F., Fruhwald M.C., Smiraglia D.J., Rush L.J., Robertson G.P., Gao X., Wright F.A., Feramisco J.D., Peltomaki P., Lang J.C., Schuller D.E., Yu L., Bloomfield C.D., Caligiuri M.A., Yates A., Nishikawa R., Su Huang H., Petrelli N.J., Zhang X., O'Dorisio M.S., Held W.A., Cavenee W.K. & Plass C. (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet.,* **24**: 132-138.

Costello J.F. & Plass C. (2001) Methylation matters. *J Med Genet.,* **38**: 285-303.

Craig J.M. & Bickmore W.A. (1994) The distribution of CpG islands in mammaliam chromosomes. *Nature Genetics,* **7**: 376-382.

Craig J.M., Kraus J. & Cremer T. (1997) Removal of repetitive sequences from FISH probes using PCR-assisted affinity chromatography. *Hum Genet.,* **100**: 472-476.

Cross S.H., Lee M., Clark V.H., Craig J.M., Bird A.P. & Bickmore W.A. (1997) The chromosomal distribution of CpG islands in the mouse: evidence for genome scrambling in the rodent lineage. *Genomics,* **40**: 454-461.

Cross S.H., Charlton J.A., Nan X. & Bird A.P. (1994) Purification of CpG islands using a methylated DNA binding column. *Nat Genet.,* **6**: 236-244.

Cross S.H. & Bird A.P. (1995) CpG islands and genes. *Curr Opin Genet Dev.,* **5**: 309-314.

Cuthbert A.P., Bond J., Trott D.A., Gill S., Broni J., Marriott A., Khoudoli G., Parkinson E.K., Cooper C.S. & Newbold R.F. (1999) Telomerase repressor sequences on chromosome 3 and induction of permanent growth arrest in human breast cancer cells. *J Natl Cancer Inst.,* **91**: 37-45.

Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. (2000) Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. *Nat Genet* 25(3), 315-319.

Del Senno L., Maestri I., Piva R., Hanau S., Reggiani A., Romano A. & Russo G. (1989) Differential hypomethylation of the c-myc protooncogene in bladder cancers at different stages and grades. *J Urol.,* **142**: 146-149.

Denissenko M., Chen J.X., Tang M.S. & Pfeifer G.P. (1997) Cytosine methylation determines hot spots of DNAb damage in the human p53 gene. *Proc Natl Acad Sci USA,* **84**: 3893-3898.

Deng G., Chen A., Hong J., Chae H.S. & Kim Y.S. (1999) Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. *Cancer Res.,* **59**: 2029-2033.

DeRisi J., Penland L., Brown P.O., Bittner M.L., Meltzer P.S., Ray M., Chen Y., Su Y.A. & Trent J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet.,* **14**: 457-460.

Dessain S.K., Yu H., Reddel R.R., Beijersbergen R.L. & Weinberg R.A. (2000) Methylation of the human telomerase gene CpG island. *Cancer Res.,* **60**: 537-541.

Devereux T.R., Horikawa I., Anna C.H., Annab L.A., Afshari C.A. & Barrett J.C. (1999) DNA methylation analysis of the promoter region of the human telomerase reverse transcriptase (hTERT) gene. *Cancer Res.,* **59**: 6087-6090.

Diatchenko L., Lau Y.F., Campbell A.P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Lukyanov K., Gurskaya N., Sverdlov E.D. & Siebert P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A,* **93**: 6025-6030.

Diatchenko L., Lukyanov S., Lau Y.F. & Siebert P.D. (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Methods Enzymol.,* **303**: 349-380.

Dodge J.E., List A.F. & Futscher B.W. (1998) Selective variegated methylation of the p15 CpG island in acute myeloid leukemia. *Int J Cancer,* **78**: 561-567.

Doerfler W. (1983) DNA methylation and gene activity. *Annu Rev Biochem.,* **52**: 93-124.

Domann F.E., Rice J.C., Hendrix M.J. & Futscher B.W. (2000) Epigenetic silencing of maspin gene expression in human breast cancers. *Int J Cancer,* **85**: 805-810.

Dowdy S.F., Fasching C.L., Araujo D., Lai K.M., Livanos E., Weissman B.E. & Stanbridge E.J. (1991) Suppression of tumorigenicity in Wilms tumor by the p15.5-p14 region of chromosome 11. *Science,* **254**: 293-295.

Drobyshev A., Mologina N., Shik V., Pobedimskaya D., Yershov G. & Mirzabekov A. (1997) Sequence analysis by hybridization with oligonucleotide microchip: identification of beta-thalassemia mutations. *Gene,* **188**: 45-52.

Eisen M.B., Spellman P.T., Brown P.O. & Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA,* **95**: 14863-14868.

Elenbaas B., Spirio L., Koerner F., Fleming M.D., Zimonjic D.B., Donaher J.L., Popescu N.C., Hahn W.C. & Weinberg R.A. (2001) Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *GENES & DEVELOPMENT,* **15**: 50-65.

Ellis L.A., Taylor G.R., Banks R. & Baumberg S. (1994) MutS binding protects heteroduplex DNA from exonuclease digestion in vitro: a simple method for detecting mutations. *Nucleic Acids Res.,* **22**: 2710-2711.

Ermolaeva O., Rastogi M., Pruitt K.D., Schuler G.D., Bittner M.L., Chen Y., Simon R., Meltzer P., Trent J.M. & Boguski M.S. (1998) Data management and analysis for gene expression arrays. *Nat Genet.,* **20**: 19-23.

Esteller M., Corn P.G., Urena J.M., Gabrielson E., Baylin S.B. & Herman J.G. (1998) Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. *Cancer Res.,* **58**: 4515-4518.

Esteller M., Avizienyte E., Corn P.G., Lothe R.A., Baylin S.B., Aaltonen L.A. & Herman J.G. (2000) Epigenetic inactivation of LKB1 in primary tumors associated with the Peutz-Jeghers syndrome. *Oncogene,* **19**: 164-168.

Esteller M., Tortola S., Toyota M., Capella G., Peinado M.A., Baylin S.B. & Herman J.G. (2000) Hypermethylation-associated inactivation of p14(ARF) is independent of p16(INK4a) methylation and p53 mutational status. *Cancer Res.,* **60**: 129-133.

Esteller M. & Herman J.G. (2002) Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J Pathol.,* **196**: 1-7.

Feinberg A.P. & Vogelstein B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal countparts. *Nature,* **301**: 89-92.

Feinberg A.P. (1999) Imprinting of a genomic domain of 11p15 and loss of imprinting in cancer: an introduction. *Cancer Res.,* **59**: 1743-1746.

Fodor S.P., Read J.L., Pirrung M.C., Stryer L., Lu A.T. & Solas D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science,* **251**: 767-773.

Frommer M., McDonald L.E., Millar D.S., Collis C.M., Watt F., Grigg G.W., Molloy P.L. & Paul C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89**: 1827-1831.

Fujii H., Biel M.A., Zhou W., Weitzman S.A., Baylin S.B. & Gabrielson E. (1998) Methylation of the HIC-1 candidate tumor suppressor gene in human breast cancer. *Oncogene,* **16**: 2159-2164.

Futreal P.A., Kasprzyk A., Birney E., Mullikin J.C., Wooster R. & Stratton M.R. (2001) Cancer and genomics. *Nature,* 409: 850-852.

Galm O., Rountree M.R., Bachman K.E., Jair K.W., Baylin S.B. & Herman J.G. (2002) Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density. *Genome Res.,* **12**: 153-157.

