Genetic Analyses of Tumor Progression in Colorectal Cancer

Kristina Lagerstedt

GÖTEBORGS UNIVERSITET

Department of Surgery/Institute of Clinical Sciences
Department of Chemistry/Institute of Biomedicine
The Sahlgrenska Academy at University of Gothenburg
Sweden, 2009

ISBN 978-91-628-7817-7

Printed by Reproservice, Chalmers, Sweden

To my wonderful family

ABSTRACT

Colorectal tumors are responsible for more than 600 000 deaths per year worldwide and thereby constitute the second most common cause of cancer related mortality. Early detection is related to improved prognosis and identification of genetic biomarkers would meliorate available diagnostic tools. Existing tumor classification systems lack precise monitoring within individual tumor stages in relation to progression. Therefore, we performed genetic characterization of tumor progression by analyses of colorectal tumors and normal colon mucosa. We used combined microarray analysis to obtain a set of candidate biomarkers, starting with genome-wide array-based DNA analyses to screen for tumor-specific aberrant DNA patterns followed by correlations to the associated changes in mRNA and microRNA expression. We also investigated the relation between functional *p53* and tumor progression as well as survival in patients with colorectal cancer (CRC). Furthermore, we used high resolution oligonucleotide array based CGH to identify nonpolymorphic structural variation in DNA from normal colon biopsies from patients with confirmed CRC to reveal candidate regions with association to putative familial CRC genetic variants.

Colorectal tumor progression is proposed to follow a step-wise transformation from normal cells into malignant tumors, and therefore we used different stages within this model to summarize our results, in terms of genetic events of potential importance. First, gain in parts of chromosome 20 encompassing *AURKA,* as well as alterations in *p53 (17p13.1)*, may be involved in the development from adenoma to carcinoma. Second, loss of 18q and gain of 8q harboring *SMAD7* and *PTP4A3* appear to rise during progression defined as early (Dukes A and B) to late (Dukes C and D) tumor stage. Third, distant metastatic potential may be associated to loss of 8p and increased expression of miR-373. Fourth, putative structural variants observed in normal colon mucosa may predispose for the onset of malignant transformation in familial sporadic CRC. Finally, there is a clear relationship between increased properties of aberrant DNA content as well as the number of combined genetic events and tumor progression.

We conclude that correlated changes in DNA and RNA abundance may represent a robust rationale for selection of genetic biomarkers. Moreover, our results also suggest that Dukes D tumors possibly develop in a way that does not fit into the stepwise progression model, illustrated by earlier onset and less genetic aberrations. These results represent a set of genetic events that can hopefully contribute to improved procedures considering diagnosis and prognosis in CRC patients by providing genetic biomarkers.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Varje år avlider mer än 600 000 människor i världen av cancer i tjock- och ändtarm, som är den tredje vanligaste cancerformen och näst vanligaste cancerrelaterade dödsorsaken i Sverige och världen. De patienter som diagnostiseras med tjock- eller ändtarmscancer på ett tidigt stadium har större möjligheter till en tidigare insatt behandling och därmed oftast en bättre prognos. Därför är det viktigt att kunna identifiera tidiga tecken på cancersjukdom samt att förstå den genetiska orsaken till varför vissa tumörer sprider sig (metastaserar).

Cancer är en genetisk sjukdom, som uppkommer genom förändringar i cellernas arvsmassa. Arvsmassan består av kromosomer som innehåller DNA. I DNA-sekvensen finns gener som fungerar som ritmall för de proteiner som cellerna behöver tillverka för sin funktion. Som budbärare mellan DNA och proteiner använder kroppen ytterligare en molekyl som kallas RNA. I denna avhandling har både kromosomer, DNA och RNA studerats i tumörceller från tjock- och ändtarmscancer och jämförts med arvsmassa från normala celler. Kromosomala avvikelser innebär att den normala sammansättningen av DNA är rubbad, vilket kan ge upphov till förändringar i de proteiner som generna kodar för. Fundamentalt för avhandlingen var att förstå tumörernas genetiska ursprung och utveckling (progression).

Vi har använt avancerade DNA-tekniker som möjliggör analys av hela det mänskliga genomet på samma gång. Tusentals fragment av arvsmassa (DNA), gener (DNA eller RNA) och kontrollelement för hur generna styrs (mikroRNA), mäts genom att man kopplar olika färger till DNA, RNA och mikroRNA och jämför skillnaderna i färgintensitet mellan tumörceller och normalceller. Denna teknik kallas microarray och har utvecklats under det senaste årtiondet.

De tumörspecifika kromosom förändringar som identifierades, kopplades sedan till olika stadier av sjukdomen, från den enklaste formen då patienten har god prognos till avancerad metastaserande cancer med dålig prognos. Vi studerade även överlevnad, genom att jämföra patienter som friskförklarats och överlevt mer än tio år efter operation med patienter som avlidit inom ett år efter operation, till följd av tumörsjukdom. Resultatet visade att mycket tydliga förändringar i stora delar av kromosom åtta var kopplade till metastaserande tumörer och tumörer från patienter med kort överlevnad. Stora avvikelser identifierades också i andra kromosomer med koppling till andra tumörstadier.

I nästa steg jämförde vi därför om det fanns en koppling mellan generna och deras proteinprodukter. För att få en uppfattning om hur mycket proteinproduktionen ökar eller minskar som en följd av genförändringen, mäter man RNA. Våra resultat visade att kromosom 20 i samtliga tumörstadier innehöll gener som var förändrade på både DNA- och RNA- nivå och därigenom troligen är kopplade till tumöruppkomst. En av generna, *AURKA,* har tidigare kopplats till ett förstadium till tumörer i tjock- och ändtarm. *p53* är en annan gen som är viktig för tumörutveckling, men som enligt vår studie saknade betydelse för prognos. Totalt identifierades 36 gener och två mikroRNA som kopplades till metastaserande tumörer

och som återfanns på kromosom 7, 8, 13 och 18. Dessa gener bör vidare studeras genom utökade analyser, för att konfirmera om de kan kandidera som prognosmarkörer.

Dagens forskare ägnar stort fokus åt att identifiera skillnader i den normala mänskliga arvsmassans sammansättning. På så vis hoppas man kunna koppla dessa skillnader, som är betydligt fler än förväntat, till olika sjukdomstillstånd, tex. cancer. I det fjärde delarbetet undersökte vi DNA från normal tarmslemhinna och blod från patienter med cancer i tjock- och ändtarm och fann ett antal DNA förändringar som potentiellt kan komma att kopplas till en ökad risk att utveckla sjukdom (familjära riskfaktorer). Dessa förändringar behöver dock verifieras i DNA från ett större antal patienter.

Sammantaget har våra analyser resulterat i en mängd fynd i arvsmassa med kopplade genetiska förändringar som förhoppningsvis kan leda till en förbättrad diagnostik och/eller prognosbedömning av patienter med tumörer i tjock- och ändtarm.

ORIGINAL PAPERS

This thesis is based on results reported in the following papers:

I. Lagerstedt, K. K., Kressner, U., Lonnroth, C., Nordgren, S., Lundholm, K. (2005) The role of combined allelic imbalance and mutations of *p53* in tumor progression and survival following surgery for colorectal carcinoma. *International Journal of Oncology 27(6)*, 1707-15.

II. Lagerstedt, K. K., Staaf, J., Jönsson, G., Hansson, E., Lonnroth, C., Kressner, U., Lindström, L., Nordgren, S., Borg, Å., Lundholm, K. (2007) Tumor genome wide alterations assessed by array CGH in patients with poor and excellent survival following operation for colorectal cancer. *Cancer Informatics, 3*, 351-365.

III. Lagerstedt, K.K., Kristiansson, E., Lönnroth, C., Andersson, M., Iresjö, B.M., Gustafsson, A., Hansson, E., Kressner, U., Nordgren, S., Enlund, F., Lundholm, K. (2009) Genes with relevance for early to late progression of colon carcinoma suggested from results in combined microarray analyses. *Manuscript.*

IV. Lagerstedt, K.K., Kristiansson, E., Lönnroth, C., Andersson, M., Gustafsson, A., Hansson, E., Kressner, U., Nordgren, S., Enlund, F., Lundholm, K. (2009) Copy number variation in normal colon mucosa from patients with primary colorectal cancer. *Manuscript.*

.

TABLE OF CONTENTS

ABBREVIATIONS

AI	Allelic Imbalance
BAC	Bacterial Artificial Chromosomes
CAM	Cell Adhesion Molecule
CGH	Comparative Genomic Hybridization
CIN	Chromosomal Instability
CNV	Copy Number Variation
CRC	Colorectal Cancer
DGV	Database of Genomic Variants
DNA	Deoxyribonucleic Acid
FAP	Familial Adenomatous Polyposis
FC	Fold Change
FDR	False Discovery Rate
GWAS	Genome Wide Association Study
HNPCC	Hereditary Non Polyposis Colorectal Cancer
Mb	Mega bases
MCR	Minimal Common Region
MMR	Mutation Mismatch Repair
MSI	Microsatellite Instability
MSS	Microsatellite Stable
LOH	Loss of Heterozygocity
microRNA	micro Ribonucleic Acid
MIAME	Minimal Information About a Microarray Experiment
mRNA	messenger Ribonucleic Acid
siRNA	small interfering Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
TSG	Tumor Suppressor Gene

SUPPLEMENTARY MATERIAL

Supplementary List 1	DNA segments with Minimal Common Regions in Dukes A-D tumors.
Supplementary List 2	Genome-wide and per chromosome interactions between DNA alterations and RNA expression in Dukes A-D tumors.
Supplementary List 3	Segment interactions of DNA alterations and mRNA expression in Dukes A-D tumors.
Supplementary List 4	Differentially expressed genes within aberrant DNA segments.

Supplementary lists are available upon request.

INTRODUCTION

Tumor Development and Progression

For more than half a century, huge efforts including considerable financial investments have been assigned investigations of tumor genetics, to understand and ultimately defeat malignant disease. These efforts provide substantial amounts of genetic data within the field of cancer, that most likely harbor numerous therapeutic targets that will hopefully be uncovered within a reasonable period of time.

Paradigms of Genetics

Tumor biology describes development and transformation of normal cells to malignant neoplastic cells – in other words, tumorigenesis. Scenarios involved in cancer formation can be described by parallels to historical paradigms from genetic research, from G Mendel and C Darwin to currently active scientists like e.g. B Vogelstein, R Weinberg and 2006 Nobel Prize winner in Medicine and Physiology, A Fire. The choice of these scientists among many successful coworkers worldwide was based on the fact that their research has provided essential evidence in the field of genetics and biology for the understanding of parts in tumor biology that are central for the specific contents in this thesis.

In synopsis, the scientific observations by Mendel and Darwin conclude the very basics of genetics and modern evolutionary biology, namely natural selection together with clonal growth advantages. Observations from human cancers and animal models argue that the progressive conversion of normal human cells into tumor cells follows a process analogous to evolutionary biology. Hence, the obtained genetic changes give rise to growth advantages subsequently leading to tumor development (Foulds 1954; Nowell 1976; Hanahan and Weinberg 2000).

