CANCER-RELATED GENE REGULATION MECHANISMS

Veerla, Srinivas

2008

CANCER-RELATED GENE REGULATION MECHANISMS

SRINIVAS VEERLA

DEPARTMENT OF CLINICAL GENETICS

LUND UNIVERSITY
Faculty of Medicine
2008

ISBN 978-91-86059-24-8

PRINTED IN SWEDEN 2008 BY MEDIA-TRYCK, LUND

This thesis is dedicated to my god father,
Hans-Joachim Hassenberg.

TABLE OF CONTENTS

ORIGINAL ARTICLES

This thesis is based on the below listed original articles, referred to in the text by their Roman numerals (**I-III**):

I. Veerla S and Höglund M. Analysis of Promoter Regions of Co-expressed Genes Identified by Microarray Analysis. *BMC Bioinformatics*, 2006, 7:384.

II. Veerla S, Panagopoulos I, Yuesheng J, Lindgren D, Höglund M. Promoter Analysis of Epigenetically Controlled Genes in Bladder Cancer. *Genes, Chromosomes and Cancer*, 2008, 47(5):368-78.

III. Veerla S, Lindgren D, Kvist A, Frigyesi A, Staaf J, Persson H, Liedberg F, Chebil G, Gudjosson S, Borg Å, Månsson W, Rovira C, Höglund M. Patterns of MicroRNA Expression in Urothelial Carcinoma; Identification of Potential Prognostic Markers. *Manuscript*.

ABBREVIATIONS

AML	Acute Myeloid Leukemia
bp	Base Pair
cDNA	Complementary DNA
CGH	Comparative Genomic Hybridization
ChIP	Chromatin Immunoprecipitation
DNAse I	Deoxyribonuclease I
DNMT	DNA Methyl Transferase
HCA	Hierarchical Clustering Analaysis
HCP	High CpG-content Promoter
ICP	Intermediate CpG-content Promoter
LCP	Low CpG-content Promoter
LOH	Loss of Heterozygosity
Mb	Mega base
MBD	Methyl-CpG Binding Domain
miRNA	MicroRNA
mRNA	Messenger RNA
nt	Nucleotide
PCR	Polymerase Chain Reaction
PTM	Pavlidis Template Matching
PWM	Positional Weight Matrices
QTC	Quality Cluster Algorithm
SAM	Significance Analysis of Microarrays
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TSS	Transcription Start Site
UC	Urothelial Carcinoma
UTR	Untranslated Region
wt	Wild Type

INTRODUCTION

CANCER GENETICS

Cancer is a collection of complex diseases caused by genetic alterations. The main characteristics of cancer are uncontrolled cell proliferation and the ability of cancer cells to migrate from the original location and spread to distant locations. These features are under the control of genes. Genes carry the information to produce proteins and are parts of DNA molecules, packaged to chromosome by interaction with chromatin proteins. In cancer, chromosomes may undergo changes resulting in chromosomal aberrations that may be observed during the late stage of the cell division process, the mitosis. Theodor Boveri (1862-1915) who studied the effect of aneuploidy in sea urchin was the first to suggest a role of abnormal chromosomes in cancer[1]. Chromosomal aberrations, such as altered chromosomal numbers and structural changes usually lead to variety of genetic disorders. The structural changes may result in the fusion of two different genes, so called fusion genes, gene amplifications, or gene deletions. The first identified chromosomal aberration in cancer was the t(9:22) translocation seen in ~95% of all patients with chronic myeloid leukemia (CML) and results in the *BCR/ABL* fusion gene[2-4].

Apart from the chromosomal changes, gene mutations are also frequently seen in cancer. Genes affected by somatic mutation are classified as dominantly acting oncogenes or as recessive tumor suppressor genes. Oncogenes contribute to the conversion of a normal cell into a cancerous cell upon activation whereas tumor suppressor genes contribute to cancerous growth by inactivation. It is known that on average six to seven successive mutations are required to convert a normal epithelial cell into a cancer cell. Point mutations i.e., mutation in one nucleotide of a gene, seems to be responsible for the activation of several oncogenes such as members of RAS family of genes, *HRAS*, *KRAS*, and *NRAS*[5,6]. In contrast to oncogenes, and in accordance to Knudson's two-hit hypothesis[7], both alleles of a tumor suppressor gene has to be inactivated. For example, in the inherited form of retinoblastoma, an embryonic malignant neoplasm of retinal origin, the patient has inherited one mutated version of *RB1* and acquired a somatic mutations in the other allele[8]. The most commonly

inactivated tumor suppressor gene is *TP53* that show mutations in more than 50% of all cancers. The *TP53* gene encodes the p53 protein that functions as a transcription factor and regulates the expression of several genes responsible for DNA repair and proper cell growth and division.

Apart from genetic lesions, cancers frequently show epigenetic changes caused by DNA methylation. Several studies have shown that aberrant DNA methylation in promoter regions may lead to silencing of genes and thus represent an alternative mechanism to inactivate tumor suppressor genes. It is believed that DNA methylation also may inhibit binding of transcription factors to promoter regions and thereby modify the gene expression. Recently, an additional mechanism has been implicated in gene regulation that involves microRNAs. MicroRNAs affect gene expression by binding to mRNAs and thus regulate genes in a post-transcription fashion. Currently, there are at least 685 known microRNAs[9]. Several investigations suggest that these microRNAs are important for maintaining normal cellular process and that improper activation or inactivation of these microRNAs may lead to diseases.

GENE REGULATION

Gene expression determines the structure and function of cells and is the basis for cellular differentiation, morphogenesis, and adaptation to environment. Gene regulation occurs at four major levels; transcriptional, post-transcriptional, translation, and post-translational level.

Gene transcription involves the polymerization of ribonucleotide precursors into an RNA molecule using DNA as a template. The enzymes that carry out this process are known as RNA polymerases. There are three types of RNA polymerases that transcribe three different groups of genes; RNA polymerase I for ribosomal RNAs, RNA polymerase II for messenger RNAs (mRNAs) that code for proteins, and RNA polymerase III for tRNAs and other small RNAs like 5.8S RNAs. At transcriptional level gene regulation is mainly mediated through transcription factors (TFs) and RNA polymerases that bind in close vicinity of the transcriptional start site, called the promoter region. Promoters are usually categorized into three types based on the type of RNA

polymerase that they bind; RNA polymerases I, II, and III promoters. These promoters together with several additional factors such as TFs, DNA methylation, and chromatin modifications, may either activate or inhibit transcription of the corresponding genes.

The activity of protein coding genes may in addition be regulated at the post-transcriptional level either through variable mRNA stability or through interactions with microRNAs. mRNAs contains at their 5'-ends a so called 5'-cap that protects mRNA from RNAses and consists of a modified guanine nucleotide added to the mRNAs shortly after transcription. After transcription a poly (A) tail is added to the 3'-end that stabilizes the mRNA. The length of the poly (A) tail effects the half life of the mRNAs and makes mRNAs with long poly (A) tail more stable[10-14]. The removal of the poly (A) tail is often the first step in mRNA degradation. In addition, the overall mRNA structure may influence the stability of the mRNA. Apart from this non-coding RNAs, microRNAs, can either bind to the coding region or to the 3' UTR of the mRNA, and affect mRNA stability or translational activity (see below).

Post-translation modification may affect the activity of proteins by attaching functional groups like acetate, phosphate, lipids, and carbohydrates. For example, phosphorylation is known to control the activity of key proteins in several signaling pathways, e.g., TP53 function and stability are regulated by phosphorylation[15] and histone proteins that control the accessibility of DNA to RNA polymerases are subjected to acetylation[16-18]. In addition, proteins may also be modified by the action of proteases. Examples of this are the production of plasmin from plasminogen and the activation of TGFβ from the latent TGFβ protein[19-21]. Hence, gene regulation involves activities at the transcriptional, post-transcription, and post-translational level and is thus highly complex. In addition, gene regulation may involve multiple transcription factors, complex interactions with small non-coding RNAs, and various protein modifications. In normal cells these interactions ensure that specific genes are expressed at appropriate levels to fulfill their proper functions. Alterations in any of these three levels may drastically change the activity of a gene and result in minor or major phenotypic effects depending on the function of the gene. Of particular importance is the alteration of regulatory genes as they may affect major cellular processes.

Gene Promoters

Genes are regulated by non-coding DNA sequences usually localized upstream of the coding sequence and upstream of the transcription start site (TSS). These regions are called promoters and may sometimes extend downstream of the TSS and include the first exon. It is normally difficult to determine the exact borders of promoters it is, however, believed that they may vary from 0.5kb to 5kb in size. For instance, one of the best characterized promoters in eukaryotes is the *Endo16* promoter in sea urchin. This promoter is 2.3kb in size and include several sequence elements that regulate *Endo16* expression[22]. Promoters are subdivided in to three main regions; basal or core, proximal, and distal promoter regions. The basal promoter region contains binding sites for common transcription factors and is located about 40bp upstream of the TSS of most genes. The basal promoters are likely to be involved in the basic process of transcription. One of the common binding sites in this region is the TATA box, an AT-rich sequence located at ~30bp upstream of the TSS. In eukaryotes, the TFIID protein complex binds to the TATA box as an initial step in the formation of a stable transcriptional initiator complex. This initial complex then recruits additional TFII factors, including one DNA helicase, the RNA polymerase II, and the SWI/SNF complex. The DNA helicase (TFIIH) unwinds the double-stranded DNA to make the DNA accessible for the RNA polymerase. The SWI/SNF complex plays an important role in remodeling the chromatin into a form that allows binding of additional transcription factors.

The proximal promoter region is located upstream of the basal promoter region, usually 50 to 250bp upstream of the transcription start site, and may contain binding sites for several TFs. Some of the common binding sites in this region are CCAAT (CCAAT boxes) and GGGGCGG (GC boxes) that are recognized by TFs such as nuclear factor I (NF-I) and nuclear factor Y (NF-Y), and specificity protein 1 (SP1), respectively. The major function of these binding sites is to regulate the basal transcriptional activity. In addition to factors that promote the activity of genes, promoter regions may also contain elements that repress the basal transcriptional activity. For example, the neural-restrictive silencer element (NRSE) bind to the neuron-restrictive silencer factor that repress neuronal gene transcription in non-neuronal cells[23]. Furthermore, proximal promoters frequently contain binding sites for TFs involved in e.g., response to external stimuli, and

for tissues specific expression. Examples are, the E2F binding sites that bind the E2F family of transcription factors that regulates cell cycle progression[24] and binding sites for the PDX1 transcription factor in the insulin gene that insures the production of insulin in pancreas beta cells only[25-28].

Apart from the binding sites in the basal and proximal promoter it is also known that TFs may bind at far distance to the TSS, up to several kb, and still influence the gene activity. These sites have either moderate or very strong influence on gene activity. Sequences with a strong influence are usually called enhancers. Enhancers consist of several repeats of consensus transcription factor binding sites (TFBSs) that are recognized by multiple transcription factors. The enhancer does not have to be located in the distal promoter, near to the TSS of genes it acts on, nor on the same chromosome. For example, the *IFNG* gene is located on chromosome 10 and its expression regulated by enhancer elements located on chromosome 11[29,30], a cell type-specific enhancer of the mouse thrombospondin 3 gene (*Thbs3*) is located within intron 6 of the adjacent metaxin gene (*Mtx*) which is located approximately 6kb upstream of the *Thbs3* gene[31]. The lymphoid enhancer binding factor 1 (LEF-1) binds to the enhancer of the T-cell receptor α gene and bends the DNA in such a way that it is positioned close to the basal promoter and facilitates interaction of other TFs bound to the enhancer with the basal promoter[32].

Identification of TFBSs using Experimental Methods

Methods used to identify DNA segments that are bound by TFs includes DNAse hypersensitive assays[33] and chromatin immunoprecipitation assays (ChIP)[34-39]. In addition there are methods to identify the precise sites or sequence motifs that actively binds transcription factor. These methods include DNA mobility shift, or gel retardation assays, and deoxyribonuclease I (DNAse I) footprinting assays.

In chromatin there are regions that are free from nucleosomes and are highly sensitive to nucleases and referred to as DNAse hypersensitive sites. It is believed that such nucleosome free regions allow for enhanced access of TFBSs[40]. Therefore, screening for DNAse hypersensitive sites may facilitate the identification of segments that are occupied by regulatory proteins[41]. Using this approach segments that take active part in the binding of TFs has been identified in several genes e.g., the binding sites for

pituitary POU domain transcription factor (Pit-1) in the promoter region of human growth hormone gene (*hGH-N*)[42].

Currently, many studies have used the chromatin immunoprecipitation (ChIP) assay to identify patterns of DNA-protein interaction. The principle behind this assay is that DNA-bound proteins are cross-linked *in vivo* to DNA, e.g., by formaldehyde fixation. Following fixation, the cells are lysed and the DNA is broken into pieces 0.2-1 kb in length by sonication. Subsequently, whole DNA-protein complexes may be immunoprecipitated using an antibody for the required TF and the DNA isolated. Regions of promoters are then analyzed by probing the immunoprecipated DNA with specific PCR primers. By combining this method with microarray technology, ChIP-chip, genome-wide maps of the *in vivo* interactions between DNA and DNA-binding proteins may be obtained[34-39]. The principle underpinning this assay is the hybridization of fluorescent labeled DNA fragments from a ChIP analysis and an appropriate reference to a DNA microarray slide. In this way a large number of possible targets for TFs may be analyzed simultaneously. This type of experiments have identified binding sites for the MYC, MAX, GATA1, E2F, and RB transcription factors in cultured cells[43-46]. The major disadvantages of ChIP based methods are the requirement for highly specific antibodies and the low mapping resolution due to the large size of the immunoprecipitated DNA fragments (0.2-1 kb). However, the advantage is that promoters which actually bind the investigated TF *in vivo* are identified.

The DNA mobility shift assay may be used to determine sequence motifs for which TFs have a large affinity. The principle behind this method is that a DNA fragment bound to protein move slower than naked DNA during gel electrophoresis. The fraction of the original DNA fragments that is shifted by the slower mobility is proportional to the binding affinity of the transcription factor to the specific DNA sequence. By using *in vitro* synthesized DNA fragments with different sequences it is possible to determine the most optimal sequence motifs for TF binding. A limitation of this method is that it only gives information about DNA-protein interaction but cannot be used to localize the area of contact between protein and DNA. To increase the resolution and to identify the exact interaction between protein and DNA, the DNAseI footprinting assay may be used. This method is similar to the DNA mobility shift method except that protein bound DNA

fragments are treated with small amounts of the DNAseI enzyme. This enzyme digests the DNA fragments except where protein has bound the DNA. After electrophoresis the digested DNA fragments will appear as ladders and nucleotide positions that have been bound to protein as blank areas, referred to as the footprint of the protein. These type of investigations have shown that a given TF may bind to dissimilar but related sequences, that together constitute binding motif. These experimentally determined binding motifs are then used to develop position weight matrices (PWMs).