Gama-Sosa M.A., Wang R.Y., Kuo K.C., Gehrke C.W. & Ehrlich M. (1983a) The 5-methylcytosine content of highly repeated sequences in human DNA. *Nucleic Acids Res.,* **11**: 3087-3095.

Gama-Sosa M.A., Slagel V.A., Trewyn R.W., Oxenhandler R., Kuo K.C., Gehrke C.W. & Ehrlich M. (1983b) The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.,* **11**: 6883-6894.

Gardiner-Garden M. & Frommer M. (1987) CpG islands invertebrate genome. *J Mo. Biol.,* **196**: 261-282.

Gemmill R.M., Chumakov I., Scott P., Waggoner B., Rigault P., Cypser J., Chen Q., Weissenbach J., Gardiner K., Wang H., et al. (1995) A second-generation YAC contig map of human chromosome 3. *Nature,* **377**: 299-319.

Gilles P.N., Wu D.J., Foster C.B., Dillon P.J. & Chanock S.J. (1999) Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nat Biotechnol.,* **17**: 365-370.

Girard L., Zochbauer-Muller S., Virmani A.K., Gazdar A.F. & Minna J.D. (2000) Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. *Cancer Res.*, **60**: 4894-4906.

Gish W. & States D.J. (1993) Identification of protein coding regions by database similarity search. *Nat Genet.,* **3**: 266-272.

Gitan R.S., Shi H., Chen C.M., Yan P.S. & Huang T.H. (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.,* **12**: 158-164.

Gonzalgo M.L. & Jones P.A. (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.,* **25**: 2529-2531.

Gossen M. & Bujard H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A,* **89**: 5547-5551.

Goyette M.C., Cho K., Fasching C.L., Levy D.B., Kinzler K.W., Paraskeva C., Vogelstein B. & Stanbridge E.J. (1992) Progression of colorectal cancer is associated with multiple tumor suppressor gene defects but inhibition of tumorigenicity is accomplished by correction of any single defect via chromosome transfer. *Mol Cell Biol.,* **12**: 1387-1395.

Greenblatt M.S., Bennett W.P., Hollstein M. & Harris C.C. (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.,* **54**: 4855-4878.

Greger V., Passarge E., Hopping W., Messmer E. & Horsthemke B. (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet.,* **83**: 155-158.

Gu Z., Hillier L. & Kwok P.Y. (1998) Single nucleotide polymorphism hunting in cyberspace. *Hum Mutat.,* **12**: 221-225.

Hacia J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet.,* **21**: 42-47.

Hahn W.C., Counter C.M., Lundberg A.S., Beijersbergen R.L., Brooks M.W. & Weinberg R.A. (1999) Creation of human tumour cells with defined genetic elements. *Nature,* **400**: 464-468.

Hannis J.C. & Muddiman D.C. (1999) Accurate characterization of the tyrosine hydroxylase allel 9.3 through development of electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom.,* **13**: 954-962.

Harris H. (1986) The genetic analysis of malignancy. *J. Cell Sci. Suppl.,* **4**: 431-444.

Harris H. (1988) The analysis of malignancy by cell fusion: the position in 1988. *Cancer Res.,* **48**: 3302-3306.

Harris L.C., Remack J.S. & Brent T.P. (1994) In vitro methylation of the human O6-methylguanine-DNA methyltransferase promoter reduces transcription. *Biochim Biophys Acta.,* **1217**: 141-146.

Harrison J., Stirzaker C. & Clark S.J. (1998) Cytosines adjacent to methylated CpG sites can be partially resistant to conversion in genomic bisulfite sequencing leading to methylation artifacts. *Anal Biochem.,* **264**: 129-132.

Hatada I., Hayashizaki Y., Hirotsune S., Komatsubara H. & Mukai T. (1991) A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc Natl Acad Sci U S A,* **88**: 9523-9527.

Hattori M., Toyoda A., Ichikawa H., Ito T., Ohgusu H., Oishi N., Kano T., Kuhara S., Ohki M. & Sakaki Y. (1993) Sequence-tagged NotI sites of human chromosome 21: sequence analysis and mapping. *Genomics,* **17**: 39-44.

Hauptmann S., Denkert C., Koch I., Petersen S., Schluns K., Reles A., Dietel M. & Petersen I. (2002) Genetic alterations in epithelial ovarian tumors analyzed by comparative genomic hybridization. *Hum Pathol.,* **33**: 632-641.

Hayashizaki Y., Hirotsune S., Okazaki Y., Hatada I., Shibata H., Kawai J., Hirose K., Watanabe S., Fushiki S., Wada S., et al. (1993) Restriction landmark genomic scanning method and its various applications. *Electrophoresis,* 14: 251-258.

Hayatsu H., Wataya Y., Kai K. & Iida S. (1970) Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry,* **9**: 2858-2865.

Heiskanen M.A., Bittner M.L., Chen Y., Khan J., Adler K.E., Trent J.M. & Meltzer P.S. (2000) Detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer Res.,* **60**: 799-802.

Herman J.G., Latif F., Weng Y., Lerman M.I., Zbar B., Liu S., Samid D., Duan D.S., Gnarra J.R., Linehan W.M., et al. (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A,* **91**: 9700-9704.

Herman J.G., Graff J.R., Myohanen S., Nelkin B.D. & Baylin S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A,* **93**: 9821-9826.

Herman J.G., Jen J., Merlo A. & Baylin S.B. (1996) Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B. *Cancer Res.,* **56**: 722-727.

Holliday R. & Pugh J.E. (1975) DNA modification mechanisma and gene activity during development. *Science,* **187**: 226-232.

Hsieh C.L. (1999) In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Mol Cell Biol.,* **19**: 8211-8218.

Huang T.H., Perry M.R. & Laux D.E. (1999) Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet.,* **8**: 459-470.

Huynh H.T., Alper L., Pollak M. (1996) Silencing of the mammary-derivedgrowth inhibitor (MDGI) in breast neoplasms is associated\with epigenetic changes. *Cancer Res.,* **56**: 4865-4870.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature,* **490**: 860-921.

Issa J.P. (2000) CpG-island methylation in aging and cancer. *Curr Top Microbiol Immunol.,* **249**: 101-118.

Ito T. & Sakaki Y. (1988) A novel procedure for selective cloning of NotI linking fragments from mammalian genomes. *Nucleic Acids Res.,* **16**: 9177-9184.

Iwata N., Yamamoto H., Sasaki S., Itoh F., Suzuki H., Kikuchi T., Kaneto H., Iku S., Ozeki I., Karino Y., Satoh T., Toyota J., Satoh M., Endo T. & Imai K. (2000) Frequent hypermethylation of CpG islands and loss of expression of the 14-3-3 sigma gene in human hepatocellular carcinoma. *Oncogene,* 19: 5298-5302.

Iyer V.R., Eisen M.B., Ross D.T., Schuler G., Moore T., Lee J.C., Trent J.M., Staudt L.M., Hudson J.Jr., Boguski M.S., Lashkari D., Shalon D., Botstein D. & Brown P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science,* **283**: 83-87.

Jahner D., Stuhlmann H., Stewart C.L., Harbers K., Lohler J., Simon I. & Jaenisch R. (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature,* **298**: 623-628.

Jarrard D.F., Kinoshita H., Shi Y., Sandefur C., Hoff D., Meisner L.F., Chang C., Herman J.G., Isaacs W.B. & Nassif N. (1998) Methylation of the androgen receptor promoter CpG island is associated with loss of androgen receptor expression in prostate cancer cells. *Cancer Res.,* **58**: 5310-5314.