Genetic Instability

Human cancer is a genetic disease and thus, understanding the molecular basis of tumor associated genetic defects is crucial. *Genetic instability* illustrates these defects and represents the focus in each of the papers described and discussed in this thesis. *Genetic instability* refers to alterations in genomic structure and is typically categorized into two major classes, comprising instability at chromosome or nucleotide level (Nowak, Komarova et al. 2002).

Tumors with *chromosomal instability, CIN,* represent the majority and display either *numerical* aberrations i.e. aneuploid karyotype, which refer to abnormal numbers of chromosomes in tumor cells (Boveri 1914) or *structural* aberrations. The later is further subdivided into *balanced* and *unbalanced instabilities.*

Balanced instabilities most commonly refer to chromosomal translocations and were until recently mainly associated to hematological malignancies, but now also reported in solid tumors (Rabbitts 1994; Mitelman, Johansson et al. 2004). Translocations are often related to a specific tumor type and either cause chimeric fusion genes or altered expression of

structurally normal genes, and can therefore also be used as diagnostic biomarkers (Frohling and Dohner 2008). No tumor specific chromosomal translocation in colorectal tumors has yet been identified.

Unbalanced instabilities are typically gains and losses of various size located throughout the genome. Many of these rearrangements are considered to be secondary events for tumor formation and sometimes involved in tumor progression. In colorectal tumors, gain of chromosome 8q is related to a more aggressive behavior and a less favorable prognosis (Ghadimi, Grade et al. 2003). The functional consequences of the majority of unbalanced instabilities are still unknown although a number of genes within altered regions, such as *p53*, *MYC* and *ERBB2*, have been identified and characterized in relation to cancer. Amplification of *ERBB2* and subsequent overexpression of the protein represents a target for the monoclonal antibody *Trastuzumab*, which is mainly used in treatment of breast cancer patients (Romond, Perez et al. 2005). Amplification of *ERBB2* also occurs in approximately 3 % of colorectal tumors (Nathanson, Culliford et al. 2003).

Several chromosomal regions of recurrent gain and loss, but without any obvious target genes, have been associated to one or several cancer types. Discovery of new tumor related genes within these regions is facilitated by comparisons to corresponding expression profiles, in order to filter out specimens without differential expression. These regions may also contain noncoding genes, such as microRNAs, possibly affected by genomic alterations and consequently future putative anti-cancer targets. These aspects are further described in Paper II and III. However, the cause of *CIN* still remains poorly understood but implications in different pathways and processes have been proposed, including telomere and centromere dysfunction and inactivation of proteins involved in mitotic spindle and DNA replication checkpoints (Grady 2004).

The *nucleotide related instability* involves small-scale point mutations, subdivided into substitution, deletion or insertion of one or a few nucleotides. Instability at the nucleotide level can arise from defective DNA repair systems, such as inactivation of DNA mismatch repair (MMR) genes *MSH2* and *MLH1* (Thibodeau, Bren et al. 1993; Peltomaki and Vasen 1997). Consequently, errors at the nucleotide level that occur during replication can not be effectively repaired, which lead to increased rates of nucleotide sequence mutations. Deficient repair is detected by analysis of *microsatellite instability (MSI)* that is used for diagnosis of patients with Hereditary Non Polyposis Colon Cancer (HNPCC) (Boland, Thibodeau et al. 1998). MSI is also involved in as many as 15% of sporadic colorectal tumors (Grady 2004). These cancers, unlike the majority of solid tumors (Lengauer, Kinzler et al. 1997), often retain a diploid karyotype, i.e. a normal set of chromosomes.

Genetic events in tumorigenesis

In 2000 Hanahan and Weinberg postulated the "*Hallmarks of Cancer*", as a panel of six main control systems that monitor homeostasis within the cell and normally act as defense mechanisms toward tumor development (*1-6* as indicated below). According to the authors,

disruption of these systems subsequently enables tumorigenesis and each of which will be presented below.

Normal cells require mitogenic i.e. growth signals to proceed into an active proliferative state in the cell cycle. Tumor cells have an acquired capability of *1) self sufficiency in growth signals,* which bring the cell to a more or less constant proliferative state. Responsible for the growth signal autonomy are excessively active oncogenes, consequently accelerating proliferation in the neoplastic cell. Yet another tumor cell characteristic is *2) insensitivity to antigrowth signals* which would normally suppress growth by induced cellular differentiation or force the cell into a quiescent state. Instead, the tumor cells monitor antiproliferative signals by modified receptors or pathways related to *Rb* or *p53* proteins and remain proliferative. The *p53* protein and its role in tumor development and colorectal cancer will later be described in more detail (Figure 1). In contrast to the direct accelerated proliferation due to excess oncogenic activation, tumor suppressor genes (TSGs) repress antigrowth signals mechanisms and thus indirectly promote growth.

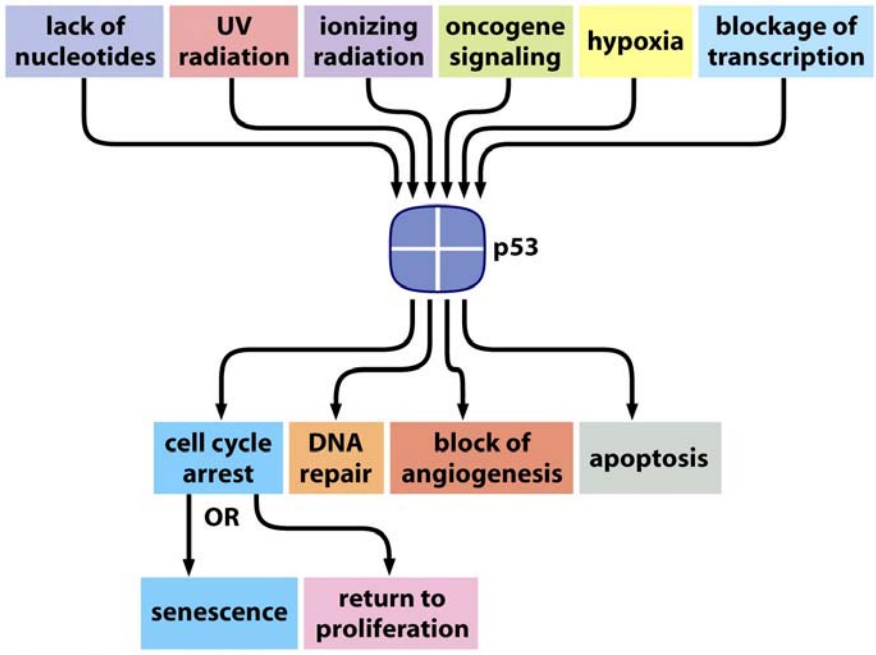


Figure 1. A variety of cell-physiological stresses cause rapid increase in p53 levels. Accumulation of p53 subsequently induces a number of cellular responses leading to downstream effects.

Moreover, the acquired resistance toward programmed cell death, to *3) evade apoptosis* allows tumor cells to survive and continue their growth despite circumstances such as cellular stress and genomic damage. Key regulators of apoptosis are the *caspase* proteins, a family of cysteine proteases that have been termed executioner proteins. In addition and independent of the growth limiting systems, tumor cells become immortalized by developing *4) limitless replicative potential.* Normal cells have a finite number of replication cycles, which is

determined by telomeric lengths denoted by the *Hayflick limit* (Hayflick 1965; Harley 2008), directly associated to successive telomeric chromosomal shortening. Tumor cells, however, develop abilities to maintain telomeres and thereby become immortalized.

Since a continuous supply of nutrients and oxygen is essential for tumor as well as normal cell growth, induced and *5) sustained angiogenesis* is necessary for the tumor to progress to a larger size. The ultimate step of tumor progression involves *6) tissue invasion and metastasis* which constitute the major cause of cancer related deaths. Metastatic spread involves cell-cell adhesion molecules (CAMs), extracellular matrix components and proteases, which mechanisms are complex and still incompletely understood (Mehlen and Puisieux 2006).

In summary, Hanahan and Weinberg suggested that six cellular processes must be disrupted to enable transformation of a normal cell into its neoplastic counterpart. However, a recent sequence evaluation of colon and breast cancer genomes indicate that the number of altered cellular processes required for tumorigenesis may be even higher (Sjoblom, Jones et al. 2006). Clearly, certain genetic instability events affect essential cell cycle activities and regulation with evident tumor promoting consequences (Hanahan and Weinberg 2000).

Oncogenes and Tumor Suppressor Genes

At the gene level, DNA alterations such as gains, losses, translocations and point mutations may give rise to *oncogenes* and affect *tumor suppressor genes*. The first *oncogene* was discovered in 1976, when *SRC* was identified to manage transformation of normal cells into tumor cells mediated by Rous sarcoma virus (Stehelin, Varmus et al. 1976). Oncogenes accelerate growth in tumor cells and are formed by amplification, translocation or point mutation of their normal analogues, called protooncogenes. In colorectal and other cancers, mutated *K-RAS* and *MYC* constitute well defined oncogenes (Forrester, Almoguera et al. 1987; He, Sparks et al. 1998) which are frequently amplified in many cancer types. A novel colorectal oncogene, *CDK8* was recently reported and locates within a recurrently amplified region on chromosome 13 (Firestein, Bass et al. 2008).

Mutations and deletions are genetic events closely associated to *tumor suppressor genes* (TSGs) and results in alterations that partly or completely deplete the normal growth inhibiting abilities of the unaltered TSG. A classic example of a TSG is the *Retinoblastoma* (*RB*) gene, described in A Knudson's twin-study of the retinoblastoma disease where the author presented his "two-hit hypothesis" (Knudson 1971). According to the hypothesis, two genetic "hits" i.e. mutations (point mutations and allelic loss) affecting both DNA strands (alleles) in the cell, are required to ablate the TSG activity and thus promote tumor progression. However, today it is clear that yet other mechanisms such as *DNA methylation* serve as efficient silencers of tumor suppressor gene activity (Jones and Baylin 2002).

In *haploinsufficient* genes, only one genetic event affecting one allele is required to cause functional inactivation and thereby enable tumorigenesis. In this case, the specific TSG need two functional copies of the gene to express a normal protein (Fero, Randel et al. 1998;

Santarosa and Ashworth 2004). Beside the *RB* gene, well characterized TSGs includes the *APC* gene that is frequently mutated in cancers and plays a central role in several cellular processes e.g. the WNT-signalling cascade (Nagase and Nakamura 1993; Fodde, Kuipers et al. 2001) and *p53* that represents an haploinsufficient TSG (Santarosa and Ashworth 2004). Aspects of *p53* is further described in Paper I. TSGs have been assigned still other important roles in tumor progression, due to their involvement in the maintenance of genomic integrity (Lengauer, Kinzler et al. 1997).

p53-"Guardian of the Genome"

The p53 protein was discovered in 1979 and has gained a lot of attention since then, titled "Molecule of the Year" in 1993 and currently, a p53 PubMed search results in more than 50 000 hits. The p53 protein is a transcription factor (Vousden and Prives 2005) and in its normal condition represents an important factor in tumor surveillance. Accordingly, p53 was denoted "guardian of the genome" in 1992 (Lane 1992). The protein is implicated in several pathways and processes, namely differentiation, senescence, antiangiogenesis, cell cycle arrest and programmed cell death, indicating its fundamental roles in maintenance of cellular homeostasis (Lane 1992; Vousden and Lu 2002; Levine, Hu et al. 2006).