Identification of TFBSs using Bioinformatic Tools

Currently, there are several *in silico* approaches to detect putative TFBSs in promoter regions of genes. In general, TFBSs are relatively short and vary between 5bp to 20bp in size and a given TF may bind to a variety of TFBSs. By analyzing several such putative TFBSs it is possible to estimate frequencies of nucleotides at a specific position, and these frequencies are then used to develop PWMs for a given TF binding profile or motif (pattern) sequences. The score of the PWM, sum of position-specific frequencies for each nucleotide, gives a weighted match to a given sequence site. These consensus representations of TF binding profiles are stored in databases like TRANSFAC[47] and JASPAR[48] which are available online. At present TRANSFAC and JASPAR contain 834 and 123 PWMs, respectively. In contrast to TRANSFAC, JASPAR only contains non-redundant TFBSs.

Several software tools have been developed to identify putative TFBSs in a given DNA sequence. Software tools such as MotifScanner[49], MATCH[50], and MatInspector[51] use PWMs to scan a given DNA sequence to map putative TFBSs. These software tools mainly differ in the type of algorithm used in calculating PWM scores. The MATCH and MatInspector use similar type of algorithms based on information vectors that uses two score values, matrix similarity score (MSS) and core similarity score (CSS) (conserved consecutive positions in a matrix) to measure the quality of a match between the PWM and the given sequence. The core of each matrix is defined as the first five most conserved consecutive positions of a matrix. The MSS is calculated using all positions of the matrix whereas CSS is calculated using the core positions only. The above two software tools differ in the number of PWMs that are used to scan a given sequence;

MATCH uses the TRANSFAC database whereas the MatInspector uses compiled matrices from published data including matrices from the TRANSFAC database. These approaches are dependent on PWMs and cannot derive novel regulatory elements. The limitation of these methods is that TFBSs are highly abundant in the genome due to their small size and therefore, the vast majority of detected TFBSs may not be functional. To overcome some of these limitations, several methods have been proposed to identify binding sites that show "significance" at some level. In most cases these approaches are based on probabilities of finding the binding site in a certain segment of the genome given a reference sequence. For a reference sequence a synthetic sequence or non-coding sequences may be used.

The obtained results are influenced by both the reference sequence and the size of the analyzed DNA segment. When large DNA segments are analyzed there will be a tendency to dilute the significance of TFBSs by noise, whereas in a shorter sequence a large number of false positives will be identified. A possible strategy to overcome some of these problems is to make use of evolutionary conserved sequences. It is known that the mutation rate in functional, including regulatory, elements is lower than in non-functional sequences[52]. Hence, conserved elements in promoter sequences of orthologous genes may be used as an approach to find functionally relevant TFBSs. This approach is known as phylogenetic footprinting. The conventional phylogenetic footprinting method is to align a given sequence set or orthologous sequences and then identify consensus sequence segments. This approach has proved successful for the discovery of regulatory elements for many genes, including ε-globin[53], cystic fibrosis transmembrane conductance regulator[54], and tumor necrosis factor-α[55]. The extension of this approach is to use PWMs to identify conserved TFBSs, which may reduce the number of false positives further. A prerequisite for phylogenetic footprinting is that promoter regions in orthologs show substantial regions of evolutionary conserved sequences, which may not always be the case. Moreover, this approach most likely misses species specific regulatory elements.

Identification of TFBSs based on Gene Expression Analyses

Transcription factors may, individually or in complexes, regulate groups of genes in a co-expressed fashion. This suggests that promoters of co-expressed genes may show

similarities with respect to the composition of TFBSs. Consequently, a possible approach to identify TFs important for gene regulation is to identify common or enriched TFBSs in promoters of co-expressed genes. To further reduce the number false positives this method may also be combined with phylogenetic footprinting. This approach will, however, only define TFBSs that are active in a given context i.e., resulting in the specific gene expression profile under investigation.

The major source for gene expression data is currently microarray studies that measure the expression of several thousands genes simultaneously[56]. Published gene expression datasets may be obtained from databases like Gene Expression Omnibus (GEO)[57], Stanford MicroArray Database (SMD), ArrayExpress, and the cancer specific ONCOMINE[58]. Even though co-expression may at first site be a simple concept it is highly dependent on the means and criteria used to establish co-expression[59-61]. It is common to identify genes that show similar behavior across series of microarray experiments by statistical means such as various clustering algorithms. However, a general problem is that the exact nature of the obtained gene clusters is dependent on the method used for cluster formation. Some algorithms, e.g., SOM and *K*-means clustering, are sensitive to initial parameter settings whereas e.g., HCA is sensitive to the agglomeration principle and distance metrics used. Hence, care must be taken when selecting means to define co-expression.

Several methods and tools have been developed to simultaneously analyze more than one promoter. The common principles of tools like TOUCAN[49,62], CONFAC[63], and oPOSSUM[64,65] is to first find high scoring TFBSs in individual promoters and then apply a principle to identify enriched or over-represented TFBSs. The TOUCAN software identifies high scoring TFBSs using MotifScanner and PWMs obtained from the TRANSFAC or the JASPAR databases. In a second step the software estimates the probability to find a given motif compared with so called background model. Background models are derived by using Markov Models from either non-coding regions of the genome or from the Eukaryotic Promoter Database (EPD)[66]. The CONFAC software starts by identifying the mouse ortholog to the investigated gene. The software then maps high scoring TFBSs in both promoters of the orthologous genes using the MATCH software and PWMs derived from the TRANSFAC database. TFBSs present on both

orthologs are then considered to be the part of the promoter. Hence, the CONFAC software uses the phylogenetic footprinting to assign significance to specific TFBS. The enrichment of a given TFBS in a set of promoters is then estimated by comparison with a set of randomly selected promoters.

The regulation of genes is usually determined by more than one transcription factor. Binding sites for such co-operative transcription factors are frequently spatially clustered and referred to as cis-regulatory modules (CRMs) or TFBSs modules[67-70]. For instance, both OCT4 and SRY are involved in the regulation of SRY target genes[71] and binds to module composed of OCT4 and SRY binding sites. Thus, by also evaluating the spatial location, and particularly identifying modules shared by several promoters, the significance of the finding may be considerably enhanced.

In summary, experimentally constructed PWMs makes it possible to design softwares that identify high scoring TFBSs in promoter regions. Their significance may be estimated using statistical methods, background models, or phylogenetic footprinting. One way to increase the significance of these findings further is to identify significantly enriched or shared TFBSs in promoter regions of co-expression genes. A further development is to only consider the evolutionary conserved binding sites or to search for shared TFBSs modules.

EPIGENETIC CHANGES IN CANCER

Epigenetics is defined as phenomenon that controls the functional state of DNA without altering the DNA nucleotide sequence. Epigenetic changes may involve alterations of chromatin or of DNA in the form of DNA modifications. Chromatin is formed by a complex of DNA and histone proteins. The function of chromatin is to package DNA into a smaller volume, to protect the DNA, and to facilitate mitosis and meiosis. The structure of chromatin plays an important role in DNA replication, DNA repair, and also serves as a mechanism to control gene expression. Changes in chromatin structure may be affected by DNA methylation and histone modifications[72,73]. It is also well known that these mechanisms are involved in the developmental programme of embryogenesis[74]. DNA methylation and histone deacetylation usually result in compact and inactive chromatin, which makes DNA inaccessible for other proteins such as transcription factors. In

addition DNA methylation is responsible for X-chromosome inactivation[75] as well as for genomic imprinting i.e., the phenomenon that different methylation patterns in egg and sperm may affect the expression of specific genes based on the parental origin. Moreover, imprinting defects cause several human diseases such as the Beckwith-Wiedemann (BWS), the Prader-Willi (PWS), and the Angelman (AS) syndromes[76]. In addition, altered methylation patterns and age-related methylation changes are known to contribute to increased cancer risk[77-80]. Overall, epigenetic mechanisms are essential for maintaining normal cellular processes.

Fundamentals in DNA Methylation

DNA methylation occurs mostly by adding methyl groups on cytosines in CpG dinucleotides and is achieved by the methyltransferases DNMT1, DNMT2, DNMT3a, and DNMT3b[81,82]. It is believed that DNMT3a and DNMT3b are involved in the initial methylation of CpG dinucleotides, a process called *de novo* methylation, whereas DNMT1 is responsible for maintaining the produced methylation patterns[83-85]. After DNA-replication the newly replicated double-stranded DNA contain methylated C's on only one of the strands, resulting hemi-methylated CpG sites. These sites are then recognized by DNMT1 that subsequently methylate the remaining C's in the CpG sites, thus maintaining the original methylation pattern[85]. Although the DNMT3 enzymes predominantly methylate CpG sites, these enzymes are believed to methylate 15-20% of cytosines that are found at sequences other than CpGs in embryonic stem cells, including CpAs, CpTs and CpCs[86]. The spontaneous deamination of methylated cytosines will yield thymine and a T:G mismatch. If the T is not replaced by a C before the following round of DNA replication the TpG will be fixed in the genome[87]. Hence, this mechanism is believed to be responsible for the observed under representation of CpG dinucleotides in mammalian genomes[86,88-90]. The remaining CpGs in the mammalian genome is distributed in a non-uniform fashion. Genomic regions rich in CpG dinucleotides are called CpG islands and are generally 0.5-4kb in length[91]. According to Gardiner-Garden and Frommer[92], a genomic region is considered as a CpG island if it has a minimum G+C content of 50%, and a minimum CpG (obs)/CpG (exp) of 0.6 in a 200bp window length, and the sequence length is at least 500bp. There are about 15,500 predicted CpG islands

in the human genome and most of them are present in the promoter regions. Hence, almost half of the genes have CpG islands[93-95]. According to Weber et al.[96] gene promoters may be classified into three classes based on CpG ratio, GC content, and the length of the CpG rich region. These promoter classes are termed high CpG-content promoters (HCPs), intermediate CpG-content promoters (ICPs), and low CpG-content promoters (LCPs). Weber et al. have shown that in normal somatic cells the vast majority of HCPs are hypomethylated. For example, *CASP2* and *BCLAF1* contain HCPs and show 7% and 0% methylation of their respective promoters. In contrast, promoters with LCPs and ICPs were shown to be heavily methylated. For example, the promoter regions of *MT1B*, *OTOR*, *EHIF*, and *FGF6* were shown to have 66%, 71%, 90%, and 71% methylation of their respective promoters. Moreover, ICPs in somatic cells were believed to be preferential targets for *de novo* methylation. The hypothesis put forward by these authors suggests that the observed differential methylation patterns are caused by protection from *de novo* methylation as a consequence by the high CpG density. In addition to this, it is also well known that in normal cells DNA methylation occurs in repetitive sequence elements such as long interspersed transposable elements (LINES) and short interspersed transposable elements (SINES)[97].

DNA methylation may affect gene transcription in different ways[98]. The methylation of promoters may physically impede the binding of DNA-binding transcriptional factors (DBTSs), thus blocking transcription[98]. Furthermore, methylated promoters may be bound by a family of proteins known as Methyl-CpG-binding domain (MBD) proteins. When bound to the promoter these proteins recruit additional proteins such as histone deacetylases (HDACs) and other chromatin remodelling proteins, that modify the histones resulting in compact and inactive chromatin[99-101]. A total of eleven MBD-containing genes exists in the human genome, of which *MBD1*, *MBD2*, *MBD3*, *MBD4* and *MeCP2* are well characterized[102-104]. MeCP2 contains a unique protein domain structure that bind to DNA sequences that contain stretches of A/T bases ([A/T]>=4) adjacent to methyl-CpG whereas MBD2 only requires methyl-CpG[73,105-107]. Studies have shown that MeCP2 repress transcription by recruiting the Sin3-HDAC1 co-repressor complex[108]. However, repression may also occur without recruiting additional proteins[109,110]. In contrast to the general inhibitory affect of DNA methylation, there are

reports that DNA methylation may increase gene expression. For instance, the binding of regulatory proteins to methylated CpG residues has been shown to increase expression of *IL-8*, *EGR2*, and *HLA-DRA*[111-113].

DNA Methylation and Cancer

In normal cells CpG islands are usually unmethylated whereas in cancer cells they frequently are hypermethylated[77,80,98,114]. Recent studies have shown that this *de novo* methylation is caused by the reactivation of *DNMT3a* and *DNMT3b* expression that normally are expressed only during embryogenesis[84]. There is a growing list of genes that show suppression of their function due to promoter hypermethylation in various cancers[98,115-120]. Genes known to be affected by DNA methylation include *VHL* in renal cell cancer, *BRAC1* and the *HOXA* gene cluster in breast cancer, and *APC*, *CDKN1C*, *RARB2*, *RASSF1A*, *RUNX3*, and *SFRP1* in bladder cancer[121-126]. The methylation in CpG-poor promoters may determine the transcriptional status of a gene by blocking TFs that are sensitive to the methylation of CpG sites in their corresponding binding sites. For example, it has been shown that methylation of CpGs within the CTCF-binding site, which controls the expression of the *Igf2* gene, eliminates binding of CTCF and thus inhibits gene expression.

According to Knudson's two-hit hypothesis, the loss of function for a tumor suppressor gene is not obtained until both alleles are inactivated[7]. The most frequent mechanism to obtain complete inactivation of a gene is to mutate one allele and to delete the wild-type allele. In rare occasions homozygous deletion of tumor suppressor genes may be seen. An alternative mechanism to mutation and losses is inactivation by promoter methylation. This alternative mechanism has been shown by several studies, for example, in colorectal cancer the tumor suppressor gene *CDKN2A* frequently show mutations and the wild-type allele promoter methylation[127] and in hereditary forms colorectal cancer the mutated version of *hMLH1* may be inherited whereas the remaining normal allele is methylated[128]. Consequently, aberrant promoter methylation may frequently substitute for loss of heterozygosity (LOH) or mutations as a "second hit" to inactivate tumor suppressor genes.