Johnson G.C. & Todd J.A. (2000) Strategies in complex disease mapping. *Curr Opin Genet Dev.,* **10**: 330-334.

Jones P.A., Wolkowicz M.J., Rideout W.M. 3rd., Gonzales F.A., Marziasz C.M., Coetzee G.A. & Tapscott S.J. (1990) De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines. *Proc Natl Acad Sci U S A,* **87**: 6117-6121.

Jones P.A. (1985) Altering gene expression with 5-azacytidine. *Cell,* **40**: 485-486.

Jones P.A. (1999) The DNA methylation paradox. *Trends Genet.,* **15**: 34-37.

Jones P.A. & Laird PW. (1999) Cancer epigenetics comes of age. *Nat Genet.,* **21**: 163-167.

Juttermann R., Li E. & Jaenisch R. (1994) Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc Natl Acad Sci U S A,* **91**: 11797-11801.

Kafri T., Ariel M., Brandeis M., Shemer R., Urven L., McCarrey J., Cedar H. & Razin A. (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.,* **6**: 705-714.

Kallioniemi A., Kallioniemi O.P., Sudar D., Rutovitz D., Gray J.W., Waldman F. & Pinkel D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science,* **258**: 818-821.

Kane M.F., Loda M., Gaida G.M., Lipman J., Mishra R., Goldman H., Jessup J.M. & Kolodner R. (1997) Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res*, **57**: 808-811.

Kanehisa M. (2000) *Post-Genome Informatics*. Oxford University Press: Oxford.

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

Kantharidis P., El-Osta A., deSilva M., Wall D.M., Hu X.F., Slater A., Nadalin G., Parkin J.D. & Zalcberg J.R. (1997) Altered methylation of the human MDR1 promoter is associated with acquired multidrug resistance. *Clin Cancer Res.,* **3**: 2025-2032.

Kashuba V., Szeles A., Allikmets R., Bergerheim U., Modi W., Grafodatsky A., Dean M., Stanbridge E.J., Winberg G., Klein G. & Zabarovsky E.R. (1995) A contig of NotI jumping and linking clones cover 2.5 Mb in 3p21-p22 region suspected to contain tumor suppressor gene. *Cancer Genetics & Cytogenet.,* **81**: 144-150.

Kashuba V.I., Gizatullin R.Z., Protopopov A.I., Allikmets R., Korolev S., Li J., Boldog F., Tory K., Zabarovska V., Marcsek Z., Sumegi J., Klein G., Zabarovsky E.R. & Kisselev L. (1997) NotI linking/jumping clones of human chromosome 3: mapping of the TFRC, RAB7 and HAUSP genes to regions rearranged in leukemia and deleted in solid tumors. *FEBS Lett.,* **419**: 181-185.

Kashuba V.I., Gizatullin R.Z., Protopopov A.I., Li J., Vorobieva N.V., Fedorova L., Zabarovska V.I., Muravenko O.V., Kost-Alimova M., Domninsky D.A., Kiss C., Allikmets R., Zakharyev V.M., Braga E.A., Sumegi J., Lerman M., Wahlestedt C., Zelenin A.V., Sheer D., Winberg G., Grafodatsky A., Kisselev L.L., Klein G. & Zabarovsky E.R. (1999) Analysis of NotI linking clones isolated from human chromosome 3 specific libraries. *Gene,* **239**: 259-271.

Kashuba V.I., Protopopov A.I., Kvasha S.M., Gizatullin R.Z., Wahlestedt C., Kisselev L.L., Klein G. & Zabarovsky E.R. (2002) hUNC93B1: a novel human gene representing a new gene family and encoding an unc-93-like protein. *Gene,* **283**: 209-217.

Kass S.U., Pruss D. & Wolffe A.P. (1997) How does DNA methylation repress transcription? *Trends Genet.,* **12**: 444-449.

Kato R. & Sasaki H. (1998) Quick identification and localization of CpG islands in large genomic fragments by partial digestion with Hpa II and Hha I. *DNA Res.,* **5**: 287-295.

Katzenellenbogen R.A., Baylin S.B. & Herman J.G. (1999) Hypermethylation of the DAP-kinase CpG island is a common alteration in B-cell malignancies. *Blood,* **93**: 4347-4353.

Killary A.M., Wolf M.E., Giambernardi T.A. & Naylor S.L. (1992) Definition of a tumor suppressor locus within human chromosome 3p21-p22. *Proc Natl Acad Sci U S A,* **89**:10877-10881.

Kim Y. (1999) Folate and cancer prevention: A new medical application of folate beyond hyperhomocysteinemia and neural tube defects. *Nutr Rev.,* **57**: 314-321.

Kim Y.I., Giuliano A., Hatch K.D., Schneider A., Nour M.A., Dallal G.E., Selhub J. & Mason J.B. (1994) Global DNA hypomethylation increases progressively in cervical dysplasia and carcinoma. *Cancer,* **74**: 893-899.

Kistner A., Gossen M., Zimmermann F., Jerecic J., Ullmer C., Lubbert H. & Bujard H. (1996) Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc Natl Acad Sci U S A,* **93**: 10933-10938.

Klein C.A., Schmidt-Kittler O., Schardt J.A., Pantel K., Speicher M.R. & Riethmuller G. (1999) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A,* **96**: 4494-4499.

Klein G. (1987) The approaching era of the tumor suppressor genes. *Science,* **238**: 1539-1545.

Knudson A.G. Jr. (1985) Hereditary cancer, oncogenes, and antioncogenes. *Cancer Res.,* **45**: 1437-1443.

Kohler B., Schumacher V., Schulte-Overberg U., Biewald W., Lennert T., l'Allemand D., Royer-Pokora B. & Gruters A. (1999) Bilateral Wilms tumor in a boy with severe hypospadias and cryptochidism due to a heterozygous mutation in the WT1 gene. *Pediatr Res.,* **45**: 187-190.

Kok K., Osinga J., Carritt B., Davis M.B., van der Hout A.H., van der Veen A.Y., Landsvater R.M., de Leij L.F., Berendsen H.H., Postmus P.E., et al. (1987) Deletion of a DNA sequence at the chromosomal region 3p21 in all major types of lung cancer. *Nature,* **330**: 578-581.

Kok K., van den Berg A., Veldhuis P.M., van der Veen A.Y., Franke M., Schoenmakers E.F., Hulsbeek M.M., van der Hout A.H., de Leij L., van de Ven W., et al. (1994) A homozygous deletion in a small cell lung cancer cell line involving a 3p21 region with a marked instability in yeast artificial chromosomes. *Cancer Res.,* **54**: 4183-4187.

Kok K., Naylor S.L. & Buys C.H. (1997) Deletions of the short arm of chromosome 3 in solid tumors and the search for suppressor genes. *Adv Cancer Res.*, **71**: 27-92.

Kovacs G., Erlandsson R., Boldog F., Ingvarsson S., Muller-Brechlin R., Klein G. & Sumegi J. (1988) Consistent chromosome 3p deletion and loss of heterozygosity in renal cell carcinoma. *Proc Natl Acad Sci U S A,* **85**: 1571-1575.

Krizman D.B. (1997) **In Birren B, Green ED, Klapholz S, Myers RM and Roskams J.** (eds), *Genome Analysis: A Laboratory Manual, Detecting Genes.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 2, pp. 191-216.