Normally, cellular stress such as genetic damage, radiation or imbalanced mitogenic signals induces increased p53 expression that promotes p53 mediated cell growth inhibition and ultimately cell death through apoptosis. In the absence of cellular stress, the p53 interact with its main regulator MDM2 (Kubbutat, Jones et al. 1997) and other proteins that keep p53 expression levels low to avoid otherwise harmful effects on normal growth and development. However, there are a number of ways by which the p53 protein is inhibited to mediate its regular activities in response to stress factors (Figure 1). *p53* is mutated in more than 50% of human cancers and since *p53* is a tumor suppressor gene, protein function is repressed or lost by mutations, most frequently missense, or deletion of the entire or parts of the gene (Harris 1993). In 1989, Baker et al. reported p53 mutation in one copy of the allele combined with deletion of the other in human colon cancers, thus fulfilling the cardinal criteria of the "two-hit"-hypothesis (Baker, Fearon et al. 1989). However, *p53* is not an ordinary TSG since it proved to be haploinsufficient in studies performed on *p53* heterozygous mice (Venkatachalam, Shi et al. 1998). Hence, only one genetic event can be enough to silence the protein and thus, *p53* does not follow the characteristic principle of TSGs. Furthermore, a mutated p53 protein may inhibit normal activity by interfering with wild-type p53 proteins during the formation of homotetramers, the functional p53 unit. The tetramer is thereby deactivated in a dominant–negative fashion, despite only one single p53 mutant protein participated in oligomerization (McLure and Lee 1998). *p53* was the target of analyses performed in Paper I.

Regulation of gene expression

Measurement of protein abundance is crucial. However, due to their complex structure and chemical properties, proteins are difficult to study in large scale experiments. Therefore, their

originators, mRNAs, are frequent targets for analysis and subsequent implication in tumors among other diseases. Apart from immediate gene expression studies, two further categories can explain or mirror altered expression, namely *epigenetic events* and *post-transcriptional regulation* by *noncoding RNAs*.

Epigenetics describe effects such as *chromatin structure remodeling* or *DNA methylation* on gene transcription (Jones and Baylin 2002). In cancer, gene silencing by hypermethylation of promoter regions occur frequently and has been described in most tumor types (Herman and Baylin 2003). Promoter hypermethylation of mismatch repair (MMR) gene *MLH1* occur frequently in MSI positive colorectal tumors (Herman, Umar et al. 1998). *Post-transcriptional regulation* involves *siRNA* and *microRNA* that specifically downregulate mRNA levels through RNA-RNA base-pairing. Until recently, siRNAs were only identified endogenously in prokaryotic organisms although synthesized siRNAs have been frequently used for eukaryotic gene silencing for several years (Fire, Xu et al. 1998) (Watanabe, Totoki et al. 2008). MicroRNAs, however, represent endogenous molecules and will be described below.

microRNA

A Fire together with C Mello received the Nobel Prize in Medicine and Physiology (2006) for discoveries about genetic silencing by small noncoding RNAs, also referred to as RNA interference. Fire's work also comprise microRNAs, a class of noncoding RNAs that have gained a lot of focus during recent years, since they represent a promising target molecule for both diagnostics and treatment in complex genetic diseases, e.g. cancer. As mentioned above, microRNAs are involved in translational regulation by binding to complementary sites in target mRNA transcripts and thereby mediate translational repression or transcript degradation (Pillai, Bhattacharyya et al. 2007).

Today 718 human microRNAs have been discovered (mirBase version 13, NCBI36, update 2009-03-08)(Griffiths-Jones, Saini et al. 2008) and the total number of microRNAs in the genome is estimated to several thousands. Furthermore, a single microRNA can be associated to and regulate a variable number of genes. Specific genes can interact with several microRNAs, for instance *p53* that have been associated to 16 different microRNAs (miRBASE Target Database, Welcome Trust Sanger Institute, microrna.sanger.ac.uk). It is likely that microRNAs play essential roles in tumor development since expression profiles demonstrate that many microRNAs are downregulated in tumors compared to normal tissue (Blenkiron and Miska 2007). MicroRNAs have been attributed oncogenic potential and may also act as tumor suppressors (Negrini and Calin 2008). The role of microRNAs in colorectal tumorigenesis was investigated in Paper III.

Tumor Progression

In 1957, J Foulds defined *tumor progression* as the irreversible development of a neoplasm by stepwise qualitative changes of one or more of its characters (Foulds 1957). Later, Foulds and other authors suggested that this biological and clinical progression may reflect a sequential increase of genetically altered cellular subpopulations of tumors, gaining new characteristics

(Cairns 1975; Nowell 1976; Nowell 1986). Today, scientists such as R Weinberg have provided robust models of the synthesized current knowledge applied in a genetic context, comprising different genetic events within altered pathways. These models either reflect tumor progression in general or are typical for a specific tumor type, like colorectal cancer (e.g. the Vogelstein model, described below). In summary, normal cells evolve into cells with increasingly neoplastic phenotype during a sequence of randomly occurring structural DNA alterations and epigenetic events, driving tumor progression. This theory is denoted the *clonal evolution theory* (Nowell 1976). Malignant cell phenotypes thereby emerges through progression based on these genetic events and subsequently altered pathways, described above (the Hallmarks of Cancer), in favor of the cancer phenotype. Beside the clonal evolution theory, the *cancer stem cell theory* has gained a lot of interest the last decade, where tumor initiation occurs in a stem or progenitor cell (Boman and Huang 2008).

Colorectal Cancer

Cancer of the colon and rectum constitutes two separate cancer types but are referred to as colorectal cancer (CRC), since they share a lot of characteristic properties. CRC is the third most common among cancers and the second cause of cancer related death worldwide and in Sweden. Risk factors for development of colorectal tumors include tobacco smoking, a diet high in total fat and meat, obesity and sedentary lifestyle. Detection of hemoglobin (F-Hb) in feces was recommended for CRC screening by the European Commission in 2003, together with colonoscopy, which is used as the main screening tool for colorectal polyps and tumors in the United States and several European countries (WHO 2002; ACS 2007; Epidemiologiskt Centrum 2007; Cancerfonden 2008). Colonoscopy decrease CRC incidence through early detection and removal of polyps and thus decrease mortality (Winawer, Zauber et al. 1993). Surgery is the primary treatment strategy and some patient groups further receive preoperative radiotherapy (rectal cancers) and adjuvant treatment by chemotherapy.

CRC can be subdivided into hereditary (< 5 %), familial (20-25 %) and sporadic (75 %) disease (www.cancer.gov) (Cardoso, Boer et al. 2007; NCI 2009). Thus, dominantly inherited variants only accounts for a small fraction of colorectal tumors, among which hereditary nonpolyposis colon cancer (HNPCC) and familiar adenomatous polyposis (FAP) are the most common and well defined. Hereditary CRCs are based on specific genetic alterations that give rise to colorectal tumors, typically with an earlier onset possibly caused by a faster development from adenoma to carcinoma (Cheah 2009). Nevertheless, patients with hereditary variants have improved survival compared to patients with sporadically occurring tumors (Lynch and de la Chapelle 2003).

Among patients with sporadic disease, several carry a familiar risk genotype (Goldgar, Easton et al. 1994; Easton and Eeles 2008), not to be confused with well characterized hereditary disease. However, genetic loci responsible for the risk genotype are largely unknown but, interestingly, sibling studies have estimated that approximately 35% of all CRC cases can be attributed to genetic susceptibility (Lichtenstein, Holm et al. 2000). Furthermore, recent studies suggest even higher rates based on the presence of rare predisposing genetic variants

(McCarroll, Kuruvilla et al. 2008) (Described below, Future CRC Prediction and Diagnostics). However, the work summarized in this thesis was primarily focused on bringing further clarity into the genetic events implicated in malignant progression of sporadic colorectal tumors.

Tumor Staging

Detailed pathological analyses have provided evidence for the multistep sequence of events that give rise to most cancer types. Colorectal tumors are of epithelial origin and pathologically classified into three major categories, namely nonneoplastic polyps, neoplastic polyps (premalignant adenomas) and cancers. The cancer category represents 95% of all colorectal tumors and constitutes the focus of this thesis. For many years, Dukes classification system (Dukes and Bussey 1958) (Table 1) was the golden standard for tumor staging by pathologists worldwide. Today, the TNM-system has replaced Dukes for classification of the anatomic extent of cancer spread (Table 1) (Sobin and Fleming 1997). In this thesis, however, the Dukes A-D classification has been used for tumor staging in relation to progression.

Table 1. TNM classification describes the local, regional and distant extent of solid tumor spread and is considered the standard cancer staging system. Dukes staging system is specific for colorectal tumor classification and was earlier considered the standard strategy for CRC staging.

Dukes	Tumor spread/localization	TNM*
A	Submucosa	T1, N0, M0
	Muscularis propria	T2, N0, M0
B	Beyond muscularis propria	T3, N0, M0
	Adjacent organs	T4, N0, M0
C	1-3 lymph node metastasis	T1-4, N1, M0
	≥ 4 lymph nodes metastasis	T1-4, N2, M0
D	Distant organ metastasis	T1-4, N0-2, M1

* T = local extent of primary tumor; N = regional lymph nodes; M = distant metastasis, N0 or M0 = no lymph nodes or distant metastases.

Genetic Instability in colorectal tumors

In sporadic CRC, CIN (85 % of all cases) and MSI (15 %) characterize genetic instability pathways and few examples of balanced instabilities, i.e. translocations, have been identified in colorectal tumors. Until recently, the reports on translocations in solid tumors were fairly limited and primarily associated to hematological malignancies (Rabbitts 1994; Mitelman, Johansson et al. 2004). This possibly reflects the more pronounced complexity in solid tumors which hence demands intricate characterization and selection among numerous genetic events of putative importance. However, MSI tumors predominantly display a diploid karyotype, while CIN type tumors are generally aneuploid (Lengauer, Kinzler et al. 1997). Nevertheless, MSI and CIN CRC tumors both display *APC* and *K-RAS* mutations that occur during transformation from normal epithelium to malignant carcinoma (Grady 2004).