Cancer cells generally show a decrease of CpG methylation in non-promoter CpG islands and other CpGs, referred to as global DNA hypomethylation[129]. It is known that hypomethylation is associated with chromosomal instability[129,130]. Hypomethylation may also occur in repetitive DNA sequences[131-133] and induce transcription of the otherwise non-transcribed repetitive sequences. In addition, hypomethylation may affect CpGs in promoters of genes resulting in their expression and, in the case of proto-oncogenes, possibly to increased cell proliferation[134]. Therefore, global DNA hypomethylation may also be involved in cancer initiation and progression.

Principles of DNA Methylation Analysis

The study of DNA methylation has taken the advantage of the fact that epigenetic changes caused by DNA methylation are reversible i.e., an epigenetically silenced gene can be reactivated. This has made it possible to study aberrant gene expression due to DNA methylation in various cancers. For these purposes, drugs like 5-aza-cytidine (5-aza-CR) or its deoxyribose analog 5-aza-deoxycytidine (5-aza-CdR)[135], and Zebularine[136] have been used. Both 5-aza-CR and 5-aza-CdR are incorporated into a DNA and form a covalent adducts with cellular DNMTs, thereby inhibiting their activity[135]. These drugs are the most widely used for studying DNA methylation both in preclinical and clinical settings[137] and 5-aza-CR has been approved as an anti-cancer agent in the USA[137-139]. In addition, inhibition of histone deacetylases (HDACs), which remove acetyl groups from histones, may lead to remodeling of chromatin to a transcriptionally active conformation. Several HDAC inhibitors for e.g., Trichostatin A, sodium phenylbutyrate, have been used to study the effects of chromatin remodeling and are currently used in clinical trials[140,141]. The use of the above drugs in combination with microarray technology has proven to be an efficient strategy to identify cancer-related genes that are repressed by epigenetic mechanisms[98].

The most common way to identify methylated CpG in a DNA sequence is to pretreat the DNA with bisulfite. By this treatment all non-methylated cytosines are deaminated to yield uracil without affecting the methylated cytosines. Using this conversion principle, several methods have been developed in order to study the exact methylation status of specific sequences e.g., Methylation-specific PCR (MSP)[142],

Combined bisulfite restriction analysis (COBRA)[143], Methylation-sensitive single-nucleotide primer extension (MsSnuPE)[144], MethylLight[145], and Methylation-specific melting curve analysis (MS-MCA)[146]. The aim for most of these methods is to detect bisulfite-induced sequence differences between treated and untreated DNA. Among these methods, MSP is the most widely used since it is rapid and suitable for screening several samples for methylation of specific CpG sites. Limitations of this method are that only one CpG is analyzed at a time, the method is not quantitative, and only previously known CpG sites may be analyzed. In contrast to this, Frommer et al.[147] have shown the usefulness of sodium bisulphite sequencing technique in conjunction to PCR amplification to identify every methylated cytosine residue within a given fragment. The sequencing can be performed directly after PCR amplification or after cloning of the PCR fragments, known as the "cloned bisulfite sequencing" technique. This method has been shown to give the highest resolution and in conjunction with sequencing of several individually cloned PCR fragments also to give quantitative information[148]. All the above methods have been successfully used to analyze individual promoters. In addition to this, there are approaches to study global methylation, also termed epigenomics. Such methods include restriction landmark genomic scanning (RLGS)[149], methylated CpG island amplification-representational difference analysis (MCA-RDA)[150], methylation-sensitive representational difference analysis (MS-RDA)[151], methylation-sensitive arbitrarily primed PCR (MS-AP-PCR)[152], and differential methylation hybridization (DMH)[153,154]. The most common principle for these approaches is to digest genomic DNA with methylation-sensitive enzymes, for e.g., *Not*I, *Sma*I, *Hpa*II, and *BstU*I, or methylation-insensitive enzymes, for e.g., *Xma*I, *Msp*I, and *Mse*I. The produced fragments are then either hybridized to arrays with immobilized CpG containing fragments, or analyzed by PCR techniques, to reveal global methylation patterns. Yan et al.[154] used a microarray containing 7776 short GC-rich tags to identify differential methylation patterns in breast tumors and identified a group of methylated CpG islands that segregated breast tumors based on their hormone-receptor status. Several additional platforms have been developed for the study of global methylation patterns[155-157] such as promoter specific methylation arrays, and whole genome CpG island specific methylation arrays. These arrays have

made it possible to analyse methylation status in promoters of thousands of genes in a large number of tumor samples[158,159].

MICRORNA

MicroRNAs (miRNAs) are short single-stranded non-coding RNA sequences, 20-22nt in length. Recent studies have shown that miRNAs are a substantial part of the transcriptional output of the genomes of plants and animals. MiRNAs have also been shown to play an important role in gene regulation at post-transcriptional and translational levels via an antisense RNA-RNA interaction. They were first discovered by Lee et al. in 1993 that identified two small transcripts of approximately 22 and 61nt in C. elegans and suggested that these non-coding transcripts regulated the *lin-14* gene[160]. This suggestion was based on the finding that the shorts RNAs contained complementarily sequences to the 3' UTR of the *lin-14* mRNA. Subsequently, similar miRNAs were shown not only to play important role in cellular process of plants but also in mammalian cells. The expression of miRNAs influences several important cellular processes such as growth and differentiation. For example, the mammalian homologues of *lin-4* and *let-7* have been shown to control cell proliferation in human cell lines[161,162], mir-181 has been shown to be involved in the differentiation of haematopoietic cells towards the B-cell lineage[163], and mir-375 in pancreatic islet-cell development and the regulation of insulin secretion[164]. In addition, miRNAs are also involved in proper organ formation such as mir-196 in mammalian limb patterning[165] and mir-1 in mammalian heart development[166].

The biogenesis of miRNA has recently been revealed[167]. Similar to conventional mRNAs, miRNAs are transcribed by polymerase II as long primary transcripts known as pri-miRNA with a cap and poly (A) tail[168]. These pri-miRNAs are then processed in the nucleus by the RNA III-type endonuclease DROSHA and the double-stranded-RNA-binding protein DGCR8, to short 70nt stem-loop structures known as pre-miRNAs[169-171]. The pre-miRNAs are then exported into the cytoplasm by exportin 5 and undergo further processing in which stem-loop pre-miRNAs are cleaved into two complementary short RNA molecules by the RNAase III-type endonuclease DICER, to yield the mature miRNAs[167,172-174]. One of the mature miRNAs is integrated into the RNA induced-silencing complex RISC and base pair with their complementary mRNA molecules. This

induces mRNA degradation by the argonaute proteins, the catalytically active members of the RISC complex[175]. After this process, the miRNA remains intact within the RISC complex and guide the recognition and destruction of additional mRNAs. Gene expression may also be affected by translational repression by the mature RISC complex. A near-perfect complementary between miRNA and the target mRNA is believed to be prerequisite for RISC-mediated cleavage but not for translational repression[176]. In animals, it is believed that the degree of complementarity between the miRNAs and mRNAs is low and therefore that translational repression is more prevalent in animals than in plants[177-179].

Identification of miRNA

Currently there are 685 human miRNAs (miRBase release 11.0)[9], but recent reports have estimated that up to 1000 more miRNAs may exists in the human genome[180,181]. A standard procedure to identify miRNAs is to use size fractionation in polyacrylamide gels, molecular cloning, and sequencing. The majority miRNAs have been identified in this way. There are several computational softwares to identify potential miRNA genes based on the initial sequencing of mature miRNAs e.g., miRseeker[182], miRScan[183], ProMiR[184], miRNAMap[185], and miRAlign[186]. These softwares employ different approaches and evaluate features like miRNA secondary structure, phylogenetic conservation, thermodynamic stability of hairpins, and similarities to known miRNAs. The mirScan software classify a short sequence as an miRNA based on its similarity to experimentally confirmed miRNAs, whereas the miRNAMap software combines thermodynamic stability and conservation of secondary structure to predict miRNAs. Recently, several alignment-based methods have been developed that search for genomic sequences that are homologs for known miRNAs at both sequence and structural levels. Many miRNAs are found to be localized to intergenic regions or within introns of either protein-coding or non-coding mRNA transcripts[187-189]. miRNAs may be found both as single genes located in some distance from other miRNA genes, or in miRNA gene cluster that may share the same promoter[190].

miRNA Promoters

Although several studies have been focused on elucidating the mechanism of miRNA target gene regulation, very few investigations have attempted to elucidate the regulation of the miRNA genes themselves. Some miRNAs form clusters that seem to be controlled by one common promoter and are transcribed as polycistronic RNAs from which more than one mature miRNAs may be produced, whereas others are transcribed as single units. To date, studies have provided experimental evidence that miRNAs are transcribed by RNA polymerase II, suggesting that miRNA transcription may be regulated by similar mechanisms as for protein-coding genes[191-193]. Zhou et al. investigated 109 human miRNAs to define a core promoter. They found that 33% of the human miRNA gene promoters contain TATA boxes and a CT-motif also seen in other species. Furthermore, they identified 5 additional motifs specific for human miRNA promoters[194]. There may be additional promoter motifs for miRNA regulation. For instance, the promoter of miRNA gene cluster miR-23a~27a~24-2 lack TATA boxes, initiator elements, and the TFIIB recognition element but a deletion of a relatively GC-rich regions in its promoter results in a moderate but reproducible reduction in expression[192]. miRNAs also contains CpG islands, however at lower frequency than mRNA promoters[194], and methylation in these islands may result in alterations of the respective miRNA expression. For example, Saito et al.[195] showed that DNA methylation can affect the expression of mir-127. The presence of miRNAs in intronic regions of protein coding as well as in non-protein coding genes indicate that miRNAs may be co-transcribed with other genes[193,196]. Consequently, mechanisms that affect the transcription of the protein-coding gene, including DNA methylation, may affect miRNA expression.

miRNA Target Prediction

As mentioned, miRNA regulates mRNA via RNA-RNA interaction and thus miRNAs sequences are complementarily to the target mRNAs. Several softwares designed to identify putative target mRNAs are based on these principles. The search for miRNA targets is restricted to evolutionarily conserved sites in almost all approaches[197,198], e.g., TargetScanS[199], and PicTar[200]. In brief, first evolutionarily conserved sites within mRNA sequences are identified and then a search for perfect Watson-Crick miRNA-target

sequence complementarity matches is preformed, which typically are located in the 5'-end of the miRNA and are called "seed" sites[197]. Experimentally, miRNA target predictions are mainly based on systematic target-site mutation experiments[197,198]. These experiments has indicated the existence of at least two-classes of miRNA targets, target sites with perfect Watson-Crick complementarity to the 5'-end of miRNA and those with imperfect 5'-end matches. In the latter case miRNA-mRNA interactions are stabilized by additional base pairing in the 3'-end of the miRNA. There are several miRNA target databases publicly available that use different methods to define target genes, e.g., miRanda[201], miRBase[9], and TarBase[202].

miRNA and Cancer

Recent studies have shown that miRNAs and miRNA processing are linked to cancer and that miRNAs may act either as tumor suppressor genes or as oncogenes. Components involved in maturation of miRNAs such as *DICER* and *DROSHA* have been shown to be down-regulated in tumors[203]. DICER has been shown to be involved in heterochromatin maintenance and centromeric silencing, and low expression of this protein is associated with poor prognosis and genomic instability[204,205]. The importance of DICER in normal cellular function is also shown by the fact that embryonic stem cells fail to differentiate normally in the absence of DICER[205]. This finding is consistent with studies of global miRNA expression in tumors, which usually show lower levels of expression than the corresponding normal tissues. Furthermore, it has been shown that miRNA expression profiles may be used to classify cancers. Lu et al.[206] used bead-based flow cytometric miRNA expression profiling to analyse 217 mammalian miRNAs in 334 samples including 8 different cancer types as well as normal samples. Apart from the finding that miRNAs are down regulated in tumors, Lu et al. found that tumors were grouped together according to their embryonic lineage; hierarchical clustering placed tumors of epithelial origin such as colon, liver, pancreas and stomach, in one cluster, and tumors of haematopoietic origin in a different cluster. Furthermore, miRNA expression profiles could identify subtypes of tumors that could not be identified using mRNA profiles[206]. The down-regulation of miRNAs in tumors has also been found in other studies. For example, down-regulation of mir-15a and mir-16-1 is seen in leukemias and

lymphomas[207] and results in increased expression of *BCL2*. Significant down-regulation of mir-143 and mir-145 has also be seen in colorectal tumors[208] and let-7 is significantly down regulated in lung cancer, resulting in increased expression of the RAS family of oncogenes[209]. As about 50% of all annotated human miRNAs are located in or in close proximity of fragile sites[210] or in genomic regions that are frequently altered in tumors, it has been suggested that the frequent down-regulation of specific miRNA expression is associated with genomic alterations of these regions[210]. Examples of such miRNAs are the let-7 family members, let-7a-1/let-7f-1/let-7d, that are located at 9q22, a region frequently lost in urothethial carcinoma[210], mirs -143 and -145 that are located within a fragile site at 5q32-33 and are frequently down regulated in colon and breast-cancer patients[211,212], and mirs -15a and -16-1 located at 13q14 recurrently lost in chronic lymphocytic leukemia (CLL)[213,214]. Even though only a moderate number of human miRNAs contains CpG islands there are reports showing that miRNAs may be regulated by epigenetic mechanisms[215,216]. Examples of such miRNAs are mir-9-1 that is inactivated due to aberrant hypermethylation in the early development of breast cancer[217], mir-127 that is epigenetically silenced in prostate and bladder cancers[195], and mir-124a that is down-regulated by epigenetic mechanisms in colon cancer[218]. Furthermore, it has been shown that down-regulation of mirs -127 and -124a contribute to the up-regulation of oncogenes such as *BCL6*[195] and *CDK6*[218], respectively. Hence, these investigations indicate that miRNAs may function as tumor suppressors.

There are reports showing that up-regulation of miRNAs also may contribute to carcinogeneis. For example, Chan et al.[219] have shown that mir-21 is up-regulated in glioblastoma to 5-100 fold higher levels than in normal tissue, and by antisense studies that mir-21 affects the growth of glioblastoma cells by inhibiting apoptosis. Consequently, mir-21 has an anti-apoptotic function in glioblastoma cells. In B-cell and mantel cell lymphomas the mir-17-92 cluster, located at 13q31, is frequently amplified and show expression levels that correlate with gene copy numbers. The importance of the mir-17-92 cluster is shown by the finding that 65% of the lymphomas show increased expression of members in this cluster[220,221]. In addition, miRNAs may also be activated by hypomethylation that results in increase of miRNAs expression. For instance, it has been shown that up-regulation of let-7a-3 by DNA hypomethylation can contribute to

human lung cancer[222]. Hence, there is sufficient evidence that miRNAs may also act as oncogenes.