Kugoh H.M., Hashiba H., Shimizu M. & Oshimura M. (1990) Suggestive evidence for functionally distinct, tumor-suppressor genes on chromosomes 1 and 11 for a human fibrosarcoma cell line, HT1080. *Oncogene,* **5**: 1637-1644.

Kuo K.C., McCune R.A., Gehrke C.W., Midgett R. & Ehrlich M. (1980) Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.,* **8**: 4763-4776.

Kutsenko A.S., Gizatullin R.Z., Al-Amin A.N., Wang F., Kvasha S.M., Podowski R.M., Matushkin Y.G., Gyanchandani A., Muravenko O.V., Levitsky V.G., Kolchanov N.A., Protopopov A.I., Kashuba V.I., Kisselev L.L., Wasserman W., Wahlestedt C. & Zabarovsky E.R. (2002) NotI flanking sequences: a tool for gene discovery and verification of the human genome. *Nucleic Acids Res.,* **30**: 3163-3170.

Kuzmin I., Geil L., Ge H., Bengtsson U., Duh F.M., Stanbridge E.J. & Lerman M.I. (1999) Analysis of aberrant methylation of the VHL gene by transgenes, monochromosome transfer, and cell fusion. *Oncogene,* **18**: 5672-5679.

Kwok P.Y. & Chen X. (1998) Detection of single nucleotide variations. *Genet Eng (N Y),* **20**: 125-134.

LaForge K.S., Shick V., Spangler R., Proudnikov D., Yuferov V., Lysov Y., Mirzabekov A. & Kreek M.J. (2000) Detection of single nucleotide polymorphisms of the human mu opioid receptor gene by hybridization or single nucleotide extension on custom oligonucleotide gelpad microchips: potential in studies of addiction. *Am J Med Genet.,* **96**: 604-615.

Landegren U., Nilsson M. & Kwok P.Y. (1998) Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res.,* **8**: 769-776.

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov J.P., Miranda C.,

Morris W., Naylor J., Raymond C., et al. (2001) Initial sequencing and analysis of the human genome. *Nature,* **409**: 860-921.

Lapeyre J.N. & Becker F.F. (1979) 5-methylcytosine content of nuclear DANN during chemical hepatocarcinogenesis and in carcinomas which result. *Biochem Biophys Res Commun.,* **87**: 698-705.

Larramendy M.L., Lushnikova T., Bj]orkqvist A.M., Wistuba I.I., Virmani A.K., Shivapurkar N., Gazdar A.F. & Knuutila S. (2000) Comparative genomic hybridization reveals complex genetic changes in primary breast cancer tumors and their cell lines. *Cancer Genet Cytogenet.,* **119**: 132-138.

Lee W.H., Morton R.A., Epstein J.I., Brooks J.D., Campbell P.A., Bova G.S., Hsieh W.S., Isaacs W.B. & Nelson W.G. (1994) Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proc Natl Acad Sci U S A,* **91**: 11733-11737.

Leonhardt H., Page A.W., Weier H.U. & Bestor T.H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell,* **71**: 865-873.

Lerman M.I. & Minna J.D. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. (2000) The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. *Cancer Res.*, **60**: 6116-6133.

Li J., Protopopov A., Wang F., Senchenko V., Petushkov V., Vorontsova O., Petrenko L., Zabarovska V., Muravenko O., Braga E., Kisselev L., Lerman M.I., Kashuba V., Klein G., Ernberg I., Wahlestedt C. & Zabarovsky E.R. (2002) NotI subtraction and NotI-specific microarrays to detect copy number and methylation changes in whole genomes. *Proc Natl Acad Sci U S A,* **99**: 10724-10729.

Li Q., Jedlicka A., Ahuja N., Gibbons M.C., Baylin S.B., Burger P.C. & Issa J.P. (1998) Concordant methylation of the ER and N33 genes in glioblastoma multiforme. *Oncogene,* **16**: 3197-3202.

Li Q., Ahuja N., Burger P.C. & Issa J.P. (1999) Methylation and silencing of the Thrombospondin-1 promoter in human cancer. *Oncogene,* **18**: 3284-3289.

Liang G., Salem C.E., Yu M.C., Nguyen H.D., Gonzales F.A., Nguyen T.T., Nichols P.W. & Jones P.A. (1998) DNA methylation differences associated with tumor tissues identified by genome scanning analysis. *Genomics,* **53**: 260-268.

Liang G., Robertson K.D., Talmadge C., Sumegi J. & Jones P.A. (2000) The gene for a novel transmembrane protein containing epidermal growth factor and follistatin domains is frequently hypermethylated in human tumor cells. *Cancer Res.,* **60**: 4907-4912.

Liang P. & Pardee A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science,* **257**: 967-971.

Lindsay S. & Bird A.P. (1987) Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature,* **327**: 336-338.

Lipshutz R.J., Fodor S.P., Gingeras T.R. & Lockhart D.J. (1999) High density synthetic oligonucleotide arrays. *Nat Genet.,* **21**: 20-24.

Lisitsyn N., Lisitsyn N. & Wigler M. (1993) Cloning the differences between two complex genomes. *Science,* **259**: 946-951.

Lisitsyn N.A., Segre J.A., Kusumi K., Lisitsyn N.M., Nadeau J.H., Frankel W.N., Wigler M.H. & Lander E.S. (1994) Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. *Nat Genet.,* **6**: 57-63.

Lisitsyn N.A., Lisitsina N.M., Dalbagni G., Barker P., Sanchez C.A., Gnarra J., Linehan W.M., Reid B.J. & Wigler M.H. (1995) Comparative genomic analysis of tumors: detection of DNA losses and amplification. *Proc Natl Acad Sci U S A,* **92**: 151-155.

Lockhart D.J., Dong H., Byrne M.C., Follettie M.T., Gallo M.V., Chee M.S., Mittmann M., Wang C., Kobayashi M., Horton H. & Brown E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol.*, **14**: 1675-1680.

Lockhart D.J. & Winzeler E.A. (2000) Genomics, gene expression and DNA arrays. *Nature,* **405**: 827-836.

Loh W.E. Jr., Scrable H.J., Livanos E., Arboleda M.J., Cavenee W.K., Oshimura M. & Weissman B.E. (1992) Human chromosome 11 contains two different growth suppressor genes for embryonal rhabdomyosarcoma. *Proc Natl Acad Sci U S A,* **89**: 1755-1759.

Loots G.G., Locksley R.M., Blakespoor C.M., et al. (2000) Identification of a coordinate regulator of intrleukins 4, 13 and 5 by cross-species sequence comparisons. *Science*, **288**: 136-140.

Lou W., Krill D., Dhir R., Becich M.J., Dong J.T., Frierson H.F. Jr., Isaacs W.B., Isaacs J.T. & Gao A.C. (1999) Methylation of the CD44 metastasis suppressor gene in human prostate cancer. *Cancer Res.*, 59: 2329-2331.

Lovett M., Kere J. & Hinton L.M. (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc Natl Acad Sci U S A,* **88**: 9628-9632.

Lu L.J., Randerath E. & Randerath K. (1983) DNA hypomethylation in morris hepatomas. *Cancer Lett.,* **19**: 231-239.

Lyamichev V., Mast A.L., Hall J.G., Prudent J.R., Kaiser M.W., Takova T., Kwiatkowski R.W., Sander T.J., de Arruda M., Arco D.A., Neri B.P. & Brow M.A. (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol.,* **17**: 292-296.