The Microsatellite Instability pathway (MSI)

MSI tumors result from inactivation of crucial genes involved in the mutation mismatch repair (MMR) system and thereby promotes tumorigenesis through generation of target gene

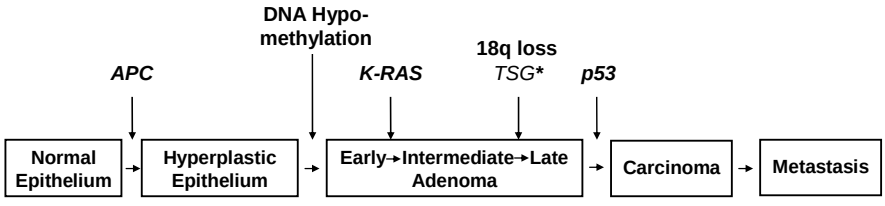
mutations. As mentioned above, MSI is characteristic for HNPCC but also found in as many as 15 % of sporadic tumors (Grady 2004). Genes involved and inactivated by either mutation or hypermethylation include *MSH2, MSH3, MSH6, MLH1, PMS1* and *PMS2* .

The Chromosomal Instability pathway (CIN)

CIN tumors carry considerable proportions of structural variation in terms of gained and lost chromosomal regions, recurrently identified on chromosomes 7, 8, 13, 18 and 20 (Ried, Heselmeyer-Haddad et al. 1999; Hermsen, Postma et al. 2002; Lagerstedt, Staaf et al. 2007; Camps, Grade et al. 2008). The origin of CIN is elusive, as previously mentioned, but involvement of the *APC* gene has been proposed (Fodde, Kuipers et al. 2001). Mutations in *p53*, loss of *SMAD7* and amplification of *AURKA* are gene specific events presented in several reports for implication in CIN mediated tumor progression (Fearon and Vogelstein 1990; Lagerstedt, Kressner et al. 2005; Broderick, Carvajal-Carmona et al. 2007; Carvalho, Postma et al. 2009). The roles and order of these genes and genetic events in relation to colorectal tumor progression are further described in the next section.

Vogelgram – a progression model for colorectal cancer

In 1990, B Vogelstein and coworkers presented the multistep model for colorectal carcinogenesis, the "Vogelgram" (Fearon and Vogelstein 1990) (Figure 2) and since then the model has been adopted and used by researchers worldwide. The *Vogelgram* is still considered valid, although a number of genetic events need to be added in order to obtain a refined picture. The adenoma to carcinoma sequence is initiated by inactivation of *APC* (5q21) which takes place in the normal epithelium, resulting in accumulation of *beta-catenin* that subsequently increase during stepwise development. The next genetic event involves hypomethylation and occurs in the state of a hyperplastic polyp. *K-ras* mutations are identified in slightly larger adenomas following loss of 18q-arm during transition to late adenomas. Several TSGs, such as *SMAD7,* involved in *TGF-β* and *WNT*-signaling (Broderick, Carvajal-Carmona et al. 2007), have been suggested as the target gene for 18q loss. Ultimately, loss of the 17p-arm includes the *p53* gene in the final progression from late adenoma to carcinoma (Fearon and Vogelstein 1990).



Adopted by
Fearon & Vogelstein, Cell 1990.

Figure 2. Transformation of normal colon epithelium into malignant carcinoma by step-wise accumulation of five genetic aberrations, as presented by Fearon and Vogelstein in 1990. This progression model for colorectal tumorigenesis is referred to as the "*Vogelgram*". * = the region is proposed to harbor a tumor suppressor gene (TSG).

Yet other genes and genetic events have been proposed to contribute to colorectal tumor development and metastasis since the presentation of the Vogelgram. Gain of chromosome 20q has been observed in adenomas and *AURKA* represents a candidate oncogene located at 20q13.2 (Carvalho, Postma et al. 2009). Recently *CDK8,* that is located within the frequently amplified chromosome 13 (13q12.3), was suggested to display oncogenic abilities in colon cancer cell lines by regulation of β-catenin activity (Firestein, Bass et al. 2008). Moreover, recurrent loss of the 18q-arm indicates the presence of potential tumor suppressor genes. *SMAD7* map to 18q21.1 and has gained a lot of attention lately since it contains a SNP-site proposed a CRC susceptibility locus (Broderick, Carvajal-Carmona et al. 2007). Late chromosomal abnormalities in colorectal tumorigenesis involve loss of the 8p-arm but no associated TSG has yet been identified. However, *PTP4A3* (*PRL-3*) is located to the 8q24.3-region that is often amplified in late stage colorectal tumors and has been related to metastasis (Saha, Bardelli et al. 2001). Some of these genes will be further discussed later in this thesis.

Structural and Copy Number Variation

Structural variation, also referred to as copy number variation (CNV), encompass single nucleotide-- to megabase-sized structural variants of genomic segments such as deletions, segmental duplications, insertions, inversions or complex chromosomal rearrangements. In recent years, it has become clear that CNVs are more common and involves a much greater proportion of the human genome than previously realized (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004; Hinds, Stuve et al. 2005; Sharp, Locke et al. 2005; Tuzun, Sharp et al. 2005; McCarroll, Hadnott et al. 2006; Redon, Ishikawa et al. 2006; Scherer, Lee et al. 2007; Wong, deLeeuw et al. 2007). The development of high resolution techniques have generated the ability to detect and catalogue CNVs *en masse* and further associate them with biological function and complex human genetic diseases (Carter 2007). Array-based analysis of single-nucleotide polymorphisms (SNPs) is widely used for identification of variation and subsequent downstream applications. Multiple genome-wide association studies (GWAS) aimed to associate specific disease genotype to phenotype (McCarroll and Altshuler 2007) have recently identified several susceptibility SNP loci proposed to predispose for CRC (Tomlinson, Webb et al. 2007; Zanke, Greenwood et al. 2007; Houlston, Webb et al. 2008; Jaeger, Webb et al. 2008; Tenesa, Farrington et al. 2008; Tomlinson, Webb et al. 2008). Patterns of structural variation in CRC patients were analyzed in Paper II and IV.

Future CRC Prediction and Diagnostics

During the last years, several genome-wide association studies (GWAS) have presented a number of potential and valid risk sites for each of the four most prevalent cancer types, namely breast, prostate, colorectal and lung cancer (Easton and Eeles 2008). In colorectal cancer, ten risk SNP loci were recently discovered at chromosomal bands; 8q23.3 (Tomlinson, Webb et al. 2008), 8q24 (Tomlinson, Webb et al. 2007; Zanke, Greenwood et al. 2007), 10p14 (Tomlinson, Webb et al. 2008), 11q23 (Houlston, Webb et al. 2008; Tenesa, Farrington et al. 2008), 14q22 (Houlston, Webb et al. 2008), 15q13 (Tomlinson, Webb et al. 2008), 16q22, (Houlston, Webb et al. 2008), 18q21 (Broderick, Carvajal-Carmona et al. 2007), 19q13 and 20p12 (Houlston, Webb et al. 2008). In these studies, large cohorts of

colorectal cancer patients and cancer free individuals were screened for novel and previously identified loci by SNP array technology. The findings of susceptibility loci represent considerable progress toward blood sample mediated screening for CRC risk, although identification of additional risk loci are required for predictive purposes.

SCIENTIFIC AIMS

The general aim of this thesis was to identify chromosomal DNA alterations and corresponding genetic transcriptional events associated to colorectal tumor progression and survival.

The specific aims were to:

- Define aberrant chromosomal DNA regions that discriminate between early and late colorectal tumors
- Determine specific genetic events associated to aberrant DNA that potentially provide biomarkers for tumor progression and survival
- Investigate the importance of *p53* gene functionality in relation to colorectal tumor progression and survival
- Identify structural variation in normal colon mucosa that may predispose for familial colorectal cancer

METHODOLOGICAL CONSIDERATIONS

Study Design and Setup

Tumor, normal colon biopsies and blood samples were collected from all patients operated for CRC during specific time periods (Figure 3) and at two Swedish hospitals specified below.

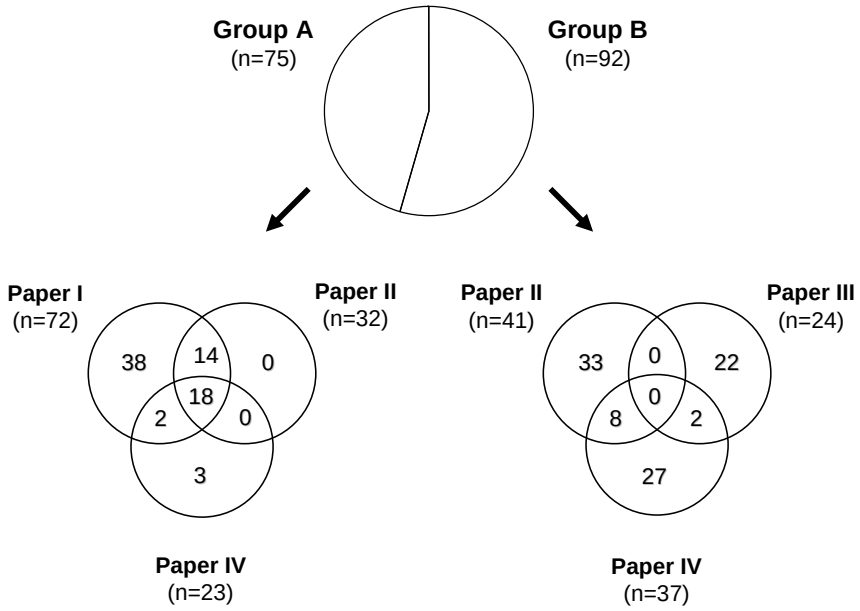


Figure 3. 167 patients diagnosed with primary colorectal carcinoma were analyzed for genetic alterations in relation to their disease, as described in the Paper I-IV. All patients were of Caucasian, Northern European origin and have had surgery as the only treatment. The patients were from 2 different cohorts, based on where and when they were operated. Group A was operated between 1988-1992 in Uppsala, Sweden and Group B was operated between 2001-2005 in Uddevalla, Sweden. The participation of patients in one or several studies is viewed in the diagrams and described in more detail in respective paper included in this thesis.

Paper I

Seventy-two CRC patients operated at Uppsala Hospital (Patient group A, Figure 3), Sweden were included and DNA for downstream analyses was isolated from primary tumors and visibly normal colon mucosa. *p53* specific mutation and LOH profile for each of the 72 patients were analyzed in relation to survival and Dukes stage.

Paper II

A total number of 73 CRC patients were included in this study. 32 were operated at Uppsala Hospital, Sweden (Patient group A, Figure 3) and subgrouped according to survival, while 41 were operated at Uddevalla Hospital, Sweden (Patient group B, Figure 3) and subgrouped according to Dukes system for tumor classification. From each patient, DNA was isolated from tumor tissue and from patients grouped according to survival, also from visibly normal colon tissue. Patients were analyzed with genome-wide arrayCGH to identify stage or survival specific major chromosomal aberrations.