BLADDER CANCER

Bladder cancer is the fifth most common cancer in men and is significantly more common in patients over 60 years of age. The highest occurrences of bladder cancer are found in industrialized countries. Risk factors strongly associated with this tumor type include tobacco smoking, occupational exposure to aromatic amines, consumption of arsenic-laced water, radiation therapy of neighboring organs, and therapeutic use of alkylating agents. In the western world bladder cancers are mostly (90%) of urothelial origin and designated urothelial carcinomas (UC).

Urothelial Carcinoma (UC)

Histopathologically, urothelial tumors are grouped into different stages Ta, Tis, T1, T2, T3, and T4, based on how deep the tumor has grown into the bladder wall. The grading system, G1, G2, and G3, are mainly based on morphological differentiation[223,224]. Ta tumors show a papillary growth and carcinoma in situ (Tis) are flat lesions, strictly confined to the epithelial layer of the bladder wall. The majority of Ta tumors are of low to moderate grade (G1-G2) whereas Tis usually is of grade 3 (G3) and poorly differentiated. Invasive tumors are T1, that involves lamina propria, T2, that invades the superficial muscle, T3, that invades the deep muscle, and T4, that involves surrounding organs.

Bladder tumors that are of stages Ta, T1, and Tis are frequently grouped together as superficial bladder cancers[225]. These tumors are mainly treated by transurethral resection and by intravesical chemotherapy using mitomycin or by treatment with Bacillus Calmette-Guérin (BCG)[226]. A characteristic of bladder cancers is the high frequency of tumor recurrence; more than half of the patients with Ta tumors show recurrences[227,228]. Hence, follow-up cystoscopies are important in managing the disease. Despite the high recurrence rate, the disease course for many of these patients is highly favorable, and only few develop a subsequent muscle-invasive disease. Muscle invasive tumors, T2 or above, are highly malignant and generally demand more aggressive

therapeutic management. Treatment of these tumors frequently includes surgical removal of the bladder or a combination of radiation therapy and chemotherapy. Even though conventional prognostic factors such as tumor stage/grade, size, and multiplicity may be useful the disease course in individual patients are unpredictable. Therefore, an improved understanding of the molecular changes in bladder cancers is important. The identification of molecular markers that defines subgroups of patients may be important when selecting treatment modalities. In addition, such markers may also serve as targets for the development of novel drugs or treatment strategies.

Genetic and Chromosomal Changes in UC

Several molecular studies have shown that UC is a complex disease associated with several genetic lesions[229,230]. Activating mutations in *FGFR3* and *HRAS*, and inactivating mutations in *TP53* and *CDKN2A* are some of the most frequent genetic changes in UC. *FGFR3* is mutated in more than 30% and *TP53* in more than 50% of the cases[231,232]. In addition, mutations in genes *CDH1*, *CTNNB1*, *KRAS*, *MSH6*, *NRAS*, and *PTEN* are also known to occur in UC. *FGFR3* (fibroblast growth factor receptor 3) belongs to the tyrosine kinase receptor family *FGFR1-FGFR4*[233]. These receptors regulate various cellular processes including cell growth, differentiation, and angiogenesis. They are activated by the fibroblast growth factors (FGFs), which are heparin-binding proteins, and via heparin sulfate proteoglycan (HSPG) induces receptor dimerization and activation of the receptor molecule[233]. The *FGFR3* mutations predominantly occur in superficial bladder cancers and are associated with low-grade tumors[234-237]. Furthermore, *FGFR3* mutations have been identified in 75% of urothelial papilomas, a benign urothelial lesion[238]. This suggests that the constitutive activation of *FGFR3* is an important event for bladder tumorigensis. Interestingly, nearly all gain-of-function mutations in *FGFR3* have previously been found to be associated with several autosomal dominant skeletal disorders[239,240]. Unlike *FGFR3* mutations, *TP53* alterations are associated with adverse tumor phenotypes and poor prognosis[241].

Several chromosomal aberrations have been associated with UC. Monosomy 9 are frequently seen in UC and is associated with early stage tumors[242] whereas gains of 1q, 3q, 5p, 8p, 17q, and 20q and losses of 5q, 6q, 8p, 11p, and 11q are involved in advanced

tumors[243]. Molecular cytogenetic studies using CGH have corroborated the frequent loss of 9p and indicated that the most frequently deleted region is 9p21[244-247]. Furthermore, the frequent loss of 9p21 has also been shown by several LOH studies[248-256] and to involve the *CDKN2A/2B/p14*^{ARF} locus[257]. So far no definite tumor suppressor genes have been identified on the q arm of chromosome 9.

THE PRESENT STUDY

AIMS OF THE STUDY

Article I

To explore the possibility to use global gene expression profiles to identify important transcription factors that determine gene signatures important for cancer development.

Article II

To identify genes that show altered gene expression in bladder cancer due to promoter methylation and to precisely map methylated CpG sites in their promoters.

Article III

To use global miRNA expression to identify miRNAs important for bladder cancer development.

MATERIALS & METHODS

Microarray Data Sets

In Article I, two cancer related microarray data sets were used. The AML (acute myeloid leukemia) dataset described by Bullinger et al.[258], downloaded from the Gene expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and a time series data set, the serum induced gene expression data, described by Chang et al.[259], downloaded from the Standford Microarray Database (SMD) (http://smd.stanford.edu/index.shtml). In Article II, the gene expression analyses were performed on oligonucleotide arrays printed with 70-mer oligos from the OPERON v 3.0 set corresponding to 11,132 unique genes. This array was obtained from the Swegene DNA microarray resource centre at Lund University.

Cell lines

In Article II eight bladder cancer cell lines were used of which five 247/01, 3396/97, 444/98, 509/98, and 770/96 were low passage cultures from primary bladder cancer cells and three, HTB-1 (J82), HTB-4 (T-24), and HTB-9 (5637) were established cell lines obtained from ATCC. These cells were grown in 1:1 DMEM: F12 medium supplemented with 10% serum and antibiotics. Cells were treated with the demethylating agent 5-aza-2'-cytidine for 48hr followed by cultivation in media complemented with Zebularine to maintain the demethylated state of the cells.

Patients and Tissues

For the miRNA studies (Article III), 34 urothelial tumor samples collected from patients undergoing transurethral resection at Lund University Hospital, Sweden, between 2001 and 2004 were used. Tumors were collectecd by cold-cup biopsies from the exophytic part of the bladder and kept at -80°C until further use. Histopathological staging and grading were reviewed according to the 2002 TNM[260] and 1999 WHO[261] classification systems by an experienced pathologist. The study included 7 Ta, 10 T1 and 17 T2/3 tumors. Of these 10 were low grade (G1 and G2) and 24 high grade (G3) tumors. This study was approved by the Research Ethics Committee of Lund University, and an informed consent was obtained from the patients.

miRNA arrays

In Article III, miRNA expression analyses were performed on the 96-sample array matrix (SAM), Illumina miRNA Expression Profiling Assays, containing 470 human miRNAs and 265 potential miRNAs obtained by computational methods. The experiments were carried at the Swegene Centre for Integrative Biology at Lund University (SCIBLU Genomics).

Clustering Algorithms

In Article I, two types of clustering algorithms were used, the Quality Cluster Algorithm (QTC)[262] and the Pavlidis Template Matching (PTM)[263] algorithm. The QTC algorithm works by forming a candidate cluster of the first gene and grouping genes with the highest correlation iteratively in a way that minimizes the cluster diameter d (d; defined by 1-Pearson correlation), until no further genes may be added without exceeding a predetermined d-value. This procedure is performed with all genes in the data set as a seed. The largest cluster is then retrieved and the procedure repeated excluding the genes selected for the preceding cluster. This makes sure that the largest and most coherent clusters of genes are formed. The PTM algorithm was used to identify genes that have similar expression profiles with genes that code for DNA binding transcription factors. In Article III, hierarchical cluster analysis (HCA) was used to group the tumor cases using Wards algorithm or average clustering as agglomeration principles and 1-Pearson correlation as distance measure.

Promoter Sequences

In Article I and II, promoter sequences were retrieved from the UCSC genome browser database and RefSeqs information was used to identify transcription start sites for the corresponding genes. For genes with no RefSeqs information, the most 5'mRNA sequences were used. For phylogenetic footprinting, orthologous mouse genes were downloaded from the JaxOrtholog table and the HomoloGene obtained from UCSC and NCBI, respectively. CpG island sequences were retrieved from the promoter sequences using the criteria described by Gardiner-Garden and Frommer[92].

TFBS Mapping

In Article I, TFBSs information was retrieved from the TRANSFAC database version 8.3 that included 243 vertebrate PWMs. To map TFBSs in the promoter sequences and to compute significances, two approaches were used, one analytical and one based on phylogenetic footprinting. For these purposes the TOUCAN and the CONFAC softwares were used, respectively. We used the TOUCAN software to screen 700bp (-600/+100bp) and the CONFAC software to screen 4000bp (-3000/+1000bp) of the respective promoters. In Article II, high scoring and evolutionarily conserved TFBSs were mapped to the promoter regions using TFBSs information from the TRANSFAC database version 9.1.

Promoter Clustering

In Article I, promoter sequences of co-expressed genes were transformed into a string of TFBSs motifs. These strings were transformed to a matrix format, in which columns corresponds to specific gene promoters and rows to presence or absence of individual TFBS. These matrices were then used to generate similarity matrices using the Jaccards algorithm. Finally, to obtain genes with similar promoters the similarity matrices were used for hierarchical cluster analysis using 1-S (S: Jaccards similarity value) as a dissimilarity measure and the Wards algorithm for cluster formation. The advantage of using the Jaccards algorithm, in contrast to Euclidian distance or Pearson correlation, is that the absence of binding sites in two promoters is not considered as an indication of similarity.

Bisulfite Sequencing

For the methylation studies (Article II), DNA was extracted from eight bladder cancer cell lines and treated with bisulfite. Promoter regions of interest were then amplified using PCR, sub-cloned, and sequenced. This type of sequence analysis gives the methylation status of every cytosine present in the target sequence. To achieve high resolution the methylation pattern in 12 randomly pick clones were determined. This method, “cloned bisulfite genomic sequencing” results in both qualitative as well as quantitative information.

RESULTS & COMMENTS

ARTICLE I

(Promoter clustering greatly enhances the identification of common and significant TFBSs in co-expressed genes)

Transcription factors may regulate the expression of several genes in a co-expressed fashion and co-expression may be studied using microarray experiments. One challenge is thus to use microarray data to identify transcription factors that are responsible for the co-regulation of genes. As transcription factors bind to the promoter regions of genes, a logical approach to identify important transcription factors would be to analyse the promoter regions of co-expressed genes. Several methods have been developed for these purposes, most of which are based on the identification of TFBSs that are over-represented in promoters of co-expressed genes. A limitation of these methods is that they cannot account for situations in which two or more groups of genes show co-expression due to regulation by more than one set of transcription factors acting in parallel. The reason is that any over-representation of TFBSs in a sub-set of genes would go undetected due to the presence of more than one regulated set of genes. This prompted us to introduce a promoter clustering principle that identifies similar promoters based on TFBSs profiles to be used before the identification of over-represented TFBSs. The aim of Article I was to apply this method to identify transcription factors that may be involved in establishing expression profiles seen in AMLs and during serum response in fibroblasts.

The original data was analyzed as follows. The QTC clustering algorithm was used to identify co-expressed genes, promoters of co-expressed genes were then retrieved and high scoring TFBSs identified. Based on the TFBSs patterns in the promoter sequences, the co-expressed genes were grouped into sub-clusters based on promoter similarities. Evolutionary conserved TFBSs were then identified within individual promoters and enriched TFBSs were identified by comparison with a reference set of promoters. Significantly enriched TFBSs that were present in the majority of promoters were then considered as good candidates for causing the respective expression signatures.

In the AML dataset a total of 13 QTC clusters were identified that were divided into 26 sub-clusters using promoter clustering. Of these, we were able to identify enriched evolutionarily conserved binding sites in 17 sub-clusters. The obtained results suggest that FOXO4, MYCMAX, and CMYB may be important to establish gene signatures in AMLs. In the time series data a total of 13 clusters divided into 43 sub-clusters using promoter clustering were identified. Of these, 26 sub-clusters showed enrichment for evolutionarily conserved TFBSs that included binding sites for MYC, SRF, TCF4, and E2F. An important cluster was composed of the genes *MCM5*, *CDC2*, *HELLS*, *USP1*, *FOXM1*, *RAMP*, *ANP32E* and *E2F1*, in which evolutionarily conserved binding sites for E2F transcription factors were identified. Importantly, all the identified genes are known to be regulated by E2F1 and several of the binding sites detected in the *E2F1*, *MCM5* and *CDC2* promoters have previously been identified experimentally. In addition, one gene cluster expressed during the early stages of serum induction showed enriched and evolutionarily conserved binding sites for MYC, including the *MYC* gene itself. Half way through the induction process a second gene cluster is induced that also shows enrichment for evolutionarily conserved MYC binding sites, indicating that these latter genes may be induced by the expression of the previous cluster. Furthermore, the second cluster, that also included the *TCF4* gene, showed enrichment for TCF4 binding sites. Interestingly, TCF4 binding sites were enriched in a late expression gene cluster indicating that genes in this cluster are induced by the expression of the previous gene cluster containing *TCF4*.

Comments

Crucial steps in the analysis of promoter sequences are to determine the TSS and the extension of the promoter sequences to be considered for the TFBSs analysis. Unfortunately, this information was, and still is, limited for most genes. In the present study we used RefSeqs or the most 5'-end of mRNA sequences to define TSS. An alternative would have been to use the DBTSS (Database for Transcription Start Sites) that contains a unique collection of precise and experimentally determined 5'-end sequences of full-length cDNAs. However, at the time of the present study the number of TSS in this database was limited. An inherent problem in promoter analyses is to

determine the extension of promoters, which is usually not known before hand. One strategy has been to use 500-700bp upstream of the putative TSS. Even though this segment is most likely to be part of all promoters, several promoters are known to be extensively larger. Hence, selecting such a small size would limit investigation to only a small proportion of the promoter in many cases. An alternative approach has been to analyse several kb upstream of the TSS and to limit the analysis to evolutionarily conserved regions. An advantage of the latter method is that binding sites present at some distance from the TSS are allowed and that the analysis is limited to evolutionarily conserved DNA sequences, i.e., "functional regions". We choose to make use of the latter principle for our analyses. In our investigation the initial sizes of the genes clusters formed by the QTC were fairly large. However, in the final steps the number of genes had in some cases been reduced due to the lack of ortholog information. Moreover, only gene clusters in which the genes on average contained 600bp or more of evolutionarily conserved sequences were amiable for analysis. The dependence of evolutionarily conserved sequences infact made the analysis of some clusters impossible. When estimating the enrichment of TFBSs the reference to be compared with is critical. In the present investigation, we based our analysis on the CONFAC software. This software uses a predefined set of genes as reference. Even though several such sets may be used, they all represent individual samples which may have a biased distribution of binding sites. Therefore, since the publication of Article I we have developed a new strategy to estimate significant enrichment (S Veerla; unpublished). In this strategy, we resample 10,000 times the same number of promoters that are present in the investigated gene clusters from all the promoters in the genome. Based on this resampling procedure we estimate p-values for obtaining the observed number of binding sites by chance in the given gene cluster. We believe that this strategy will produce much more stringent and reliable estimates of enrichment. Nonetheless, using the available information the method described in Article I proved to unveil a number of biologically relevant patterns in the experimental systems. Even though this approach remains at the level of statistical inferences we believe that the obtained results may be good starting points for experimental validation.