Maekawa M., Sugano K., Kashiwabara H., Ushiama M., Fujita S., Yoshimori M. & Kakizoe T. (1999) DNA methylation analysis using bisulfite treatment and PCR-single-strand conformation polymorphism in colorectal cancer showing microsatellite instability. *Biochem Biophys Res Commun.,* **262**: 671-676.

Malik K., Salpekar A., Hancock A., Moorwood K., Jackson S., Charles A. & Brown K.W. (2000) Identification of differential methylation of the WT1 antisense regulatory region and relaxation of imprinting in Wilms' tumor. *Cancer Res.,* **60**: 2356-2360.

McAllister L., Penland L. & Brown P.O. (1998) Enrichment for loci identical-by-descent between pairs of mouse or human genomes by genomic mismatch scanning. *Genomics,* **47**: 7-11.

McConnell B.B. & Vertino P.M. (2000) Activation of a caspase-9-mediated apoptotic pathway by subcellular redistribution of the novel caspase recruitment domain protein TMS1. *Cancer Res.,* **60**: 6243-6247.

McQueen H.A., Fantes J., Cross S.H., Clark V.H., Archibald A.L. & Bird A.P. (1996) CpG islands of chicken are concentrated on microchromosomes. *Nat Genet.,* **12**: 321-324.

Merlo A., Herman J.G., Mao L., Lee D.J., Gabrielson E., Burger P.C., Baylin S.B. & Sidransky D. (1995) 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med.,* **1**: 686-692.

Miller G., Fuchs R. & Lai E. (1997) IMAGE cDNA clones, UniGene clustering, and ACeDB: an integrated resource for expressed sequence information. *Genome Res.*, **7**: 1027-1032.

Mirzayans F., Mears A.J., Guo S.W., Pearce W.G. & Walter M.A. (1997) Identification of the human chromosomal region containing the iridogoniodysgenesis anomaly locus by genomic-mismatch scanning. *Am J Hum Genet.,* **61**: 111-119.

Miwa W., Yashima K., Sekine T. & Sekiya T. (1995) Demethylation of a repetitive DNA sequence in human cancers. *Electrophoresis,* **16**: 227-232.

Monk M., Boubelik M. & Lehnert S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development,* **99**: 371-382.

Munger K. (2002) Disruption of oncogene/tumor suppressor networks during human carcinogenesis. *Cancer Invest.*, **20**: 71-81.

Narayan A., Ji W., Zhang X.Y., Marrogi A., Graff J.R., Baylin S.B. & Ehrlich M. (1998) Hypomethylation of pericentromeric DANN in breast adenocarcinomas. *Int J cancer,* **77**: 833-838.

Negrini M., Castagnoli A., Sabbioni S., Recanatini E., Giovannini G., Possati L., Stanbridge E.J., Nenci I. & Barbanti-Brodano G. (1992) Suppression of tumorigenesis by the breast cancer cell line MCF-7 following transfer of a normal human chromosome 11. *Oncogene,* **7**: 2013-2018.

Nelkin B.D., Przepiorka D., Burke P.J., Thomas E.D. & Baylin S.B. (1991) Abnormal methylation of the calcitonin gene marks progression of chronic myelogenous leukemia. *Blood,* **77**: 2431-2434.

Nelson S.F., McCusker J.H., Sander M.A., Kee Y., Modrich P. & Brown P.O. (1993) Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nat Genet.,* **4**: 11-18.

Nicolaides N.C., Littman S.J., Modrich P., Kinzler K.W. & Vogelstein B. (1998) A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. *Mol Cell Biol.,* **18**: 1635-1641.

Ohmura H., Tahara H., Suzuki M., Ide T., Shimizu M., Yoshida M.A., Tahara E., Shay J.W., Barrett J.C. & Oshimura M. (1995) Restoration of the cellular senescence program and repression of telomerase by human chromosome 3. *Jpn J Cancer Res.,* 86: 899-904.

Okano M., Bell D.W., Haber D.A. & Li E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* **99**: 247-257.

Okazaki Y., Okuizumi H., Sasaki N., Ohsumi T., Kuromitsu J., Hirota N., Muramatsu M. & Hayashizaki Y. (1995) An expanded system of restriction landmark genomic scanning (RLGS Ver. 1.8). *Electrophoresis,* **16**: 197-202.

Okazaki Y., Okuizumi H., Ohsumi T., Nomura O., Takada S., Kamiya M., Sasaki N., Matsuda Y., Nishimura M., Tagaya O., Muramatsu M. & Hayashizaki Y. (1996) A genetic linkage map of the Syrian hamster and localization of cardiomyopathy locus on chromosome 9qa2.1-b1 using RLGS spot-mapping. *Nat Genet.,* **13**: 87-90.

Onyango P., Miller W., Lehoczky J., Leung C.T., Birren B., Wheelan S., Dewar K. & Feinberg A.P. (2000) Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.,* **10**: 1697-1710.

Oshimura M., Kugoh H.M., Shimizu M., Yamada H., Hashiba H., Horikawa I. & Sasaki M. (1989) Multiple chromosomes carrying tumor suppressor activity, via microcell-mediated chromosome transfer, for various tumor cell lines. *Princess Takamatsu Symp.,* **20**: 249-257.

Oshimura M., Kugoh H., Koi M., Shimizu M., Yamada H., Satoh H. & Barrett J.C. (1990) Transfer of a normal human chromosome 11 suppresses tumorigenicity of some but not all tumor cell lines. *J Cell Biochem.,* **42**: 135-142.

Ottaviano Y.L., Issa J.P., Parl F.F., Smith H.S., Baylin S.B. & Davidson N.E. (1994) Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells. *Cancer Res.,* **54**: 2552-2555.

Palmisano W.A., Divine K.K., Saccomanno G., Gilliland F.D., Baylin S.B., Herman J.G. & Belinsky S.A. (2000) Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res.,* 60: 5954-5958.

Parimoo S., Patanjali S.R., Shukla H., Chaplin D.D. & Weissman S.M. (1991) cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc Natl Acad Sci U S A,* **88**: 9623-9627.

Pinkel D., Segraves R., Sudar D., Clark S., Poole I., Kowbel D., Collins C., Kuo W.L., Chen C., Zhai Y., Dairkee S.H., Ljung B.M., Gray J.W. & Albertson D.G. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet.,* **20**: 207-211.

Plass C., Shibata H., Kalcheva I., Mullins L., Kotelevtseva N., Mullins J., Kato R., Sasaki H., Hirotsune S., Okazaki Y., Held W.A., Hayashizaki Y. & Chapman V.M. (1996) Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nat Genet.,* **14**: 106-109.

Pollack J.R., Perou C.M., Alizadeh A.A., Eisen M.B., Pergamenschikov A., Williams C.F., Jeffrey S.S., Botstein D. & Brown P.O. (1998) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet.,* **23**: 41-46.

Poustka A. & Lehrach H. (1988) Chromosome jumping: A long range cloning technique. In ''Genetic Engineering Principles and Methods'' (J.K. Setlow, ED), Vol. 10, pp. 169-193, Brookhaven National laboratory, Upton, New York, Plenum Press New York/London.

Protopopov A.I., Gizatullin R.Z., Vorobieva N.V., Protopopova M.V., Kiss C., Kashuba V.I., Klein G., Kisselev L.L., Graphodatsky A.S. & Zabarovsky E.R. (1996) Human chromosome 3: high-resolution fluorescence in situ hybridization mapping of 40 unique NotI linking clones homologous to genes and cDNAs. *Chromosome Res.,* **4**: 443-447.