Nineteen patients with recurrence-free, long-term survival (> 10 years after primary surgery), 13 patients with very short survival time (< 1 year after primary surgery) (referred to as "alive" and "dead" in Paper II) and 32 patients according to Dukes A-D stages (n=8) were sub-grouped prior to analysis. For each of the groups described above, DNA was pooled prior to analysis. All hybridizations were performed in competition to a commercially available reference DNA (Clontech, BD Biosciences, Palo Alto, CA, USA) with one exception, where tumor DNA from long and short-term survivors was hybridized together. In addition, RNA from tumor tissue was extracted from 9 patients and subsequently analyzed for global expression patterns.

Paper III

Twenty-four CRC patients who underwent primary surgery at Uddevalla Hospital, Sweden were included in this study (Patient group B, Figure 3). Each Dukes A-D stage was represented by 6 randomly selected patients. From each patient, genomic DNA and total RNA was extracted from the same piece of tumor and visibly normal colon tissue. Patients were analyzed with 3 types of microarrays to identify specific and combined alterations in DNA, RNA and microRNA expression. DNA or total RNA was subsequently pooled according to Dukes staging prior to microarray analyses. The comparison of matched tumor to normal DNA or RNA enabled tumor specific interpretations, since tumor DNA was hybridized in competition with normal colon DNA.

Paper IV

Sixty CRC patients operated at either Uppsala Hospital or Uddevalla County Hospital, Sweden (Patient group A and B, Figure 3) were included and subgrouped (n=15) according to Dukes A, B, C and D histopathological classification system. Patients in Dukes A and B groups had at least 5 years recurrence free survival after primary surgery and patients in Dukes C and D died from colorectal cancer within 38 months after primary surgery. Dukes A and B patients represented a good prognosis group and Dukes C and D patients represented a poor prognosis group. Genomic DNA from visibly normal colon mucosa was isolated from each patient and in addition DNA was also isolated from blood derived from 5 of the Dukes A patients. These blood samples were collected at least 5 years after primary operation (December 2008).

Genomic DNA was pooled in sets of 5 patients, 3 sets (n=5) per Dukes group. Genomic DNA from 5 blood samples was also pooled. Altogether, 12 groups of pooled DNA from normal colon mucosa and 1 group of pooled DNA from blood were analyzed. A standard reference sample, NA10851 (Coriell Cell Repositories, Camden, NJ, USA) was included in the study design to minimize false positive calls in array-based CGH analysis (Carter 2007; Scherer, Lee et al. 2007). Exclusion of reference specific CNVs were enabled by hybridization of the NA10851 reference DNA to another purchased reference sample, a pool of DNA from normal

colon mucosa derived from six human cancer free donors (Biochain Institute Inc, Hayward, CA, USA).

Genetic Analysis

LOH and MSI detection by Microsatellite Allelotyping

Microsatellites constitute highly polymorphic DNA regions composed of tandem repeats and are widely used as molecular markers for genetic applications, such as forensics, population linkage and gene dosage studies (Weissenbach, Gyapay et al. 1992; Urquhart, Kimpton et al. 1994; Goldstein, Ruiz Linares et al. 1995). In cancer research, microsatellites are used to determine Loss of Heterozygocity (LOH) or more correctly, Allelic Imbalance (AI) to evaluate deletion or amplification patterns in specified DNA regions. Moreover, the discovery of Microsatellite Instability (MSI) in CRC and its linkage to HNPCC in 1993 (Thibodeau, Bren et al. 1993) emphasized the clinical importance of microsatellites as biological markers.

In allelotype analysis (i.e. identification of LOH) tumor specific allelic patterns are compared to the corresponding patterns in normal tissue DNA (Skotheim, Diep et al. 2001). Heterozygous alleles display repetitive microsatellite motifs of different lengths and this particular repeat variation is determined by PCR amplification of microsatellite containing regions and subsequent detection of these PCR fragments by capillary electrophoresis. If the relative amounts of tumor and normal alleles are skewed, the relationship is defined as allelic imbalance. Microsatellite Instability (MSI) is characterized by the appearance of novel microsatellite alleles in tumor DNA where the actual number of microsatellite repeats differs from that of the two alleles present in normal tissue DNA. Microsatellite markers simultaneously provide information about MSI as well as LOH. Our intention was to analyze the patients for LOH within the area surrounding p53. A number of tumors displayed DNA with novel alleles that appeared in at least one of the analyzed microsatellite loci. Subsequently, BAT26, a well defined marker for MSI detection (Thibodeau, Bren et al. 1993) (Zhou, Hoang et al. 1998) was added to verify the occurrence of MSI. See Paper I for experimental details (Lagerstedt, Kressner et al. 2005).

Mutation Analysis by DGGE and sequencing

p53 mutations were primarily identified by Denaturant Gradient Gel Electrophoresis (DGGE) (Fischer and Lerman 1983) screening and further characterized by genomic DNA sequencing followed by a second sequencing procedure performed on cDNA (Lagerkvist, Stewart et al. 1994). The principle of DGGE is based on double stranded DNA fragment melting behavior. Mutated DNA fragments denature at different points on an acrylamide gel compared to normal DNA and hence, give rise to different banding patterns. Each mutated fragment was reamplified by PCR and forwarded to sequencing in both sense and antisense direction by capillary electrophoresis separation. For each genomic DNA PCR fragment with verified p53 mutation, the procedure was repeated by sequencing of the corresponding cDNA fragment.

Microarray Analysis

Since the first expression microarray was described in 1995 (Schena, Shalon et al. 1995), DNA microarray technology has become a common and powerful tool for large scale and high resolution genetic analyses. Different applications of the microarray technology have been developed through the years to enable investigations within various research areas such as population genetics and characterization of complex genetic diseases such as cancer. Subsequently, different microarray platforms for different applications have raised and five types of microarrays to study genetic variation in DNA, RNA and microRNA were used in Papers II, III and IV, included in this thesis (described below).

Principles of microarrays

Technically, there are two major microarray categories, namely one-color (e.g. Affymetrix arrays) and two-color arrays (e.g. Agilent arrays). The difference essentially refers to hybridization performance and type of data generated from scanned images. One-color arrays are based on hybridization of one source of DNA or RNA while two-color arrays are based on hybridization of two sample sources and thereby generate a *ratio* derived from the two analyzed samples. One–color performance involves fluorescent dye labeling (often Cy-3, Cy-5 or biotin) of one sample which is then solely hybridized to the array (Paper III). In two-color performance, two samples are labeled with two different dyes (typically Cy-3 and Cy-5) which are simultaneously hybridized to the array in a competitive manner (Paper II, III and IV). After hybridization one- and two-color arrays are treated similarly during scanning where the fluorescence of one or two fluorophores is visualized by laser beam excitation at defined wave lengths and subsequently detected as images of the array slide. Image analysis and quantification is described below.

Array-based Comparative Genomic Hybridization

Array-based Comparative Genomic Hybridization (arrayCGH) comprises high resolution techniques that enable genome-wide identification of genetic aberrations such as amplifications and deletions in genomic DNA samples. In this thesis arrayCGH constitutes the central analysis tool for the purpose of defining tumor progression on the basis of genetic DNA aberrations. Originally, CGH was developed to evaluate DNA copy number changes across the genome by co-hybridization of differently labeled sample and reference DNAs to normal metaphase chromosomes (Kallioniemi, Kallioniemi et al. 1992). Today CGH is essentially improved and replaced by array-based CGH, which was introduced in the late 1990s (Pinkel, Segraves et al. 1998; Pollack, Perou et al. 1999). cDNA, tiling BAC and oligonucleotide arrayCGH platforms were used and details about performance and analysis are described in Paper II, III and IV.

Gene and microRNA expression microarrays

Gene expression analysis in Paper II and III were performed on 44k whole genome ***oligonucleotide microarrays*** (Agilent Technologies) (Skotheim, Lind et al. 2005) and one-

color 15k human ***miRNA microarrays*** (Agilent Technologies) (Wang, Ach et al. 2007) were used for oligonucleotide-based microRNA expression analysis (Paper III).

Preprocessing of Microarray Data

Microarray preprocessing includes several steps where fluorescence intensities from each probe specific spot location on the array are converted into applicable data. Image analysis tools, such as Genepix (Axon Instruments Inc, Foster City, CA, USA) (Paper II) and Feature Extraction (Agilent Technologies) (Paper II, III and IV) were used to extract probe-specific values from the array images. Spotted arrays (BAC and cDNA arrays) require verification of probe location on the array, which is performed by alignment of a grid (a probe locus map) to the microarray image. This step is not necessary for *in situ* synthesized oligo arrays, since each probe position is defined by the array design. Next, signal and background intensities are calculated for each probe and imported into appropriate software for subsequent quantification, normalization and further data analysis. Preprocessing details for microarrays and analysis formats are summarized in Table 2.

Statistical Analysis of Microarray Data

The frequently discussed and well known complexity of microarray data is due to high dimensionality and high levels of noise - which in essence is a statistical problem (Kristiansson 2007). To accomplish the aims of microarray studies it is therefore of great importance to establish clearly defined hypotheses and select appropriate statistical tools for data analysis. Analysis of arrayCGH data aims to determine DNA segment alterations in terms of gain and loss of subchromosomal regions of various sizes. Expression analyses measure genes or microRNA transcripts that are differentially expressed between samples or conditions. Different strategies for data analysis were used depending on the array format and are described in Table 3.

Strategies to combine and correlate microarray data

In Paper III, aberrant tumor specific DNA segments were combined with the corresponding differentially expressed genes and microRNAs. The aim was to sort out regions and genes that represent genetic events present in more than one genetic level (e.g. DNA and mRNA or microRNA). In addition, stage specific alterations were evaluated in order to determine genetic events that discriminate between early (Dukes A and B) and late (Dukes C and D) tumor progression. All combined analyses were based on DNA segments called as aberrant by the CBS DNA segmentation algorithm (Olshen, Venkatraman et al. 2004) and these segments were subsequently tested for overrepresentation of differentially expressed genes or microRNAs as follows.

First, probes from the microarray were mapped to NCBI Entrez (build 18) genes or microRNAs within the region. The proportion of differentially expressed genes was compared to the entire genome and enrichment was then tested using Fisher's exact test. The test of *interaction* was performed for significant alterations over the entire genome, each

Table 2. Methods and software for microarray preprocessing and data analysis presented for each of the 7 formats used in our studies. "BG Sub" is short for background subtraction and refers to correction for background signal intensities for each probe locus prior to analysis. Image analysis software converts probe specific signal intensities of each array into applicable data, summarized in a data result file that is imported and subsequently analyzed in another software or database. "Manufacturer" refers to where microarray slides are produced.