ARTICLE II

(Epigenetically controlled genes in bladder cancer)

In this study, our aim was to identify epigenetically controlled genes in bladder cancer. For this purpose we performed genome wide expression analyses of eight bladder cancer cell lines that were treated with the demethylating agents 5-aza-2'-cytidine and Zebularine. This identified a total of 1092 genes that showed ≥2-fold altered expression in at least one cell line; 710 showed up-regulation and 382 down-regulation. To perform clone-based bisulfite sequencing analysis we selected 15 up- and 2 down-regulated CpG island genes, and 8 up-regulated non-CpG island genes. We identified methylated C-residues in 15 genes. Of these, thirteen genes showed an association between promoter methylation and expression. The observed methylation patterns revealed a patchy appearance with short segments having high level of methylation separated by larger segments with no methylation. We investigated if the patchy pattern could be correlated with presence of the repetitive sequences or with binding sites for MeCP2, but no such correlations were observed. It is known that methylated C-residues in TFBSs may affect the interactions with transcription factors. We therefore investigated if frequently methylated CpGs or patches of methylated CpGs coincided with evolutionarily conserved TFBSs. We indeed found high scoring and evolutionarily conserved TFBSs that included CpGs in the recognition sites that were affected by methylation. The identified TFBSs included sites that were previously known to be important for regulation of the respective genes. For example, *CDKN1C* is known to be regulated by SMAD proteins and we found a high scoring SMAD binding site in the methylated segment of *CDKN1C* promoter. In addition, to identify genes that could be activated by promoter methylation we screened for genes that showed reduced expression as a consequence of demethylation. Two genes were identified, *FGF18* and *MMP11*, which both were down-regulated as response to 5-aza-2'-cytidine treatment and showed methylation at specific sites in the untreated cells. In the *FGF18* promoter a single CpG site corresponding to a RFX binding site was methylated in 7 out of 8 cases and this methylation was associated with increased expression. The RFX has been shown to mediate transcriptional activation by binding methylated RFX sites. Consequently, this might be the cause for the decreased expression seen after demethylation.

Comments

Several methods have been described to study the level of promoter methylation. The majority of these is based on the study of single CpG sites previously known to be methylated and hence is not suitable for studying overall methylation patterns in promoters. On the other hand these methods may be very useful when critical sites have been identified and the aim is to characterize larger sets of tumors. To this end there exist several array platforms to identify methylated cytosines in several promoters simultaneously. For example, Illumina provides DNA methylation arrays which give information of approximately one CpG per gene. In the present investigation, however, our aim was to identify new methylation patterns of promoters and therefore the established bisulfite sequencing method was employed. This method has been used in many previous investigations and proved to be highly reliable. The methodology is highly standardized and accepted as "gold standard" for the detection of new CpG sites that are methylated. The method is however associated with some difficulties, of which finding primers suitable for PCR of the bisulfite treated DNA is one.

ARTICLE III

(miRNAs and Urothelial Carcinomas (UC))

The aim of Article III was to explore the possibility to use miRNA expression profiles to classify bladder tumor subtypes and to investigate possible associations of miRNA expression with frequent genetic alterations in UC. For these purposes we analyzed 34 cases of UC by miRNA profiling. The tumors comprised of Ta, T1, and muscle invasive tumors (T2 and T3) and cases with *FGFR3* and *TP53* mutations. In addition, gene expression profiles (mRNA) and genomic profiles (array CGH) were available for these cases. Of the 735 miRNAs present on the array information for 300 remained after an appropriate data filtering.

Unsupervised HCA based on the 300 miRNAs resulted in three major clusters, one that contained all the Ta tumors, a second that contained all but two T1 cases, and a third cluster containing all muscle invasive tumors T2/3 except three. A three-class SAM analysis was performed to identify miRNAs that showed significant association with Ta, T1, and T2/3 cases, respectively and a total of 51 miRNAs were identified. A subsequent

HCA using these miRNAs produced three distinct groups of tumors that showed 82% concordance with tumor stage. A score based on expression of the 51 miRNAs was produced for each tumor and used for class predication purposes. The obtained profile could classify the tumors into Ta, T1, and T2/3 cases at high precision and sensitivity (AUC=0.96). Three miRNAs hsa-mirs -125b, -109b, and -100 were shown to be significantly over-expressed in muscle invasive tumors and hsa-mir-10a was shown to be specifically expressed in Ta tumors. Hsa-mir-10a is located in the *HOXB* gene cluster located in 17q21, between *HOXB4* and *HOXB5*. The expression data for the *HOXB4* and the *HOXB5* genes was not available, however, hsa-mir-10a expression showed significant correlation with expression of the flanking *HOXB3* and *HOXB6* genes. A similar SAM analysis between *FGFR3* mutated and non-mutated cases identified a miRNA signature consisting of 10 miRNAs, of which hsa-mir-7 was found to be associated with mutated cases. This signature showed significant association with *FGFR3* mutated cases and could be used to identify *FGFR3* mutated cases with good precision and sensitivity (AUC=0.90). In addition, three of the cases showed homozygous deletion at *CDKN2A/2B/P14*^{ARF} locus that also included hsa-mir-31. Finally, we could show that hsa-mirs -452 and 452* was associated with lymph node metastases and could be used as a markers for bad prognosis.

Comments

Currently, there is information for 685 miRNAs but it is believed that more than 1000 miRNAs may exist in human genome. The major problems in identifying miRNAs are their small size and the existence of miRNA families with very similar sequences. To date, only a limited number of array platforms are available to perform miRNA expression profiling. In our study we used a platform produced by Illumina that contains 735 human miRNAs, including 265 potential miRNAs identified by computational methods. We indeed found that miRNA expression profiling could classify bladder tumors into pathological stages and could be used for prognostic purpose. In an attempt to identify miRNA target genes we used TaregetScan release 4.0 software that identified high scoring mRNAs for each miRNA. As the main function of miRNAs is to inhibit gene expression, often by the degradation of target mRNAs, one would expect a negative

correlation between miRNAs and its target genes, such correlations were however not found. A likely explanation for this is that most of the inhibition occurs at translational level and thus does not include mRNA degradation. Nevertheless, we identified several miRNAs that could be of possible importance for UC. However, to gain insights how these miRNAs contribute to initiation and progression of bladder cancer experimental studies are needed.

CONCLUSIONS

The main findings of the present thesis can be summarized as follows:

Article I

We developed a new approach to analyse promoter sequences for the presence of enriched TFBSs in co-expressed genes. This approach unveiled biological relevant patterns of TFBSs using two experimental data sets.

Article II

Promoters of both CpG-island and non-CpG-island genes may show a non-uniform distribution of methylated CpGs resulting in patchy methylation patterns. Such patches may often contain high and evolutionarily conserved binding sites. In addition, several genes that may be of importance for bladder cancer development were identified.

Article III

Urothelial carcinomas show distinct miRNA expression signatures associated with tumor stages and frequent genetic changes. Important findings are that hsa-mir-10a expression is associated with Ta tumors, that hsa-mir-31 is frequently deleted, and that hsa-mir-7 expression is associated with *FGFR3* mutated cases. In addition hsa-mirs -452 and 452* is associated with lymph node metastases.

SUMMARY IN SWEDISH

Populärvetenskaplig sammanfattning

Cancer är ett samlingsnamn på en grupp sjukdomar som alla karakteriseras av att reglering cellers tillväxt på något sätt har åsidosatts. Förutom att cancerceller har en förmåga att växer mer än normala celler så har de ofta förlorar sina tidigare egenskaper och funktioner. Sättet på vilket dessa förändringar sker varierar mellan olika cancertyper och är avhängigt i vilken celltyp cancerväxten uppstår. Så till exempel visar blodtumörer (leukemier) helt annorlunda förändringar än t.ex. lungtumörer. Gemensamt för alla cancerformer är dock att de har förändrat de genetiska program som bestämmer deras egenskaper. Ett genetiskt program kan definieras som uttrycket av en grupp gener, allt från ett 20-tal till flera hundra, som ger cellen en viss egenskap. Kombinationen av fler sådana program ger sammantaget cellen sina typiska egenskaper, så kan t.ex. kombinationen av sådana program resultera i en levercell som utför de funktioner den är ämnad för. Av detta framgår att reglering av genetiska program är viktig för cellens funktioner. Genetiska program regleras huvudsakligen av två funktioner, sk. promotorer och transkription faktorer. Promotorn utgörs av en DNA sekvens som sitter framför genen och styr dess aktivitet. Promotorn innehåller ett flertal korta DNA sekvenser till vilket transkriptionsfaktorer binder. Dessa faktorer kan både aktivera och inaktivera en given gen. Till detta kommer att en och samma faktor kan påverka ett flertal gener som tillsammans kan utgöra ett genetiskt program. Förändringar av transkriptionsfaktorers aktivitet och eventuella modifieringar promotorer är därför viktiga för uppkomst av cancer. I mitt avhandlingsarbete har jag därför fokuserat på dessa faktorer.

I mitt första arbete utgår jag från publicerad information om förändringar i genuttryck i skilda tumörtyper. Den centrala frågeställningen är om det går att utifrån sådana data sluta sig till vilka transkriptionsfaktorer som är viktiga för den givna tumörtypen. I ett första steg skapar vi med hjälp av statistiska metoder profiler för grupper av gener som samvarierar över ett stort antal tumörer. Sådana profiler motsvarar i viss mån genetiska program. Därefter analyserar vi promotorerna i sådana grupper av gener för att utröna vilka transkriptionsfaktorer som kan tänkas styra dessa gener. För

detta ändamål utvecklade vi en metod som stegvis bearbetar informationen för att till slut identifiera ett flertal faktorer. Vi visar med hjälp av tidigare kända samband att våran metod lyckas väl.

I mitt andra arbete fokuserar jag på själva promotorregionen. När transkriptionsfaktorer binder till dessa sker det genom att faktorn, som är ett protein, känner igen en specifik DNA sekvens. Dessa sekvenser kan modifieras genom så kallad metylering. I de flesta fall leder en sådan modifiering till att faktorn inte längre kan binda till sekvensen och att genens uttryck därmed stängs av. Denna undersökning inleddes med att jag identifierade ett flertal gener vars uttryck stängdes av genom metylering. Därefter kartlade jag de exakta positionerna i DNA sekvensen där modifieringarna skett. Vi kunde visa att många gener uppvisade klara mönster av modifieringar, mönster som inte var slumpmässiga. Framför allt kunde vi visa att vissa delar av promotorerna visade metylering oftare än andra och att dessa segment för det mesta innehöll igenkänningssekvenser för transkriptionsfaktorer. I denna undersökning arbetade jag med blåscancer som modellsystem och vi identifierade ett flertal gener som kan vara av avgörande betydelse för uppkomst och progression av blåscancer.

I mitt tredje arbete fokuserande jag på en nyligen upptäckt typ av reglerande molekyler sk. microRNAn. Dessa fungerar analogt med transkriptionsfaktorer genom att de reglerar aktiviteten av ett flertal gener, men skiljer genom att inte vara proteiner utan små RNA molekyler. Då mycket lite är känt om vilka gener ett givet microRNA reglerar var vår frågeställning huruvida uttrycket av dessa varierar i blåscancer och om denna variation sammanfaller med egenskaper som aggressivitet och bildandet av metastaser. I denna studie och kunde visa att olika patologiska undergrupper av blåscancer visade specifika microRNA mönster. Detta skulle kunna betyda att microRNAn, i analogi med transkriptionsfaktorer, på ett central sätt bestämmer egenskaper hos tumörerna.

ACKNOWLEDGEMENTS

Although it is stimulating and motivating experience, completing a PhD takes a lot of effort and energy. There are therefore, several people I would like to thank for helping me going through this process.

Mattias Höglund, my supervisor, thank you for accepting me as your student and giving me timely support, encouragement, and personal advices whenever needed.

Felix Mitelman, former Head of the Department, thank you for giving me the opportunity to study at the Department of Clinical Genetics and for creating scientifically inspiring environment.

Ioannis Panagopoulos, thank you for sharing knowledge and timely support in my Methylation project.

Thanks to my senior research staff at the Department of Clinical Genetics, *Bertil*, *Charlotte*, *David G*, *Fredrik*, *Kristina*, *Karin*, *Lyudmila*, *Mia*, *Nils*, *Nina*, *Samuel*, *Thoas*, *Ulf*, and *Yuesheng* for sharing knowledge.

Thanks to all my past and present colleagues at the department: *Anna A*, *Anna C*, *Björn*, *David L*, *Emely*, *Hammurabi*, *Helena Å*, *Henrik*, *Josef*, *Kajsa*, *Karolin*, *Linda*, *Malin*, *Markus H*, *Markus J*, *Petra*, *Therese*, *Tord* and *Ylva* for pleasant and enjoyable moments! Thank you all for lots of fun over the years! Love you guys!!! A special thanks to *David L* for supporting me through out my PhD studies and for being my "Best Man" (it was a cultural shock and a difficult role but you pulled it☺).

Carin Lassen and *Magarette Isaksson* (my gal friends☺), thank you for your amazing kindness and for your technical advice in the lab.

Anette Welin, thank you for your help in every bureaucratic problem.

Thanks to all my collegeaus in SCIBLU, *Anna K*, *Anders*, *Carlos*, *Cecilia*, *Eleonor*, *Göran*, *Helena P*, *Ingrid*, *Jeanette*, *Jari*, *Joachim*, *Johan V*, *Johan S*, *Markus R*, and *Widet*, its a real pleasure to share & enjoy the nice atmosphere & coffee breaks.