Qian X., von Wronski M.A. & Brent T.P. (1995) Localization of methylation sites in the human O6-methylguanine-DNA methyltransferase promoter: correlation with gene suppression. *Carcinogenesis,* **16**: 1385-1390.

Qu G.Z., Dubeau L., Narayan A., Yu M.C. & Ehrlich M. (1999) Satellite DNA hypomethylation vs overall genomic hypomethylation in ovian epithelial tumors of different malignant potential. Mutat *Res Fund Mol Mech Mutagens,* **423**: 91-101.

Ramsahoye B.H., Biniszkiewicz D., Lyko F., Clark V., Bird A.P. & Jaenisch R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A,* **97**: 5237-5242.

Razin A. & Riggs A.D. (1980) DNA methylation and gene function. *Science,* **210**: 604-610.

Rice J.C., Massey-Brown K.S. & Futscher B.W. (1998) Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. *Oncogene,* **17**: 1807-1812.

Rice J.C. & Futscher B.W. (2000) Transcriptional repression of BRCA1 by aberrant cytosine methylation, histone hypoacetylation and chromatin condensation of the BRCA1 promoter. *Nucleic Acids Res.,* **28**: 3233-3239.

Rideout W.M.I., Coetzee G.A., Olumi A.F. & Jones P.A.. (1990) 5-methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science*, **249**: 1288-1290.

Riggs A.D. & Jones P.A. (1983) 5-methylcytosine, gene regulation, and cancer. *Adv Cancer,* **40**: 1-30.

Rimessi P., Gualandi F., Morelli C., Trabanelli C., Wu Q., Possati L., Montesi M., Barrett J.C. & Barbanti-Brodano G. (1994) Transfer of human chromosome 3 to an ovarian carcinoma cell line identifies three regions on 3p involved in ovarian cancer. *Oncogene,* **9**: 3467-3474.

Ringwald M., Eppig J.T., Kadin J.A. & Richardson J.E. (2000) GXD: a Gene Expression Database for the laboratory mouse: current status and recent enhancements. The Gene Expresison Database group. *Nucleic Acids Res.,* **28**: 115-119.

Risch N.J. (2000) Searching for genetic determinants in the new millennium. *Nature,* **405**: 847-856.

Robertson K.D., Uzvolgyi E., Liang G., Talmadge C., Suegi J., Gonzales F.A. & Jones P.A. (1999) The human DNA methyltransferases (DNMTS) 1,3a, 3b: coordinate mRNA expression in normal and overexpression in tumors. *Nucleic Acids Res.,* **27**: 2291-2298.

Robertson K.D., Wolffe A.P. (2000) DNA methylation in health and disease. *Nat Rev Genet.,* **1**: 11-19.

Rosenberg M., Przybylska M. & Straus D. (1994) "RFLP subtraction": a method for making libraries of polymorphic markers. *Proc Natl Acad Sci U S A,* **91**: 6113-6117.

Rountree M.R., Bachman K.E. & Baylin S.B. (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet.,* **25**: 269-277.

Rountree M.R., Bachman K.E., Herman J.G. & Baylin S.B. (2001) DNA methylation, chromatin inheritance, and cancer. *Oncogene,* **20**: 3156-3165.

Sachidadandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J., Willey D.L., Hunt S.E., Cole C.G., Coggill P.C., Rice C.M., Ning Z., Rogers J., Bentley D.R., Kwok P.Y., Mardis E.R., Yeh R.T., Schultz B., Cook L., Davenport R., Dante M., Fulton L., Hillier L., Waterston R.H., McPherson J.D., Gilman B., Schaffner S., Van Etten W.J., Reich D., Higgins J., Daly M.J., Blumenstiel B., Baldwin J., Stange-Thomann N., Zody M.C., Linton L., Lander E.S. & Altshuler D.; The International SNP Map Working Group. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409: 928-933.

Sager R. (1985) Genetic suppression of tumor formation. *Adv Cancer Res.,* **44**: 43-68.

Sager R. (1986) Genetic suppression of tumor formation: a new frontier in cancer research. *Cancer Res.,* **46**: 1573-1580.

Saito A., Abad J.P., Wang D.N., Ohki M., Cantor C.R. & Smith C.L. (1991) Construction and characterization of a NotI linking library of human chromosome 21. *Genomics,* **10**: 618-630.

Saito Y., Takazawa H., Uzawa K., Tanzawa H. & Sato K. (1998) Reduced expression of E-cadherin in oral squamous cell carcinoma: relationship with DNA methylation of 5' CpG island. *Int J Oncol.,* **12**: 293-298.

Salem C.E., Markl I.D., Bender C.M., Gonzales F.A., Jones P.A. & Liang G. (2000) PAX6 methylation and ectopic expression in human tumor cells. *Int J Cancer,* **87**: 179-185.

Salvesen H.B., MacDonald N., Ryan A., Jacobs I.J., Lynch E.D., Akslen L.A. & Das S. (2001) PTEN methylation is associated with advanced stage and microsatellite instability in endometrial carcinoma. *Int J Cancer,* **91**: 22-26.

Sanchez Y., el-Naggar A., Pathak S. & Killary A.M. (1994) A tumor suppressor locus within 3p14-p12 mediates rapid cell death of renal cell carcinoma *in vivo. Proc Natl Acad Sci U S A,* **91**: 3383-3387.

Satoh H., Lamb P.W., Dong J.T., Everitt J., Boreiko C., Oshimura M. & Barrett J.C. (1993) Suppression of tumorigenicity of A549 lung adenocarcinoma cells by human chromosomes 3 and 11 introduced via microcell-mediated chromosome transfer. *Mol Carcinog.,* **7**: 157-164.

Saxon P.J., Srivatsan E.S. & Stanbridge E.J. (1986) Introduction of human chromosome 11 via microcell transfer controls tumorigenic expression of HeLa cells. *EMBO J.,* **5**: 3461-3466.

Schena M., Shalon D., Davis R.W. & Brown P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science,* **270**: 467-470.

Schena M., Shalon D., Heller R., Chai A., Brown P.O. & Davis R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci* U S A, **93**: 10614-10619.

Schmutte C. & Jones P.A. (1998) Involvement of DNA methylation in human carcinogenesis. *Biol Chem.,* **379**: 377-388.

Sekido Y., Bader S., Latif F., Gnarra J.R., Gazdar A.F., Linehan W.M., Zbar B., Lerman M.I. & Minna J.D. (1994) Molecular analysis of the von Hippel-Lindau disease tumor suppressor gene in human lung cancer cell lines. *Oncogene*, **9**: 1599-1604.

Shalon D., Smith S.J. & Brown P.O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.,* **6**: 639-645.

Sharrard R.M., Royds J.A., Rogers S. & Shorthouse A.J. (1992) Patterns of methylation of the c-myc gene in human colorectal cancer progression. *Blood,* **82**: 1820-1828.

Shen J.C., Rideout W.M. & Jones P.A. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-strand DNA. *Nucleic Acids Res.,* **22**: 972-976.

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

Shi H., Yan P.S., Chen C., Rahmatpanah F., Lofton-Day C., Caldwell C.W. & Huang T.H. (2002) Expressed CpG island sequence tag microarray for dual screening of DNA hypermethylation and gene silencing in cancer cells. *Cancer Res.,* **62**: 3214-3220.

Shibata H., Yoshino K., Muramatsu M., Plass C., Chapman V.M. & Hayashizaki Y. (1995) The use of restriction landmark genomic scanning to scan the mouse genome for endogenous loci with imprinted patterns of methylation. *Electrophoresis,* **16**: 210-217.