			Preprocessing		Software		
	Microarray	**Channel**	**BG Sub**	**Normalisation**	**Image analysis**	**Statistical Analysis**	**Manufacturer**
Paper II							
27k cDNA	arrayCGH	2-color	Yes	Lowess	GenePix Pro 4.0[1]	BASE[3]	Swegene[5]
32k tiling BAC	arrayCGH	2-color	Yes	Lowess	GenePix Pro 4.0[1]	BASE[3]	Swegene[5]
44k oligo	expression	2-color	Yes	Lowess	Feature Extraction 7.5[2]	GeneSpring 7.2[2]	Agilent
Paper III							
4x44k oligo	arrayCGH	2-color	No	Median	Feature Extraction 9.1[2]	Bioconductor[4]	Agilent
4x44k oligo	expression	2-color	No	Lowess	Feature Extraction 9.1[2]	Bioconductor[4]	Agilent
8x15k oligo	microRNA	1-color	No	Quantile-quantile	Feature Extraction 9.5[2]	Bioconductor[4]	Agilent
Paper IV							
105k oligo	arrayCGH	2-color	No	Lowess	Feature Extraction 9.5[2]	CGH Analytics 3.4[2]	Agilent

Normalisation methods: Lowess (Yang, Dudoit et al. 2002), Median (Smyth and Speed 2003) or quantile-quantile (Bolstad, Irizarry et al. 2003) 1 Axon Instruments Inc, 2 Agilent Technologies 3 BioArray Software Enviroment, http://base.thep.lu.se/ (Saal, Troein et al. 2002), 4 statistical language R 2.7.2 (R Development Core Team 2009), 5 Swegene Centre for Integrative Biology at Lund University (SCIBLU).

Table 3. Preprocessed microarray data analyses were performed by application of suitable strategies for specified datatypes. Further details are available in Paper II-IV.

		Analysis			Array coverage	
	Microarray	**Type**	**Tool**	**Ref**	**Number of clones/probes**	**Human genes/transcripts**
Paper II						
27k cDNA	arrayCGH	DNA aberration	CGH Plotter	1	27k	10978
32k tiling BAC	arrayCGH	DNA aberration	CGH Plotter	1	32k	18900
44k oligo	expression	mRNA expression	No gene specific analysis applied		44k	33000
Paper III						
4x44k oligo	arrayCGH	DNA aberration	CBS	2	44k	19003
4x44k oligo	expression	mRNA expression	Moderated T-statistics, FDR	4	44k	38896
8x15k oligo	microRNA	microRNA expression	Moderated T-statistics, FDR	4	15k	509
Paper IV						
2x105k oligo	arrayCGH	CNV	ADM-2	3	105k	19003

CNV, Copy Number Variation; FDR, False Discovery Rate; CBS, Circular Binary Segmentation; ADM, Aberration Detection Method 2; Ref, References: 1) (Autio, Hautaniemi et al. 2003) 2) (Olshen, Venkatraman et al. 2004) 3) (Lipson, Aumann et al. 2006) 4) (Benjamini and Hochberg 1995; Smyth 2004)

chromosome and each segment according to the CGH microarray. Gene expression patterns for genes within the candidate regions were also evaluated by linear regression (described above), followed by selection of the regulated genes with significant difference between Dukes A, B and C, D or A, B, C and D (Lagerstedt, Kristiansson et al. 2009).

Ethical approval

All analyses performed in Paper I-IV were approved by the Ethics Committee at the Sahlgrenska Academy, Sahlgrenska University Hospital, University of Gothenburg, Sweden.

RESULTS

Combined p53 mutation and LOH in relation to colorectal tumorigenesis and progression (Paper I)

Among the 72 patients included in this study, Dukes stage was a predictor of disease specific mortality. Mutation analyses of all *p53* exons revealed that 60% of the tumors contained mutation in at least one exon and several tumors had more than one mutation. *p53* mutation did not predict survival. Four polymorphic markers, two within (I1p53 and Cd72), one upstream (D17S938) and one downstream (D17S720) of the *p53* gene were analyzed for Loss of Heterozygocity (LOH). 59% of the analyzed tumors showed AI in one or more loci when all four markers were taken into account.

Patients with LOH indicated by the D17S720 microsatellite marker, located downstream of the *p53* gene had decreased survival compared to patients with retained heterozygocity at this locus ($p<0.05$). No differences detected by any other marker or combination of markers were related to survival. 34 % (24/71) of the tumors displayed combined *p53* mutation and LOH, which was considered equal to ablated p53 function by loss of both alleles (non-functional $p53^{-/-}$). The *p53* gene was altered by mutation or LOH (functional, $p53^{+/-}$) or the combination of both in 76 % (54/71) of the analyzed tumors. Subsequently, 24 % of the tumors displayed two intact copies of the *p53* gene (functional $p53^{+/+}$) as defined by our approach. Neither of the groups characterized by different p53 functional status, in terms of mutation and/or LOH, predicted survival in the current material.

Recurrent DNA aberrations in relation to colorectal tumorigenesis and progression (Paper II and III)

In this section, results from both papers are presented together. There are two notable differences between the studies, namely hybridization conditions and data analysis strategies, which are described in detail in the Methodological Considerations section.

Summary of genome-wide DNA aberrations (Paper II and III)

The proportion of aberrant DNA increased with tumor progression, defined as early (Dukes A and B) versus late stage tumors (Dukes C and D) (Figure 4d). Accordingly, the number of altered chromosomes also increased with tumor progression. Dukes C showed the highest figures considering the proportion of aberrant DNA as well as the number of altered chromosomes (Figure 4d). The most frequent aberrations identified in colorectal tumors and independent of tumor stage were gain on chromosomes 7, 13 and 20 and loss of 18, although quantitative DNA alterations were identified in each chromosome. Aberrations that appeared to discriminate between early and late colorectal tumors and thus recurrently and exclusively observed in Dukes C and/or Dukes D, were loss of 8p and gain of 7q, 8q and 13q.

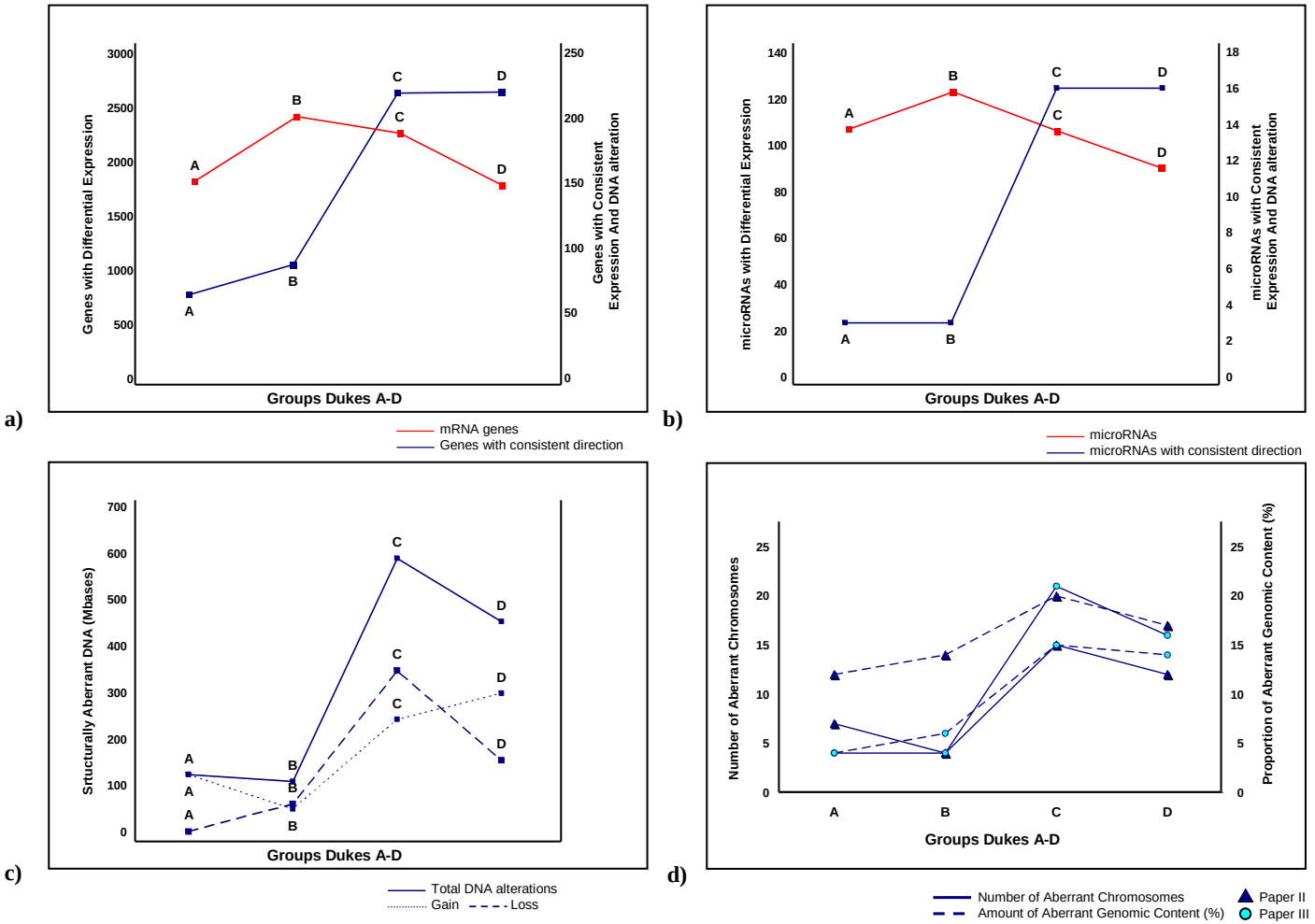


Fig 4. Colorectal tumor progression visualized by individual or combinations of DNA aberrations and mRNA or microRNA expression. Figure a-d summarize Dukes A, B, C and D specific: a) mRNA gene distribution (red line) and genes with combined DNA and mRNA alterations (blue line), b) microRNA distribution (red line) and genes with combined DNA and microRNA alterations (blue line), c) DNA alteration in terms of total number of aberrant DNA sequences (blue line), lost sequences (dashed line) and gained sequences (semidashed line), d) genetic DNA aberrations presented as the proportion of aberrations per whole genome (line) and as the number of aberrant chromosomes (dashed line). Figures from Paper II and III are indicated by blue triangles and turqoise circles, respectively.

Genome-wide DNA aberrations (Paper II)

DNA aberrations in Dukes A-D

7 %, 4 %, 15 % and 12 % of BAC-clones representing autosomal chromosomes were altered in Dukes A, B, C and D tumors, respectively and the number of altered chromosomes was 12, 14, 20 and 17. Equal amount of loss was observed in chromosome 18 in both Dukes A and B and C and D tumors, while tumor DNA from Dukes C and/or D displayed additional gains in chromosome 7, 8, 19 and 20 compared to Dukes A and B.

DNA aberrations in late stage tumors (Dukes C and D)

Major losses of the short chromosomal arm 8p and a region on 21q were observed in Dukes D but not in Dukes A, B or C. Gains specific for late stage tumors were parts of both 7p and 7q, the majority of chromosome 8q as well as parts of 19p and 20p (Table 4).