A very special thanks to *Ylva*, *Denial*, *Anna A*, *Simon*, *Henrik*, *Helena*, *Emmy*, *Micky*, *Joseeee*, *Anna K*, *Jeanette*, *David & Peter* for making my wedding the most memorable one with your presence and *Markus R*, *Johan V*, *Johan S*, *David G*, *Markus J*, *Hammu*, *Göran* for the wonderful and chilling bachelors party.

Thanks to all my friends outside the Lab, *Benny*, *Victor*, *Esko*, *Kalle*, *Anna-Marie*, *Michell*, *Rohit (Mera chota bhai)*, *Rahul (Raul)*, *Sriky*, *Sinan (best neighbor)*, *Sanjeev Bhai*, *Deepak*, *Asheesh*, *Daniel B*, *Tobi*, *Sonja*, *Nathan*, *Prem*, *Vijay K (Tarjan), Ram (spicy), Balaji* and to all Lund Cricket Club members, for showing me that there is life outside the lab. Thanks to my little friends, *Pranav*, *Pooja*, and *Asavari*, for cheering me up in Indian parties.

Thanks to my parents, father, *Kishan Veerla*, mother, *Yada Laxmi Veerla*, for always believing in and supporting me no matter what decision I take. Last but not the least I would like to thank my wife, *Dhana Laxmi Veerla*, and my brothers, *Naveen* and *Kranthi*, for endless love.

This study was supported by grants from the Swedish Cancer Society, the Swedish Research Council, the Gunnar, Arvid and Elisabeth Nilsson Foundation, the Crafoord Foundation, the John and Augusta Persson Foundation, the Inga Britt and Arne Lundberg Foundation, the Maud and Birger Gustavsson Foundation, the Petrus and Augusta Hedlund Foundation. The microarray facility was supported by the Knut and Alice Wallenberg Foundation via the Swegene Program.

REFERENCES

1. Boveri, T. Anton Dohbn. *Science* **36**, 453-468 (1912).
2. Mitelman, F. & Levan, G. Clustering of aberrations to specific chromosomes in human neoplasms. III. Incidence and geographic distribution of chromosome aberrations in 856 cases. *Hereditas* **89**, 207-32 (1978).
3. Rowley, J.D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-3 (1973).
4. Shtivelman, E., Lifshitz, B., Gale, R.P. & Canaani, E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **315**, 550-4 (1985).
5. Knobbe, C.B., Reifenberger, J. & Reifenberger, G. Mutation analysis of the Ras pathway genes NRAS, HRAS, KRAS and BRAF in glioblastomas. *Acta Neuropathol* **108**, 467-70 (2004).
6. Macaluso, M., Russo, G., Cinti, C., Bazan, V., Gebbia, N. & Russo, A. Ras family genes: an interesting link between cell cycle and cancer. *J Cell Physiol* **192**, 125-30 (2002).
7. Knudson, A.G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-3 (1971).
8. Knudson, A.G., Jr. Genetics and cancer. *Postgrad Med* **48**, 70-4 (1970).
9. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-4 (2006).
10. Inge-Vechtomov, S., Zhouravleva, G. & Philippe, M. Eukaryotic release factors (eRFs) history. *Biol Cell* **95**, 195-209 (2003).
11. Kean, K.M. The role of mRNA 5'-noncoding and 3'-end sequences on 40S ribosomal subunit recruitment, and how RNA viruses successfully compete with cellular mRNAs to ensure their own protein synthesis. *Biol Cell* **95**, 129-39 (2003).
12. Linder, P. Yeast RNA helicases of the DEAD-box family involved in translation initiation. *Biol Cell* **95**, 157-67 (2003).
13. Osborne, H.B. An insight into the post-transcriptional control of gene expression in cell function. *Biol Cell* **95**, 125-7 (2003).
14. Paillard, L. & Osborne, H.B. East of EDEN was a poly(A) tail. *Biol Cell* **95**, 211-9 (2003).
15. Ashcroft, M., Kubbutat, M.H. & Vousden, K.H. Regulation of p53 function and stability by phosphorylation. *Mol Cell Biol* **19**, 1751-8 (1999).
16. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-52 (1997).
17. Turner, B.M. Histone acetylation and control of gene expression. *J Cell Sci* **99 (Pt 1)**, 13-20 (1991).
18. Nightingale, K.P., Wellinger, R.E., Sogo, J.M. & Becker, P.B. Histone acetylation facilitates RNA polymerase II transcription of the Drosophila hsp26 gene in chromatin. *Embo J* **17**, 2865-76 (1998).

19. Annes, J.P., Munger, J.S. & Rifkin, D.B. Making sense of latent TGFbeta activation. *J Cell Sci* **116**, 217-24 (2003).
20. Sato, Y. & Rifkin, D.B. Inhibition of endothelial cell movement by pericytes and smooth muscle cells: activation of a latent transforming growth factor-beta 1-like molecule by plasmin during co-culture. *J Cell Biol* **109**, 309-15 (1989).
21. Yu, Q. & Stamenkovic, I. Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-beta and promotes tumor invasion and angiogenesis. *Genes Dev* **14**, 163-76 (2000).
22. Yuh, C.H., Ransick, A., Martinez, P., Britten, R.J. & Davidson, E.H. Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech Dev* **47**, 165-86 (1994).
23. Schoenherr, C.J. & Anderson, D.J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360-3 (1995).
24. Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A. & Dynlacht, B.D. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* **16**, 245-56 (2002).
25. Clark, A.R., Petersen, H.V., Read, M.L., Scott, V., Michelsen, B.K. & Docherty, K. Human insulin gene enhancer-binding proteins in pancreatic alpha and beta cell lines. *FEBS Lett* **329**, 139-43 (1993).
26. Peers, B., Leonard, J., Sharma, S., Teitelman, G. & Montminy, M.R. Insulin expression in pancreatic islet cells relies on cooperative interactions between the helix loop helix factor E47 and the homeobox factor STF-1. *Mol Endocrinol* **8**, 1798-806 (1994).
27. Petersen, H.V., Serup, P., Leonard, J., Michelsen, B.K. & Madsen, O.D. Transcriptional regulation of the human insulin gene is dependent on the homeodomain protein STF1/IPF1 acting through the CT boxes. *Proc Natl Acad Sci U S A* **91**, 10465-9 (1994).
28. Serup, P., Jensen, J., Andersen, F.G., Jorgensen, M.C., Blume, N., Holst, J.J. & Madsen, O.D. Induction of insulin and islet amyloid polypeptide production in pancreatic islet glucagonoma cells by insulin promoter factor 1. *Proc Natl Acad Sci U S A* **93**, 9015-20 (1996).
29. Dean, A. On a chromosome far, far away: LCRs and gene expression. *Trends Genet* **22**, 38-45 (2006).
30. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. & Flavell, R.A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637-45 (2005).
31. Collins, M., Rojnuckarin, P., Zhu, Y.H. & Bornstein, P. A far upstream, cell type-specific enhancer of the mouse thrombospondin 3 gene is located within intron 6 of the adjacent metaxin gene. *J Biol Chem* **273**, 21816-24 (1998).
32. Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. & Wright, P.E. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-5 (1995).
33. Lu, Q. & Richardson, B. DNaseI hypersensitivity analysis of chromatin structure. *Methods Mol Biol* **287**, 77-86 (2004).

34. Buck, M.J. & Lieb, J.D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349-60 (2004).
35. Hanlon, S.E. & Lieb, J.D. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr Opin Genet Dev* **14**, 697-705 (2004).
36. Horak, C.E. & Snyder, M. ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol* **350**, 469-83 (2002).
37. Kirmizis, A. & Farnham, P.J. Genomic approaches that aid in the identification of transcription factor target genes. *Exp Biol Med (Maywood)* **229**, 705-21 (2004).
38. Lieb, J.D. Genome-wide mapping of protein-DNA interactions by chromatin immunoprecipitation and DNA microarray hybridization. *Methods Mol Biol* **224**, 99-109 (2003).
39. Rodriguez, B.A. & Huang, T.H. Tilling the chromatin landscape: emerging methods for the discovery and profiling of protein-DNA interactions. *Biochem Cell Biol* **83**, 525-34 (2005).
40. Gross, D.S. & Garrard, W.T. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57**, 159-97 (1988).
41. Crawford, G.E., Holt, I.E., Mullikin, J.C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E.D., Wolfsberg, T.G. & Collins, F.S. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci U S A* **101**, 992-7 (2004).
42. Shewchuk, B.M., Asa, S.L., Cooke, N.E. & Liebhaber, S.A. Pit-1 binding sites at the somatotrope-specific DNase I hypersensitive sites I, II of the human growth hormone locus control region are essential for in vivo hGH-N gene activation. *J Biol Chem* **274**, 35725-33 (1999).
43. Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M. & Snyder, M. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. *Proc Natl Acad Sci U S A* **99**, 2924-9 (2002).
44. Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q. & Ren, B. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci U S A* **100**, 8164-9 (2003).
45. Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H. & Farnham, P.J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* **16**, 235-44 (2002).
46. Wells, J., Yan, P.S., Cechvala, M., Huang, T. & Farnham, P.J. Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene* **22**, 1445-60 (2003).
47. Knuppel, R., Dietze, P., Lehnberg, W., Frech, K. & Wingender, E. TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol* **1**, 191-8 (1994).
48. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4 (2004).

49. Aerts, S., Thijs, G., Coessens, B., Staes, M., Moreau, Y. & De Moor, B. Toucan: deciphering the cis-regulatory logic of coregulated genes. *Nucleic Acids Res* **31**, 1753-64 (2003).
50. Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V. & Wingender, E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* **31**, 3576-9 (2003).
51. Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. & Werner, T. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933-42 (2005).
52. Tompa, M. Identifying functional elements by comparative DNA sequence analysis. *Genome Res* **11**, 1143-4 (2001).
53. Gumucio, D.L., Shelton, D.A., Bailey, W.J., Slightom, J.L. & Goodman, M. Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the epsilon-globin gene. *Proc Natl Acad Sci U S A* **90**, 6018-22 (1993).
54. Vuillaumier, S., Dixmeras, I., Messai, H., Lapoumeroulie, C., Lallemand, D., Gekas, J., Chehab, F.F., Perret, C., Elion, J. & Denamur, E. Cross-species characterization of the promoter region of the cystic fibrosis transmembrane conductance regulator gene reveals multiple levels of regulation. *Biochem J* **327 (Pt 3)**, 651-62 (1997).
55. Leung, J.Y., McKenzie, F.E., Uglialoro, A.M., Flores-Villanueva, P.O., Sorkin, B.C., Yunis, E.J., Hartl, D.L. & Goldfeld, A.E. Identification of phylogenetic footprints in primate tumor necrosis factor-alpha promoters. *Proc Natl Acad Sci U S A* **97**, 6614-8 (2000).
56. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-70 (1995).
57. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-10 (2002).
58. Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. & Chinnaiyan, A.M. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1-6 (2004).
59. Bussemaker, H.J., Li, H. & Siggia, E.D. Regulatory element detection using correlation with expression. *Nat Genet* **27**, 167-71 (2001).
60. Haverty, P.M., Hansen, U. & Weng, Z. Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification. *Nucleic Acids Res* **32**, 179-88 (2004).
61. Kim, S.Y. & Kim, Y. Genome-wide prediction of transcriptional regulatory elements of human promoters using gene expression and promoter analysis data. *BMC Bioinformatics* **7**, 330 (2006).
62. Aerts, S., Van Loo, P., Thijs, G., Mayer, H., de Martin, R., Moreau, Y. & De Moor, B. TOUCAN 2: the all-inclusive open source workbench for regulatory sequence analysis. *Nucleic Acids Res* **33**, W393-6 (2005).

63. Karanam, S. & Moreno, C.S. CONFAC: automated application of comparative genomic promoter analysis to DNA microarray datasets. *Nucleic Acids Res* **32**, W475-84 (2004).
64. Ho Sui, S.J., Fulton, D.L., Arenillas, D.J., Kwon, A.T. & Wasserman, W.W. oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res* **35**, W245-52 (2007).
65. Ho Sui, S.J., Mortimer, J.R., Arenillas, D.J., Brumm, J., Walsh, C.J., Kennedy, B.P. & Wasserman, W.W. oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res* **33**, 3154-64 (2005).
66. Perier, R.C., Praz, V., Junier, T., Bonnard, C. & Bucher, P. The eukaryotic promoter database (EPD). *Nucleic Acids Res* **28**, 302-3 (2000).
67. Krivan, W. & Wasserman, W.W. A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res* **11**, 1559-66 (2001).
68. Sharan, R., Ben-Hur, A., Loots, G.G. & Ovcharenko, I. CREME: Cis-Regulatory Module Explorer for the human genome. *Nucleic Acids Res* **32**, W253-6 (2004).
69. Sharan, R., Ovcharenko, I., Ben-Hur, A. & Karp, R.M. CREME: a framework for identifying cis-regulatory modules in human-mouse conserved segments. *Bioinformatics* **19 Suppl 1**, i283-91 (2003).
70. Yuh, C.H., Bolouri, H. & Davidson, E.H. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**, 1896-902 (1998).
71. Jin, V.X., O'Geen, H., Iyengar, S., Green, R. & Farnham, P.J. Identification of an OCT4 and SRY regulatory module using integrated computational and experimental genomics approaches. *Genome Res* **17**, 807-17 (2007).
72. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **293**, 1074-80 (2001).
73. Bird, A.P. & Wolffe, A.P. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**, 451-4 (1999).
74. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**, 662-73 (2002).
75. Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644-9 (2004).
76. Horsthemke, B. & Buiting, K. Imprinting defects on human chromosome 15. *Cytogenet Genome Res* **113**, 292-9 (2006).
77. Jones, P.A. DNA methylation and cancer. *Oncogene* **21**, 5358-60 (2002).
78. Robertson, K.D. DNA methylation, methyltransferases, and cancer. *Oncogene* **20**, 3139-55 (2001).
79. Ahuja, N. & Issa, J.P. Aging, methylation and cancer. *Histol Histopathol* **15**, 835-42 (2000).
80. Jones, P.A. & Laird, P.W. Cancer epigenetics comes of age. *Nat Genet* **21**, 163-7 (1999).
81. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. 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**, 971-83 (1988).
82. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**, 219-20 (1998).
83. Oka, M., Meacham, A.M., Hamazaki, T., Rodic, N., Chang, L.J. & Terada, N. De novo DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine. *Oncogene* **24**, 3091-9 (2005).
84. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57 (1999).
85. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26 (1992).
86. Antequera, F. Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci* **60**, 1647-58 (2003).
87. Bird, A.P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**, 1499-504 (1980).
88. Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P. & Jaenisch, R. 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**, 5237-42 (2000).
89. Russell, G.J., Walker, P.M., Elton, R.A. & Subak-Sharpe, J.H. Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J Mol Biol* **108**, 1-23 (1976).
90. Swartz, M.N., Trautner, T.A. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol Chem* **237**, 1961-7 (1962).
91. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
92. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-82 (1987).
93. Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095-107 (1992).
94. McClelland, M. & Ivarie, R. Asymmetrical distribution of CpG in an 'average' mammalian gene. *Nucleic Acids Res* **10**, 7865-77 (1982).
95. Takai, D. & Jones, P.A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99**, 3740-5 (2002).
96. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M. & Schubeler, D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39**, 457-66 (2007).
97. Yoder, J.A., Walsh, C.P. & Bestor, T.H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**, 335-40 (1997).
98. Jones, P.A. & Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-28 (2002).
99. Murzina, N., Verreault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol Cell* **4**, 529-40 (1999).

100. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* **12**, 599-606 (1998).
101. Taddei, A., Maison, C., Roche, D. & Almouzni, G. Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. *Nat Cell Biol* **3**, 114-20 (2001).
102. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-47 (1998).
103. Roloff, T.C., Ropers, H.H. & Nuber, U.A. Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics* **4**, 1 (2003).
104. Hung, M.S. & Shen, C.K. Eukaryotic methyl-CpG-binding domain proteins and chromatin modification. *Eukaryot Cell* **2**, 841-6 (2003).
105. Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I. & Bird, A.P. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol Cell* **19**, 667-78 (2005).
106. Wade, P.A. Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. *Oncogene* **20**, 3166-73 (2001).
107. Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A. & Hansen, J.C. Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J Biol Chem* **278**, 32181-8 (2003).
108. Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. & Wolffe, A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-91 (1998).
109. Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. & Bird, A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-9 (1998).
110. Wade, P.A. Methyl CpG-binding proteins and transcriptional repression. *Bioessays* **23**, 1131-7 (2001).
111. De Larco, J.E., Wuertz, B.R., Yee, D., Rickert, B.L. & Furcht, L.T. Atypical methylation of the interleukin-8 gene correlates strongly with the metastatic potential of breast carcinoma cells. *Proc Natl Acad Sci U S A* **100**, 13988-93 (2003).
112. Niesen, M.I., Osborne, A.R., Yang, H., Rastogi, S., Chellappan, S., Cheng, J.Q., Boss, J.M. & Blanck, G. Activation of a methylated promoter mediated by a sequence-specific DNA-binding protein, RFX. *J Biol Chem* **280**, 38914-22 (2005).
113. Unoki, M. & Nakamura, Y. Methylation at CpG islands in intron 1 of EGR2 confers enhancer-like activity. *FEBS Lett* **554**, 67-72 (2003).
114. Baylin, S.B. & Herman, J.G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* **16**, 168-74 (2000).
115. Ha, P.K. & Califano, J.A. Promoter methylation and inactivation of tumour-suppressor genes in oral squamous-cell carcinoma. *Lancet Oncol* **7**, 77-82 (2006).
116. Calmon, M.F., Colombo, J., Carvalho, F., Souza, F.P., Filho, J.F., Fukuyama, E.E., Camargo, A.A., Caballero, O.L., Tajara, E.H., Cordeiro, J.A. & Rahal, P. Methylation profile of genes CDKN2A (p14 and p16), DAPK1, CDH1, and ADAM23 in head and neck cancer. *Cancer Genet Cytogenet* **173**, 31-7 (2007).

117. Herman, J.G. & Baylin, S.B. Promoter-region hypermethylation and gene silencing in human cancer. *Curr Top Microbiol Immunol* **249**, 35-54 (2000).
118. Kane, M.F., Loda, M., Gaida, G.M., Lipman, J., Mishra, R., Goldman, H., Jessup, J.M. & Kolodner, R. 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**, 808-11 (1997).
119. Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W., Kane, M.F., Kolodner, R.D., Vogelstein, B., Kunkel, T.A. & Baylin, S.B. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* **95**, 6870-5 (1998).
120. El-Osta, A., Kantharidis, P., Zalcberg, J.R. & Wolffe, A.P. Precipitous release of methyl-CpG binding protein 2 and histone deacetylase 1 from the methylated human multidrug resistance gene (MDR1) on activation. *Mol Cell Biol* **22**, 1844-57 (2002).
121. 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. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**, 9700-4 (1994).
122. Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., Gabrielson, E., Schutte, M., Baylin, S.B. & Herman, J.G. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* **92**, 564-9 (2000).
123. Novak, P., Jensen, T., Oshiro, M.M., Wozniak, R.J., Nouzova, M., Watts, G.S., Klimecki, W.T., Kim, C. & Futscher, B.W. Epigenetic inactivation of the HOXA gene cluster in breast cancer. *Cancer Res* **66**, 10664-70 (2006).
124. Maruyama, R., Toyooka, S., Toyooka, K.O., Harada, K., Virmani, A.K., Zochbauer-Muller, S., Farinas, A.J., Vakar-Lopez, F., Minna, J.D., Sagalowsky, A., Czerniak, B. & Gazdar, A.F. Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features. *Cancer Res* **61**, 8659-63 (2001).
125. Hoffmann, M.J., Florl, A.R., Seifert, H.H. & Schulz, W.A. Multiple mechanisms downregulate CDKN1C in human bladder cancer. *Int J Cancer* **114**, 406-13 (2005).
126. Kim, W.J., Kim, E.J., Jeong, P., Quan, C., Kim, J., Li, Q.L., Yang, J.O., Ito, Y. & Bae, S.C. RUNX3 inactivation by point mutations and aberrant DNA methylation in bladder tumors. *Cancer Res* **65**, 9347-54 (2005).
127. Myohanen, S.K., Baylin, S.B. & Herman, J.G. Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia. *Cancer Res* **58**, 591-3 (1998).
128. Esteller, M., Fraga, M.F., Guo, M., Garcia-Foncillas, J., Hedenfalk, I., Godwin, A.K., Trojan, J., Vaurs-Barriere, C., Bignon, Y.J., Ramus, S., Benitez, J., Caldes, T., Akiyama, Y., Yuasa, Y., Launonen, V., Canal, M.J., Rodriguez, R., Capella, G., Peinado, M.A., Borg, A., Aaltonen, L.A., Ponder, B.A., Baylin, S.B. & Herman, J.G. DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. *Hum Mol Genet* **10**, 3001-7 (2001).

129. Kaneda, A., Tsukamoto, T., Takamura-Enya, T., Watanabe, N., Kaminishi, M., Sugimura, T., Tatematsu, M. & Ushijima, T. Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation. *Cancer Sci* **95**, 58-64 (2004).
130. Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455 (2003).
131. Narayan, A., Ji, W., Zhang, X.Y., Marrogi, A., Graff, J.R., Baylin, S.B. & Ehrlich, M. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer* **77**, 833-8 (1998).
132. Qu, G., Dubeau, L., Narayan, A., Yu, M.C. & Ehrlich, M. Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. *Mutat Res* **423**, 91-101 (1999).
133. Qu, G.Z., Grundy, P.E., Narayan, A. & Ehrlich, M. Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. *Cancer Genet Cytogenet* **109**, 34-9 (1999).
134. Lengauer, C. Cancer. An unstable liaison. *Science* **300**, 442-3 (2003).
135. Christman, J.K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483-95 (2002).
136. Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J. & Hornby, D.P. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J Mol Biol* **321**, 591-9 (2002).
137. Santini, V., Kantarjian, H.M. & Issa, J.P. Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. *Ann Intern Med* **134**, 573-86 (2001).
138. Kornblith, A.B., Herndon, J.E., 2nd, Silverman, L.R., Demakos, E.P., Odchimar-Reissig, R., Holland, J.F., Powell, B.L., DeCastro, C., Ellerton, J., Larson, R.A., Schiffer, C.A. & Holland, J.C. Impact of azacytidine on the quality of life of patients with myelodysplastic syndrome treated in a randomized phase III trial: a Cancer and Leukemia Group B study. *J Clin Oncol* **20**, 2441-52 (2002).
139. Silverman, L.R., Demakos, E.P., Peterson, B.L., Kornblith, A.B., Holland, J.C., Odchimar-Reissig, R., Stone, R.M., Nelson, D., Powell, B.L., DeCastro, C.M., Ellerton, J., Larson, R.A., Schiffer, C.A. & Holland, J.F. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol* **20**, 2429-40 (2002).
140. Butler, L.M. Suberoylanilide hydroxamic acid, an inhibitor of histone deacetylase, suppresses the growth of prostate cancer cells in vitro and in vivo. *Cancer Res.* **60**, 5165-5170 (2000).
141. Butler, L.M. Inhibition of transformed cell growth and induction of cellular differentiation by pyroxamide, an inhibitor of histone deacetylase. *Clin. Cancer Res.* **7**, 962-970 (2001).
142. Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. & Baylin, S.B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* **93**, 9821-6 (1996).
143. Eads, C.A. & Laird, P.W. Combined bisulfite restriction analysis (COBRA). *Methods Mol Biol* **200**, 71-85 (2002).

144. Gonzalgo, M.L. & Jones, P.A. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res* **25**, 2529-31 (1997).
145. Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P.V. & Laird, P.W. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* **28**, E32 (2000).
146. Worm, J., Aggerholm, A. & Guldberg, P. In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin Chem* **47**, 1183-9 (2001).
147. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. & Paul, C.L. 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**, 1827-31 (1992).
148. Dahl, C. & Guldberg, P. DNA methylation analysis techniques. *Biogerontology* **4**, 233-50 (2003).
149. 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. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**, 132-8 (2000).
150. Toyota, M., Ho, C., Ahuja, N., Jair, K.W., Li, Q., Ohe-Toyota, M., Baylin, S.B. & Issa, J.P. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* **59**, 2307-12 (1999).
151. Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T. & Nagao, M. 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**, 2284-9 (1997).
152. Gonzalgo, M.L., Liang, G., Spruck, C.H., 3rd, Zingg, J.M., Rideout, W.M., 3rd & Jones, P.A. Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer Res* **57**, 594-9 (1997).
153. Huang, T.H., Laux, D.E., Hamlin, B.C., Tran, P., Tran, H. & Lubahn, D.B. Identification of DNA methylation markers for human breast carcinomas using the methylation-sensitive restriction fingerprinting technique. *Cancer Res* **57**, 1030-4 (1997).
154. Yan, P.S., Chen, C.M., Shi, H., Rahmatpanah, F., Wei, S.H., Caldwell, C.W. & Huang, T.H. Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res* **61**, 8375-80 (2001).
155. Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E. & Ecker, J.R. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1-15 (2005).
156. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E. & Ecker, J.R. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189-201 (2006).

157. Chan, S.W., Henderson, I.R. & Jacobsen, S.E. Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat Rev Genet* **6**, 351-60 (2005).
158. Melnikov, A.A., Scholtens, D.M., Wiley, E.L., Khan, S.A. & Levenson, V.V. Array-based multiplex analysis of DNA methylation in breast cancer tissues. *J Mol Diagn* **10**, 93-101 (2008).
159. Yu, Y.P., Paranjpe, S., Nelson, J., Finkelstein, S., Ren, B., Kokkinakis, D., Michalopoulos, G. & Luo, J.H. High throughput screening of methylation status of genes in prostate cancer using an oligonucleotide methylation array. *Carcinogenesis* **26**, 471-9 (2005).
160. Lee, R.C., Feinbaum, R.L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54 (1993).
161. Lee, Y.S., Kim, H.K., Chung, S., Kim, K.S. & Dutta, A. Depletion of human micro-RNA miR-125b reveals that it is critical for the proliferation of differentiated cells but not for the down-regulation of putative targets during differentiation. *J Biol Chem* **280**, 16635-41 (2005).
162. Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y., Mitsudomi, T. & Takahashi, T. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* **64**, 3753-6 (2004).
163. Chen, C.Z., Li, L., Lodish, H.F. & Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83-6 (2004).
164. Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P. & Stoffel, M. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**, 226-30 (2004).
165. Hornstein, E., Mansfield, J.H., Yekta, S., Hu, J.K., Harfe, B.D., McManus, M.T., Baskerville, S., Bartel, D.P. & Tabin, C.J. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature* **438**, 671-4 (2005).
166. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214-20 (2005).
167. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
168. Gu, J., He, T., Pei, Y., Li, F., Wang, X., Zhang, J., Zhang, X. & Li, Y. Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences. *Mamm Genome* **17**, 1033-41 (2006).
169. Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. & Shiekhattar, R. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-40 (2004).
170. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. & Kim, V.N. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
171. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. & Hannon, G.J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231-5 (2004).

172. Yi, R., Qin, Y., Macara, I.G. & Cullen, B.R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-6 (2003).
173. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95-8 (2004).
174. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. & Zamore, P.D. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-8 (2001).
175. Schwarz, D.S., Hutvagner, G., Haley, B. & Zamore, P.D. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol Cell* **10**, 537-48 (2002).
176. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. & Bartel, D.P. Prediction of plant microRNA targets. *Cell* **110**, 513-20 (2002).
177. Carrington, J.C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336-8 (2003).
178. Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E. & Ambros, V. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol* **5**, R13 (2004).
179. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**, 735-9 (2002).
180. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y. & Bentwich, Z. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**, 766-70 (2005).
181. Berezikov, E., van Tetering, G., Verheul, M., van de Belt, J., van Laake, L., Vos, J., Verloop, R., van de Wetering, M., Guryev, V., Takada, S., van Zonneveld, A.J., Mano, H., Plasterk, R. & Cuppen, E. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res* **16**, 1289-98 (2006).
182. Lai, E.C., Tomancak, P., Williams, R.W. & Rubin, G.M. Computational identification of Drosophila microRNA genes. *Genome Biol* **4**, R42 (2003).
183. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. & Bartel, D.P. The microRNAs of Caenorhabditis elegans. *Genes Dev* **17**, 991-1008 (2003).
184. Nam, J.W., Kim, J., Kim, S.K. & Zhang, B.T. ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucleic Acids Res* **34**, W455-8 (2006).
185. Hsu, S.D., Chu, C.H., Tsou, A.P., Chen, S.J., Chen, H.C., Hsu, P.W., Wong, Y.H., Chen, Y.H., Chen, G.H. & Huang, H.D. miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* **36**, D165-9 (2008).
186. Wang, X., Zhang, J., Li, F., Gu, J., He, T., Zhang, X. & Li, Y. MicroRNA identification based on sequence and structure alignment. *Bioinformatics* **21**, 3610-4 (2005).
187. Saini, H.K., Griffiths-Jones, S. & Enright, A.J. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci U S A* **104**, 17719-24 (2007).

188. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-8 (2001).
189. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. & Dreyfuss, G. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**, 720-8 (2002).
190. Baskerville, S. & Bartel, D.P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna* **11**, 241-7 (2005).
191. Bracht, J., Hunter, S., Eachus, R., Weeks, P. & Pasquinelli, A.E. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *Rna* **10**, 1586-94 (2004).
192. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H. & Kim, V.N. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**, 4051-60 (2004).
193. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**, 1957-66 (2004).
194. Zhou, X., Ruan, J., Wang, G. & Zhang, W. Characterization and identification of microRNA core promoters in four model species. *PLoS Comput Biol* **3**, e37 (2007).
195. Saito, Y., Liang, G., Egger, G., Friedman, J.M., Chuang, J.C., Coetzee, G.A. & Jones, P.A. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* **9**, 435-43 (2006).
196. Kim, V.N. & Nam, J.W. Genomics of microRNA. *Trends Genet* **22**, 165-73 (2006).
197. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-98 (2003).
198. Rajewsky, N. microRNA target predictions in animals. *Nat Genet* **38 Suppl**, S8-13 (2006).
199. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).
200. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. & Rajewsky, N. Combinatorial microRNA target predictions. *Nat Genet* **37**, 495-500 (2005).
201. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. & Marks, D.S. Human MicroRNA targets. *PLoS Biol* **2**, e363 (2004).
202. Sethupathy, P., Corda, B. & Hatzigeorgiou, A.G. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *Rna* **12**, 192-7 (2006).
203. Karube, Y., Tanaka, H., Osada, H., Tomida, S., Tatematsu, Y., Yanagisawa, K., Yatabe, Y., Takamizawa, J., Miyoshi, S., Mitsudomi, T. & Takahashi, T. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Sci* **96**, 111-5 (2005).

204. Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T. & Oshimura, M. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol* **6**, 784-91 (2004).
205. Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M. & Rajewsky, K. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**, 489-501 (2005).
206. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R. & Golub, T.R. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-8 (2005).
207. Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M. & Croce, C.M. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944-9 (2005).
208. Michael, M.Z., SM, O.C., van Holst Pellekaan, N.G., Young, G.P. & James, R.J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res* **1**, 882-91 (2003).
209. Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. & Slack, F.J. RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635-47 (2005).
210. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. & Croce, C.M. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* **101**, 2999-3004 (2004).
211. Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Menard, S., Palazzo, J.P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G.A., Querzoli, P., Negrini, M. & Croce, C.M. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* **65**, 7065-70 (2005).
212. Esquela-Kerscher, A. & Slack, F.J. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-69 (2006).
213. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F. & Croce, C.M. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **99**, 15524-9 (2002).
214. Calin, G.A., Liu, C.G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M., Dell'Aquila, M.L., Alder, H., Rassenti, L., Kipps, T.J., Bullrich, F., Negrini, M. & Croce, C.M. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A* **101**, 11755-60 (2004).
215. Chuang, J.C. & Jones, P.A. Epigenetics and microRNAs. *Pediatr Res* **61**, 24R-29R (2007).

216. Yang, N., Coukos, G. & Zhang, L. MicroRNA epigenetic alterations in human cancer: one step forward in diagnosis and treatment. *Int J Cancer* **122**, 963-8 (2008).
217. Lehmann, U., Hasemeier, B., Christgen, M., Muller, M., Romermann, D., Langer, F. & Kreipe, H. Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J Pathol* **214**, 17-24 (2008).
218. Lujambio, A., Ropero, S., Ballestar, E., Fraga, M.F., Cerrato, C., Setien, F., Casado, S., Suarez-Gauthier, A., Sanchez-Cespedes, M., Git, A., Spiteri, I., Das, P.P., Caldas, C., Miska, E. & Esteller, M. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res* **67**, 1424-9 (2007).
219. Chan, J.A., Krichevsky, A.M. & Kosik, K.S. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* **65**, 6029-33 (2005).
220. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J. & Hammond, S.M. A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828-33 (2005).
221. O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. & Mendell, J.T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839-43 (2005).
222. Brueckner, B., Stresemann, C., Kuner, R., Mund, C., Musch, T., Meister, M., Sultmann, H. & Lyko, F. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res* **67**, 1419-23 (2007).
223. Droller, M.J. Biological considerations in the assessment of urothelial cancer: a retrospective. *Urology* **66**, 66-75 (2005).
224. Reuter, V.E. The pathology of bladder cancer. *Urology* **67**, 11-7; discussion 17-8 (2006).
225. Bryan, R.T. & Wallace, D.M. 'Superficial' bladder cancer - time to uncouple pT1 tumours from pTa tumours. *BJU Int* **90**, 846-52 (2002).
226. Sengupta, S. & Blute, M.L. The management of superficial transitional cell carcinoma of the bladder. *Urology* **67**, 48-54; discussion 54-5 (2006).
227. Larsson, P., Wijkstrom, H., Thorstenson, A., Adolfsson, J., Norming, U., Wiklund, P., Onelov, E. & Steineck, G. A population-based study of 538 patients with newly detected urinary bladder neoplasms followed during 5 years. *Scand J Urol Nephrol* **37**, 195-201 (2003).
228. Leblanc, B., Duclos, A.J., Benard, F., Cote, J., Valiquette, L., Paquin, J.M., Mauffette, F., Faucher, R. & Perreault, J.P. Long-term followup of initial Ta grade 1 transitional cell carcinoma of the bladder. *J Urol* **162**, 1946-50 (1999).
229. Knowles, M.A. Molecular subtypes of bladder cancer: Jekyll and Hyde or chalk and cheese? *Carcinogenesis* **27**, 361-73 (2006).
230. Wu, X.R. Urothelial tumorigenesis: a tale of divergent pathways. *Nat Rev Cancer* **5**, 713-25 (2005).
231. Cappellen, D., De Oliveira, C., Ricol, D., de Medina, S., Bourdin, J., Sastre-Garau, X., Chopin, D., Thiery, J.P. & Radvanyi, F. Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nat Genet* **23**, 18-20 (1999).
232. Sidransky, D., Von Eschenbach, A., Tsai, Y.C., Jones, P., Summerhayes, I., Marshall, F., Paul, M., Green, P., Hamilton, S.R., Frost, P. & et al. Identification

of p53 gene mutations in bladder cancers and urine samples. *Science* **252**, 706-9 (1991).
233. Eswarakumar, V.P., Lax, I. & Schlessinger, J. Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev* **16**, 139-49 (2005).
234. Bakkar, A.A., Wallerand, H., Radvanyi, F., Lahaye, J.B., Pissard, S., Lecerf, L., Kouyoumdjian, J.C., Abbou, C.C., Pairon, J.C., Jaurand, M.C., Thiery, J.P., Chopin, D.K. & de Medina, S.G. FGFR3 and TP53 gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder. *Cancer Res* **63**, 8108-12 (2003).
235. Billerey, C., Chopin, D., Aubriot-Lorton, M.H., Ricol, D., Gil Diez de Medina, S., Van Rhijn, B., Bralet, M.P., Lefrere-Belda, M.A., Lahaye, J.B., Abbou, C.C., Bonaventure, J., Zafrani, E.S., van der Kwast, T., Thiery, J.P. & Radvanyi, F. Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. *Am J Pathol* **158**, 1955-9 (2001).
236. van Rhijn, B.W., van der Kwast, T.H., Vis, A.N., Kirkels, W.J., Boeve, E.R., Jobsis, A.C. & Zwarthoff, E.C. FGFR3 and P53 characterize alternative genetic pathways in the pathogenesis of urothelial cell carcinoma. *Cancer Res* **64**, 1911-4 (2004).
237. van Rhijn, B.W., Vis, A.N., van der Kwast, T.H., Kirkels, W.J., Radvanyi, F., Ooms, E.C., Chopin, D.K., Boeve, E.R., Jobsis, A.C. & Zwarthoff, E.C. Molecular grading of urothelial cell carcinoma with fibroblast growth factor receptor 3 and MIB-1 is superior to pathologic grade for the prediction of clinical outcome. *J Clin Oncol* **21**, 1912-21 (2003).
238. van Rhijn, B.W., Montironi, R., Zwarthoff, E.C., Jobsis, A.C. & van der Kwast, T.H. Frequent FGFR3 mutations in urothelial papilloma. *J Pathol* **198**, 245-51 (2002).
239. Vajo, Z., Francomano, C.A. & Wilkin, D.J. The molecular and genetic basis of fibroblast growth factor receptor 3 disorders: the achondroplasia family of skeletal dysplasias, Muenke craniosynostosis, and Crouzon syndrome with acanthosis nigricans. *Endocr Rev* **21**, 23-39 (2000).
240. Webster, M.K. & Donoghue, D.J. FGFR activation in skeletal disorders: too much of a good thing. *Trends Genet* **13**, 178-82 (1997).
241. Esrig, D., Elmajian, D., Groshen, S., Freeman, J.A., Stein, J.P., Chen, S.C., Nichols, P.W., Skinner, D.G., Jones, P.A. & Cote, R.J. Accumulation of nuclear p53 and tumor progression in bladder cancer. *N Engl J Med* **331**, 1259-64 (1994).
242. Fadl-Elmula, I. Chromosomal changes in uroepithelial carcinomas. *Cell Chromosome* **4**, 1 (2005).
243. Hoglund, M., Sall, T., Heim, S., Mitelman, F., Mandahl, N. & Fadl-Elmula, I. Identification of cytogenetic subgroups and karyotypic pathways in transitional cell carcinoma. *Cancer Res* **61**, 8241-6 (2001).
244. Bruch, J., Wohr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W. & Paiss, T. Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* **23**, 167-74 (1998).

245. Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T.C., Moch, H., Mihatsch, M.J. & Sauter, G. Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am J Pathol* **153**, 1615-21 (1998).
246. Richter, J., Wagner, U., Schraml, P., Maurer, R., Alund, G., Knonagel, H., Moch, H., Mihatsch, M.J., Gasser, T.C. & Sauter, G. Chromosomal imbalances are associated with a high risk of progression in early invasive (pT1) urinary bladder cancer. *Cancer Res* **59**, 5687-91 (1999).
247. Zhao, J., Richter, J., Wagner, U., Roth, B., Schraml, P., Zellweger, T., Ackermann, D., Schmid, U., Moch, H., Mihatsch, M.J., Gasser, T.C. & Sauter, G. Chromosomal imbalances in noninvasive papillary bladder neoplasms (pTa). *Cancer Res* **59**, 4658-61 (1999).
248. Habuchi, T., Devlin, J., Elder, P.A. & Knowles, M.A. Detailed deletion mapping of chromosome 9q in bladder cancer: evidence for two tumour suppressor loci. *Oncogene* **11**, 1671-4 (1995).
249. Habuchi, T., Yoshida, O. & Knowles, M.A. A novel candidate tumour suppressor locus at 9q32-33 in bladder cancer: localization of the candidate region within a single 840 kb YAC. *Hum Mol Genet* **6**, 913-9 (1997).
250. Keen, A.J. & Knowles, M.A. Definition of two regions of deletion on chromosome 9 in carcinoma of the bladder. *Oncogene* **9**, 2083-8 (1994).
251. Kimura, F., Florl, A.R., Seifert, H.H., Louhelainen, J., Maas, S., Knowles, M.A. & Schulz, W.A. Destabilization of chromosome 9 in transitional cell carcinoma of the urinary bladder. *Br J Cancer* **85**, 1887-93 (2001).
252. Olumi, A.F., Tsai, Y.C., Nichols, P.W., Skinner, D.G., Cain, D.R., Bender, L.I. & Jones, P.A. Allelic loss of chromosome 17p distinguishes high grade from low grade transitional cell carcinomas of the bladder. *Cancer Res* **50**, 7081-3 (1990).
253. Orlow, I., Lianes, P., Lacombe, L., Dalbagni, G., Reuter, V.E. & Cordon-Cardo, C. Chromosome 9 allelic losses and microsatellite alterations in human bladder tumors. *Cancer Res* **54**, 2848-51 (1994).
254. Simoneau, A.R., Spruck, C.H., 3rd, Gonzalez-Zulueta, M., Gonzalgo, M.L., Chan, M.F., Tsai, Y.C., Dean, M., Steven, K., Horn, T. & Jones, P.A. Evidence for two tumor suppressor loci associated with proximal chromosome 9p to q and distal chromosome 9q in bladder cancer and the initial screening for GAS1 and PTC mutations. *Cancer Res* **56**, 5039-43 (1996).
255. Simoneau, M., Aboulkassim, T.O., LaRue, H., Rousseau, F. & Fradet, Y. Four tumor suppressor loci on chromosome 9q in bladder cancer: evidence for two novel candidate regions at 9q22.3 and 9q31. *Oncogene* **18**, 157-63 (1999).
256. Wada, T., Berggren, P., Steineck, G., Adolfsson, J., Wijkstrom, H., Norming, U., Hansson, J., Hemminki, K. & Larsson, P. Bladder neoplasms--regions at chromosome 9 with putative tumour suppressor genes. *Scand J Urol Nephrol* **37**, 106-11 (2003).
257. Devlin, J., Keen, A.J. & Knowles, M.A. Homozygous deletion mapping at 9p21 in bladder carcinoma defines a critical region within 2cM of IFNA. *Oncogene* **9**, 2757-60 (1994).
258. Bullinger, L., Dohner, K., Bair, E., Frohling, S., Schlenk, R.F., Tibshirani, R., Dohner, H. & Pollack, J.R. Use of gene-expression profiling to identify

prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* **350**, 1605-16 (2004).

259. Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., Montgomery, K., Chi, J.T., van de Rijn, M., Botstein, D. & Brown, P.O. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* **2**, E7 (2004).
260. Sobin LH, W.C. TNM classification of malignant tumors (ed 6th). (New York: John Wiley & Sons, 2002).
261. Mostofi FK, D.C., Sesterhenn I. Histological typing of urinary bladder tumors. International classification of tumors (ed 2). (Springer, Berlin, 1999).
262. Heyer, L.J., Kruglyak, S. & Yooseph, S. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* **9**, 1106-15 (1999).
263. Pavlidis, P. & Noble, W.S. Analysis of strain and regional variation in gene expression in mouse brain. *Genome Biol* **2**, RESEARCH0042 (2001).