Shimizu M., Yokota J., Mori N., Shuin T., Shinoda M., Terada M. & Oshimura M. (1990) Introduction of normal chromosome 3p modulates the tumorigenicity of a human renal cell carcinoma cell line YCR. *Oncogene,* **5**:185-194.

Shiraishi M., Oates A.J., Li X., Hosoda F., Ohki M., Alitalo T., Lerman L.S. & Sekiya T. (1998) The isolation of CpG islands from human chromosomal regions 11q13 and Xp22 by segregation of partlymelted molecules. *Nucleic Acids Res.,* **26**: 5544-5550.

Silva A.J. & White R. (1988) Inheritance of allelic blueprints for methylation patterns. *Cell,* **54**: 145-152.

Simpkins S.B., Bocker T., Swisher E.M., Mutch D.G., Gersell D.J., Kovatich A.J., Palazzo J.P., Fishel R. & Goodfellow P.J. (1999) MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum Mol Genet.,* **8**: 661-666.

Singh B., Gogineni S.K., Sacks P.G., Shaha A.R., Shah J.P., Stoffel A. & Rao P.H. (2001) Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. *Cancer Res.,* **61**: 4506-4513.

Smiraglia D.J. & Plass C. (2002) The study of aberrant methylation in cancer via restriction landmark genomic scanning. *Oncogene,* **21**: 5414-5426.

Snijders A.M., Nowak N., Segraves R., Blackwood S., Brown N., Conroy J., Hamilton G., Hindle A.K., Huey B., Kimura K., Law S., Myambo K., Palmer J., Ylstra B., Yue J.P., Gray J.W., Jain A.N., Pinkel D. & Albertson D.G. (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet.,* **29**: 263-264.

Solinas-Toldo S., Lampel S., Stilgenbauer S., Nickolenko J., Benner A., Dohner H., Cremer T. & Lichter P. (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chrom Cancer,* **20**: 399-407.

Somogyi R. & Sniegoski C.A. (1996) Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation. *Complexity,* **1**: 45-63.

Spellman P.T., Sherlock G., Zhang M.Q., Iyer V.R., Anders K., Eisen M.B., Brown P.O., Botstein D. & Futcher B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell,* **9**: 3273-3297.

Stirzaker C., Millar D.S., Paul C.L., Warnecke P.M., Harrison J., Vincent P.C., Frommer M. & Clark S.J. (1997) Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res.,* **57**: 2229-2237.

Sugai M., Kondo S., Shimizu A. & Honjo T. (1998) Isolation of differentially expressed genes upon immunoglobulin class switching by a subtractive hybridization method using uracil DNA glycosylase. *Nucleic Acids Res.,* **26**: 911-918.

Suzuki H., Itoh F., Toyota M., Kikuchi T., Kakiuchi H. & Imai K. (2000) Inactivation of the 14-3-3 sigma gene is associated with 5' CpG island hypermethylation in human cancers. *Cancer Res.,* **60**: 4353-4357.

Szeles A., Bajalica-Lagercrantz S., Lindblom A., Lushnikova T., Kashuba V.I., Imreh S., Nordenskjold M., Klein G. & Zabarovsky E.R. (1996) Mapping of a new MAP kinase activated protein kinase gene (3PK) to human chromosome band 3p21.2 and ordering of 3PK and two cosmid markers in the 3p22-p21 tumour-suppressor region by two-colour fluorescence in situ hybridization. *Chromosome Res.,* **4**: 310-313.

Taillon-Miller P., Gu Z., Li Q., Hillier L. & Kwok P.Y. (1998) Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res.*, **8**: 748-754.

Tanaka H., Shimizu M., Horikawa I., Kugoh H., Yokota J., Barrett J.C. & Oshimura M. (1998) Evidence for a putative telomerase repressor gene in the 3p14.2-p21.1 region. *Genes Chrom Cancer*, **23**: 123-133.

Tanaka K., Oshimura M., Kikuchi R., Seki M., Hayashi T. & Miyaki M. (1991) Suppression of tumorigenicity in human colon carcinoma cells by introduction of normal chromosome 5 or 18. *Nature*, **349**: 340-342.

Teitz T., Wei T., Valentine M.B., Vanin E.F., Grenet J., Valentine V.A., Behm F.G., Look A.T., Lahti J.M. & Kidd V.J. (2000) Caspase 8 is deleted or silenced preferentially in childhood neuroblastomas with amplification of MYCN. *Nat Med.*, **6**: 529-535.

Thayer R.E., Singer M.F. & Fanning T.G. (1993) Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. *Gene*, **133**: 273-277.

Todd M.C., Xiang R.H., Garcia D.K., Kerbacher K.E., Moore S.L., Hensel C.H., Liu P., Siciliano M.J., Kok K., van den Berg A., Veldhuis P., Buys C.H., Killary A.M. & Naylor S.L. (1996) An 80 Kb P1 clone from chromosome 3p21.3 suppresses tumor growth in vivo. *Oncogene*, 13: 2387-2396.

Tompa R., McCallum C.M., Delrow J., Henikoff J.G., van Steensel B. & Henikoff S. (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr Biol.*, **12**: 65-68.

Toyota M., Ho C., Ohe-Toyota M., Baylin S.B. & Issa J.P. (1999) Inactivation of CACNA1G, a T-type calcium channel gene, by aberrant methylation of its 5' CpG island in human tumors. *Cancer Res.*, **59**: 4535-4541.

Toyota M., Shen L., Ohe-Toyota M., Hamilton S.R., Sinicrope F.A. & Issa J.P. (2000) Aberrant methylation of the Cyclooxygenase 2 CpG island in colorectal tumors. *Cancer Res.*, **60**: 4044-4048.

Trent J.M., Stanbridge E.J., McBride H.L., Meese E.U., Casey G., Araujo D.E., Witkowski C.M. & Nagle R.B. (1990) Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science*, **247**: 568-571.

Tsuchiya T., Tamura G., Sato K., Endoh Y., Sakata K., Jin Z., Motoyama T., Usuba O., Kimura W., Nishizuka S., Wilson K.T., James S.P., Yin J., Fleisher A.S., Zou T., Silverberg S.G., Kong D. & Meltzer S.J. (2000) Distinct methylation patterns of two APC gene promoters in normal and cancerous gastric epithelia. *Oncogene*, **19**: 3642-3646.

Tulchinsky E., Grigorian M., Tkatch T., Georgiev G. & Lukanidin E. (1995) Transcriptional regulation of the mts1 gene in human lymphoma cells: the role of DNA methylation. *Biochem Biophys Acta.*, **1261**: 243-248.

Tycko B. (2000) Epigenetic gene silencing in cancer. *J Clin Invest.*, **105**: 401-407.

Ushijima T., Morimura K., Hosoya Y., Okonogi H., Tatematsu M., Sugimura T. & Nagao M. (1997) Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. *Proc Natl Acad Sci U S A*, **94**: 2284-2289.

Vachtenheim J., Horakova I. & Novotna H. (1994) Hypomethylation of CCGG sites in the 3' region of H-ras protooncogene is frequent and is associated with H-ras allele loss in non-small cell lung cancer. *Cancer Res.*, **54**: 1145-1148.

Valdes J.M., Tagle D.A. & Collins F.S. (1994) Island rescue PCR: a rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. *Proc Natl Acad Sci U S A*, **91**: 5377-5381.