DNA aberrations in tumors from long and short-term survivors

Major alterations were found on chromosomes 8, 13, 18 and 20, but DNA from tumors derived from long-term (>10 years survival after CRC operation) patients only displayed large-scale aberrations on chromosomes 13 and 20 (Fig. 2 in Paper II). In tumor DNA from short-term survivors (< 1 year survival after CRC operation), chromosomes 8q and 13q showed major regions with gain, while losses were observed at chromosomes 8p, 18 and 21 (Fig. 2 in Paper II) (data not shown). When tumor DNA from long- and short-term survivors were hybridized in competition to each other, three chromosomes displayed major DNA aberrations. Gained regions were observed on chromosome 8, 9, 13 and a major loss on chromosome 8 (Table 4).

Genome-wide DNA aberrations (Paper III)

Tumor specific DNA alterations increase with progression, defined as early (Dukes A and B) versus late tumors (Dukes C and D) (Figure 4c and Fig. 2 in Paper II). Dukes A, B C and D tumors displayed DNA alterations in 4, 4, 21 and 16 percent of the genome when compared to normal colon tissue (Table 2 in Paper III) where 4 chromosomes were altered in Dukes A, 6 in Dukes B, 15 in Dukes C and 14 in Dukes D (Table 2 in Paper III). Distributions of genes that map to these regions are summarized in Figure 3.

Minimal Common Regions (MCR) in Dukes A, B, C and D

In Dukes A, B, C and D groups, chromosomes 1-11, 13-18 and 20-21 showed 102 Minimal Common Regions (MCRs); 78 % (80) were gained and 22% (22) were lost regions (Supplementary List 1). Overall, the aberrations correspond to 30 % of the genome (X and Y chromosomes excluded). 14 % of the aberrant DNA bases covered by MCR regions were altered in at least 3 out of 4 Dukes groups (ABCD, ABC, ACD or BCD) and located to chromosomes 7p, 13q, 18 and 20. Chromosomes 20 (41 Mb) and 13 (1 Mb) showed aberrations (gain) in all Dukes A, B, C and D stages. All combinations of MCRs that include either Dukes A, B or both, represented 55% of the MCRs (56) and were not considered relevant for tumor progression.

Table 4. Chromosomal regions with DNA aberrations (> 2 Mb) in tumor DNA in relation to survival or tumor progression defined by Dukes staging of colorectal cancer. The indicated cytobands refer to aberration start points. Regions in **bold** are aberrant in late stage tumors.

	Chromosome									
Hybridization	**7**	**8**	**9**	**13**	**14**	**17**	**18**	**19**	**20**	**21**
Paper II: Long vs Short Survival										
Gain	-	8p11.21 (6Mb) **8q11.21 (16Mb)** **8q12.3 (18 Mb)** **8q21.3 (34 Mb)** **8q24.13 (22Mb)**	**9q33.2 (17Mb)**	13q12.2 (9Mb)	-	-	-	-	-	-
Loss	-	**8p21.3 (13 Mb)**	-		-	-	-	-	-	-
Paper II : Dukes A+B										
Gain	7q36.1 (11 Mb)	8p11.21 (6 Mb)		13q12.11 (33 Mb) 13q21.33 (8 Mb) 13q31.3 (22 Mb)				19p13.3 (2 Mb)	20q11.1 (34 Mb)	-
Loss	-	-	-	-	-	-	18p11.32 (8 Mb) 18q11.2 (21 Mb) 18q21.2 (21 Mb)	-	-	-
Dukes C+D										
Gain	**7p22.3 (5 Mb)** **7p15.3 (33 Mb)** 7q36.1 (11 Mb)	8p11.21 (6 Mb)b **8q22.1 (4 Mb)** **8q23.3 (28 Mb)**	-	13q12.11 (33 Mb) 13q21.33 (8 Mb) 13q31.3 (22Mb)	-	-	-	19p13.3 (2 Mb) 19p13.3 (17 Mb)	20p13 (7 Mb) 20p12.1 (7.4 Mb) 20q11.1 (34 Mb)	-
Loss	-	**8p23.3 (36 Mb)a**	-	-	-	-	18p11.32 (8 Mb) 18q11.2 (21 Mb) 18q21.2 (21 Mb)	-	-	**21q21.1 (5 Mb)a**
Paper III: Dukes A+B										
Gain	-	-	-	13q12.11 (4 Mb) 13q12.13 (3 Mb) 13q21.1 (2 Mb)	-	-	-	-	20p13 (62 Mb)	-
Loss	-	-	-	-	-	-	18q11.2 (54 Mb)	-	-	-
Dukes C+D										
Gain	**7p11.1(20Mb)a,b**	**8q11.1 (91 Mb)**	**9q34.3 (3 Mb)a**	13q12.11 (4 Mb) 13q12.12 (3 Mb) 13q12.13 (3 Mb) **13q12.3 (29 Mb)c** 13q21.1 (2 Mb) **13q21.31 (11 Mb)c** **13q22.1 (40 Mb)c**	-	-	-	-	20p13 (62 Mb)	-
Loss		**8p23.2 (33 Mb)a**	-	-	**14q12 (3 Mb)a**	**17p13.3 (22 Mb)a**	18q11.2 (54 Mb) **18p11.32 (17Mb)b,c**	-	-	**21q21.1 (14 Mb)** **21q11.22 (7 Mb)a**

a) The region was only indicated as aberrant in Dukes D, b) the aberration starts on the p-arm, spans over the centromere and covers parts of the q-arm. c) a part of the aberrant DNA region is only observed in late Duke stages.

Minimal Common Regions (MCRs) in late stage tumors (Dukes C and D)
Overall, 75 % of the MCRs were found in Dukes C and/or D: 23 % of aberrant bases covered by MCR regions were only present in Dukes C and D, 39 % only in Dukes C and 13 % of the aberrant bases were only present in Dukes D. MCRs in Dukes C and D were located to chromosomes 8q, 13q, 18p and 21q and MCRs in Dukes D only included 1p, 7q, 8p, 9q, 11q, 13q, 14q, 15q, 17p and 21q (Supplementary List 1) (Table 4).

mRNA and microRNA expression in colorectal tumorigenesis and progression (Paper III)

Summary of genome-wide mRNA changes between tumor and normal colon tissue
There was no clear relation between the number of differentially expressed genes and tumor progression (Figure 4a) where 6, 8, 8 and 6 percent of the genes showed differential expression (FC>1, FDR<0.5) in tumor tissue compared to normal colon tissue in Dukes A, B, C and D. The gene distribution between Dukes groups is summarized in Figure 5 (details in Table 3 in Paper III). In all Dukes groups, the number of differentially expressed genes was consistent between the different chromosomes, except for chromosome 13 where Dukes D showed more than twice as many up-regulated genes compared to Dukes A-C ($p=10^{-11}$, Fisher's exact test) (Table 3 in Paper III).

Summary of genome-wide microRNA changes between tumor and normal colon tissue
There was no relation between tumor stage and the number of regulated microRNAs (Figure 4b). Dukes A, B, C and D tumors showed 17, 21, 18 and 15 percent differential expression (FC>0.5, FDR<0.05) compared to normal colon tissue. Dukes B and C showed microRNAs with differential expression in all chromosomes. Dukes A showed no such expression in chromosome 10 and Dukes D lacked differentially expressed microRNAs in chromosomes 10 and 12 (details in Table 4 in Paper III). 173 microRNAs showed differential expression in one or combinations of Dukes groups and 55 were regulated in all Dukes groups (Figure 5).

microRNA expression in relation to late Dukes stages (C and D)
Six microRNAs showed significant change in expression levels between Dukes A and B and Dukes C and D stages (Table 5 in Paper III). Two were altered in Dukes D only and 4 in Dukes A and B but not in C and D.

Combination of DNA aberrations and mRNA/microRNA expression in relation to colorectal tumorigenesis and progression (Paper III)

Genome-wide interaction of DNA alteration and mRNA expression
Each Dukes group showed genome-wide correlation between genomic DNA aberrations and transcriptional differential expression (Supplementary List 2). 31% (6498/21261) of human genes were altered in any of Dukes A, B, C or D stages at either the DNA or RNA level (or both, Figure 5). 1231 of these genes (19 %, 1231/6498) were altered in all Dukes stages and

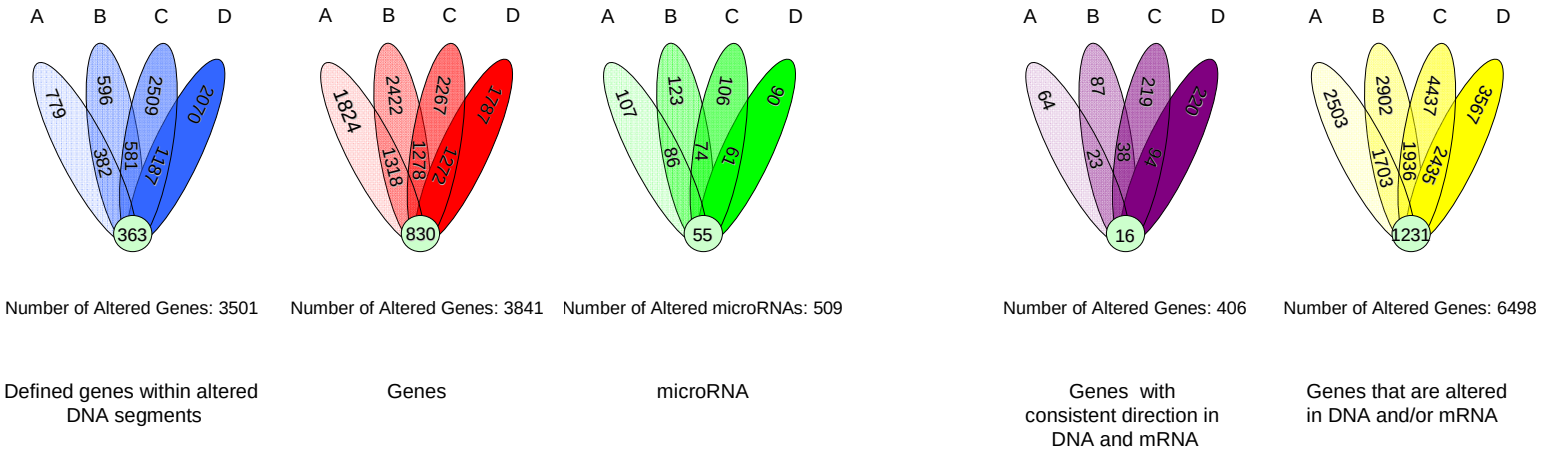


Figure 5. Gene distribution in Dukes A, B, C and D groups in five subsets dependent of analysis; DNA (blue), mRNA (red), microRNA (green), combined DNA and mRNA with consistent direction (purple) and total number of combined DNA and mRNA genes (yellow). The number of genes or microRNAs that are common in Dukes A, B, C and D are presented in the light green circles (◯) at the base of each "Vennflower" and those that are common between A and B, B and C, C and D are displayed between overlaid petals. Observe that the values indicated by each individual leave refer to the total number of genes or microRNAs for that Dukes stage.