Veigl M.L., Kasturi L., Olechnowicz J., Ma A.H., Lutterbaugh J.D., Periyasamy S., Li G.M., Drummond J., Modrich P.L., Sedwick W.D. & Markowitz S.D. (1998) Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci U S A*, 95: 8698-8702.

Velculescu V.E., Zhang L., Vogelstein B. & Kinzler K.W. (1995) Serial analysis of gene expression. *Science,* **270**: 484-487.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Wortman J.R., Zhang Q., Kodira C.D., Zheng X.H., Chen L., Skupski M., Subramanian G., Thomas P.D., Zhang J., Gabor Miklos G.L., Nelson C., et al. (2001) The sequence of the human genome. *Science,* **291**: 1304-1351.

Vogelstein B., Lane D. & Levine A.J. (2000) Surfing the p53 network. *Nature*, **408**: 307-310.

Vos MD, Ellis CA, Bell A, Birrer MJ and Clark GJ. (2000). Ras uses the novel tumor suppressor RASSF1 as an effector to mediate apoptosis. *J. Biol. Chem.*, **275**, 35669-35672.

Wajed S.A., Laird P.W. & DeMeester T.R. (2001) DNA methylation: an alternative pathway to cancer. *Ann Surg.,* **243**: 10-20.

Wales M.M., Biel M.A., el Deiry W., Nelkin B.D., Issa J.P., Cavenee W.K., Kuerbitz S.J. & Baylin S.B. (1995) p53 activates expression of HIC-1, a new candidate tumour suppressor gene on 17p13.3. *Nat Med.,* **1**: 570-577.

Wallace M.R., Fountain J.W., Brereton A.M. & Collins F.S. (1989) Direct construction of a chromosome-specific NotI linking library from flow-sorted chromosomes. *Nucleic Acids Res.,* **17**: 1665-1677.

Weber J., Salgaller M., Samid D., Johnson B., Herlyn M., Lassam N., Treisman J. & Rosenberg S.A. (1994) Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine. *Cancer Res.,* **54**: 1766-1771.

Weber J.L. (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. *Genomics,* **7**: 524-530.

Weiss K.M. (1998) In search of human variation. *Genome Res.,* **8**: 691-697.

Welsh K.M., Lu A.L., Clark S. & Modrich P. (1987) Isolation and characterization of the Escherichia coli mutH gene product. *J Biol Chem.,* **262**: 15624-15629.

Wistuba I.I., Behrens C., Virmani A.K., Mele G., Milchgrub S., Girard L., Fondon J.W. 3rd., Garner H.R., McKay B., Latif F., Lerman M.I., Lam S., Gazdar A.F. & Minna J.D. (2000) High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res*., 60: 1949-1960.

Wolffe A.P. (1996) Histone deacetylase: a regulator of transcription. *Science,* **272**: 371-372.

Woodcock D.M., Lawler C.B., Linsenmeyer M.E., Doherty J.P. & Warren W.D. (1997) Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J Biol Chem.,* **272**: 7810-7816.

Worm J., Bartkova J., Kirkin A.F., Straten P., Zeuthen J., Bartek J. & Guldberg P. (2000) Aberrant p27Kip1 promoter methylation in malignant melanoma. *Oncogene,* **19**: 5111-5115.

Xie S., Wang Z., Okano M., Nogami M., Li Y., He W.W., Okumura K. & Li E. (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene,* **236**: 87-95.

Xiong Z. & Laird P.W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.,* **25**: 2532-2534.

Yamada H., Wake N., Fujimoto S., Barrett J.C. & Oshimura M. (1990) Multiple chromosomes carrying tumor suppressor activity for a uterine endometrial carcinoma cell line identified by microcell-mediated chromosome transfer. *Oncogene,* **5**: 1141-1147.

Yamakawa K., Takahashi T., Horio Y., Murata Y., Takahashi E., Hibi K., Yokoyama S., Ueda R., Takahashi T. & Nakamura Y. (1993) Frequent homozygous deletions in lung cancer cell lines detected by a DNA marker located at 3p21.3-p22. *Oncogene,* **8**: 327-330.

Yan P.S., Chen C.M., Shi H., Rahmatpanah F., Wei S.H., Caldwell C.W. & Huang T.H. (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.,* **61**: 8375-8380.

Yen C.C., Chen Y.J., Chen J.T., Hsia J.Y., Chen P.M., Liu J.H., Fan F.S., Chiou T.J., Wang W.S. & Lin C.H. (2001) Comparative genomic hybridization of esophageal squamous cell carcinoma: correlations between chromosomal aberrations and disease progression/prognosis. *Cancer,* **92**: 2769-2777.

Yoder J.A., Walsh C.P. & Bestor T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trend Genet.,* **13**: 335-340.

Yoshiura K., Kanai Y., Ochiai A., Shimoyama Y., Sugimura T. & Hirohashi S. (1995) Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc Natl Acad Sci U S A,* **92**: 7416-7419.

Youil R., Kemper B. & Cotton R.G.H. (1995) Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc Natl Acad Sci U S A,* **92**: 87-91.

Zabarovsky E.R., Boldog F., Thompson T., Scanlon D., Winberg G., Marcsek Z., Erlandsson R., Stanbridge E.J., Klein G. & Sumegi J. (1990) Construction of a human chromosome 3 specific NotI linking library using a novel cloning procedure. *Nucleic Acids Res.,* **18**: 6319-6324.

Zabarovsky E.R., Boldog F., Erlandsson R., Kashuba V.I., Allikmets R.L., Marcsek Z., Kisselev L.L., Stanbridge E., Klein G., Sumegi J., et al. (1991) New strategy for mapping the human genome based on a novel procedure for construction of jumping libraries. *Genomics,* **11**: 1030-1039.

Zabarovsky E.R., Allikmets R., Kholodnyuk I., Zabarovska V.I., Paulsson N., Bannikov V.M., Kashuba V.I., Dean M., Kisselev L.L. & Klein G. (1994) Construction of representative NotI linking libraries specific for the total human genome and for human chromosome 3. *Genomics,* **20**: 312-316.

Zabarovsky E.R., Gizatullin R., Podowski R.M., Zabarovska V.V., Xie L., Muravenko O.V., Kozyrev S., Petrenko L., Skobeleva N., Li J., Protopopov A., Kashuba V., Ernberg I., Winberg G. & Wahlestedt C. (2000) NotI clones in the analysis of the human genome. *Nucleic Acids Res.,* **28**: 1635-1639.

Zabarovsky ER, Lerman MI, Minna JD. (2002) Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene,* **21**: 6915-6935.

Zardo G. & Caiafa P. (1998) The unmethylated state of CpG islands in mouse fibroblasts depends on the poly(ADP-ribosyl)ation process. *J Biol Chem.,* **273**: 16517-16520.

Zheng B.H., Mills A.A. & Bradley A. (2001) Introducing defined chromosomal rearrangements into the mouse genome. *Methods,* **24**: 81-94.

Zheng S., Chen P., McMillan A., Lafuente A., Lafuente M.J., Ballesta A., Trias M. & Wiencke J.K. (2000) Correlations of partial and extensive methylation at the p14(ARF) locus with reduced mRNA expression in colorectal cancer cell lines and clinicopathological features in primary tumors. *Carcinogenesis,* **21**: 2057-2064.

Zhou Y., Li J., Xu K., Hu S.X., Benedict W.F. & Xu H.J. (1994) Further characterization of retinoblastoma gene-mediated cell growth and tumor suppression in human cancer cells. *Proc Natl Acad Sci U S A,* **91**: 4165-4169.