406 genes (6 %, 406/6498) showed alterations with consistent direction, i.e. up-regulation in regions with gain or down-regulation in regions with loss (Figure 5).

Chromosomal interactions of DNA alterations and mRNA expression

Chromosomes 4, 8, 13 and 20 displayed interactions between DNA alterations and differentially expressed genes when whole chromosomes were tested separately (Supplementary List 2). The number of chromosomes with significant interaction increased with tumor progression according to Dukes staging system; Dukes A showed the least number of interactions (1) and Dukes D the most (4). Chromosome 20 displayed interactions in Dukes A, B and D ($p<0.05$, Fishers exact test) and a trend in Dukes C ($p<0.1$) while chromosome 8 showed interactions in Dukes C and D ($p<0.05$, Fishers exact test) (Fig. 2 in Paper III). Chromosome 4 showed interactions in Dukes C and chromosome 13 in Dukes A and D ($p<0.05$) (Supplementary List 2).

Segment interactions of DNA alterations and mRNA expression

The number of DNA segments with significant interaction to mRNA differential expression increased with progression, according to the Dukes staging system (Figure 6). Dukes A comprised 3 segments (66 Mb), Dukes B 3 (23 Mb), Dukes C 5 (358 Mb) and Dukes D 7 segments (244 Mb) with interaction. Eleven of the segments were present in at least two of the four Dukes groups. These segments were located to chromosomes 7, 8, 13, 18 and 20.

Segment interactions of DNA alterations and mRNA expression in Dukes C and D or only Dukes D

Three DNA segments with significant interaction to mRNA differential expression were identified only in Dukes C and D; gain of 8q22.1-22.3 (10 Mb), gain of 8q23.3-24.23 (25 Mb) and loss of 18q12.3-21.1 (5.4 Mb). Five regions were found only in Dukes D; gain of 7p11.1-q21.11 (20 Mb), loss of 8p23.2-23.1 (3.5 Mb), loss of 8p21.2-12 (4 Mb), gain of 13q12.3-q21.2 (29 Mb) and gain of 13q22.1-q34 (41 Mb) (Supplementary List 3).

Genes of potential importance for carcinogenesis

Sixteen up-regulated genes within segments with significant interaction were found in all Dukes groups and were all located to chromosome 20 (Supplementary List 4). The segment covers 40 Mb on chromosome 20p11.21-20q13.33 (66% of the entire chromosome 20).

Genes of potential importance for tumor progression

The DNA segments with significant interaction to mRNA expression contained 41 genes that discriminated between early and late tumor stages such that they were differentially expressed in either Dukes A and B or Dukes C and D. Sixteen genes were regulated in only Dukes C and D, 21 genes in only Dukes D and 5 genes were regulated in Dukes AB or ABC but not in D. A full 90% (36) of these genes showed a consistent mRNA expression in relation to DNA alteration, i.e. up-regulation in regions with gain or down-regulation in regions with loss. Four of these genes were located to chromosome 7p (regulated in Dukes D), three at 7q (Dukes C

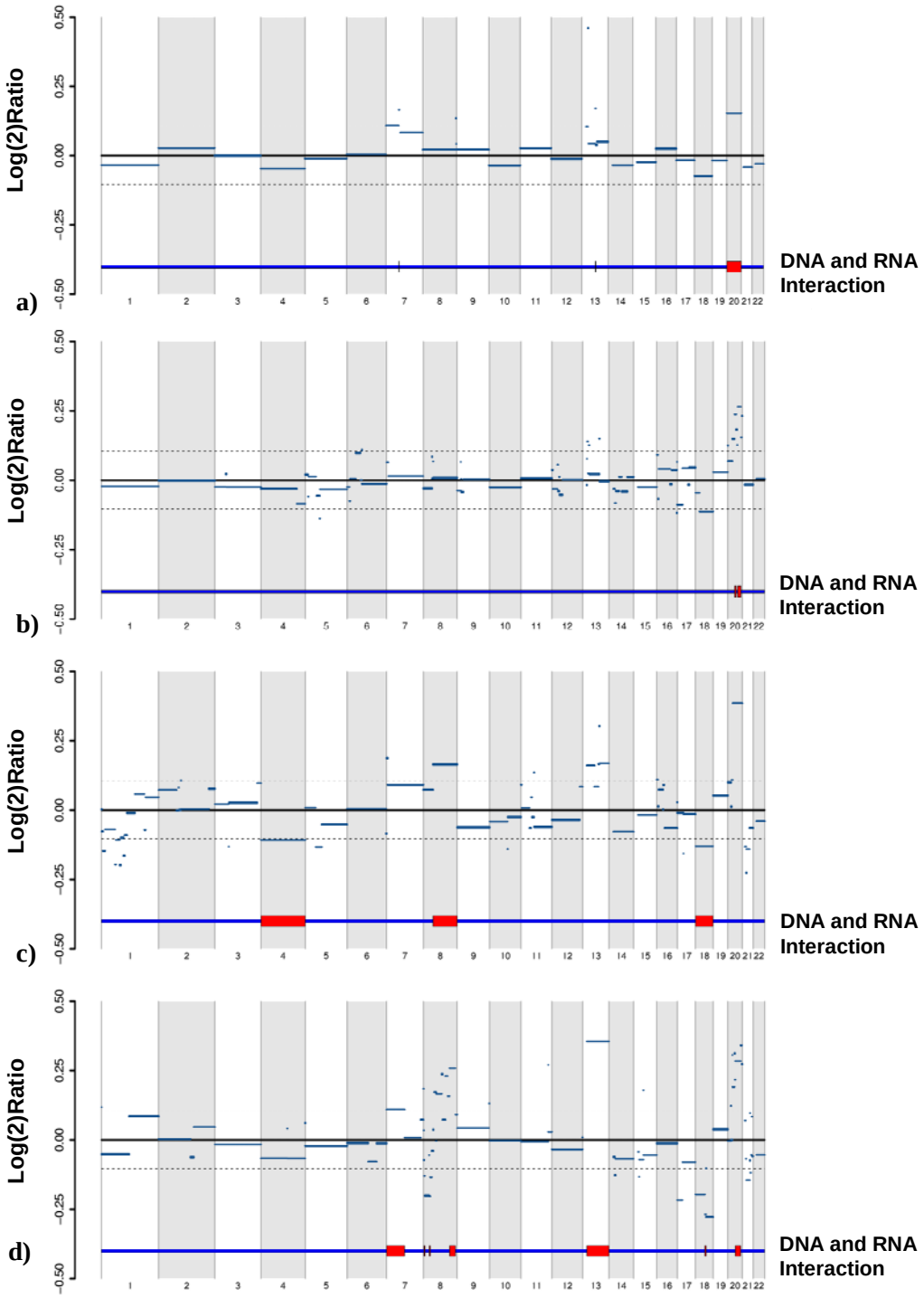


Figure 6. Genome-wide DNA aberrations in Dukes A, B, C and D. Thresholds for aberrant DNA segments with gain and loss are indicated by dashed lines and segments with significant interaction between DNA and differential expression are highlighted in red at the base of each plot.

and D), five at 8p (Dukes D), six at 8q (Dukes C and D), 19 at 13q (14 Dukes C and D, 5 Dukes D), and five at 18q (2 Dukes C and D, 3 Dukes D) (Table 5).

Interaction between DNA alternations and microRNA expression

Twenty-three microRNAs were located to significantly altered DNA segments in Dukes A, B, C or D in chromosomes 1, 4, 7-9, 13, 17, 18 and 20. Three, 3, 16 and 16 microRNAs were altered in Dukes A, B, C and D respectively. One microRNA (miR-663 at 20p11.1) was upregulated within a gained region in all Dukes stages. All microRNAs in segments with significant interaction in Dukes C and D showed differential expression in Dukes A and B, only that these microRNAs were located in regions without altered DNA segments in Dukes A and B. No regions with interaction between microRNAs and DNA alterations were found to discriminate between early and late Dukes tumor stages.

Structural DNA variation in normal colon mucosa from colorectal cancer patients (Paper II and IV)

In this section, results from Paper II and IV are presented together. There are two notable differences between the studies, namely experimental setup and data analysis strategies which are described in detail in the Methodological Considerations section.

Structural variation in normal colon DNA, in summary (Paper II and IV)

Nine verified CNV regions were common between the total amount of 205 (5 %) CNVs found by oligo microarrays (Paper IV) and 28 (36 %) CNVs observed by tiling BAC arrays (Paper II). All overlapping CNVs were verified CNV regions by comparison to the Database of Genomic Variants (DGV).

Structural Variation in normal colon DNA from long and short-term survivors (Paper II)

Twenty-eight verified and 2 novel CNV loci were identified in normal colon DNA from long and short-term survivors (Table 3 in Paper II). The 2 novel CNV loci identified at the time of analysis (2006) were, however, confirmed CNV loci according to the current database information (DGV updated March 11, 2009, Build 36 (Mar. 2006)).

Structural Variation in normal colon DNA (Paper IV)

Genome-wide analyses of DNA from 12 normal colon mucosa and one blood sample specimen from patients with colorectal tumors, revealed a total number of 774 potential CNV loci calls. 205 calls were identified in at least two specimens out of which 118 (57%) calls were verified CNVs, according to the *Database of Genomic Variants* (DGV) (Iafrate, Feuk et al. 2004; Zhang, Feuk et al. 2006) and 87 were novel at the time of analysis. Eight potential CNVs were found in more than 50% and 5 in more than 75% of the specimens (Table 1A in Paper IV). 67% (58) of the CNVs covered regions that harbor at least one gene. Three novel variations, present in 3 out of 12 groups, potentially discriminate between early (Dukes A and B) and late (Dukes C and D) stage. One CNV, identified at 10q21.1, was present in late

Table 5. 41 genes located within altered DNA segments in late CRC tumors (Dukes C and/or D). Genes showed differential expression in either Dukes A, A-C, Dukes C and D or only Dukes D. "Alteration DNA/RNA" refers to alterations in late Dukes stages.

	Genes	Location	Alteration DNA/RNA	Protein Function
Dukes A				
	ITM2B	13q14.2	↑ -	Membrane
	A_32_P80587	13q13.1	↑ -	Unknown Function
Dukes A - C				
	GTF2E2	8p12	↓ -	Transcription Factor -Initiation
	ESCO2	8p21.1	↓ -	Acetyltransferase
	PBK	8p21.1	↓ -	Phosphorylation
Dukes C & D				
	STX1A[b]	7q11.23	↑ ↑	Transport
	CLDN4	7q11.23	↑ ↓	Membrane-Development
	CLDN3	7q11.23	↑ ↓	Membrane-Development
	NAT2 [a]	8p22	↓ ↓	Associated to Cancer and Drug Toxicity
	BLK [a]	8p23.1	↓ ↓	Signal transduction in B-lymphoid Cells
	RPL7 [a]	8q21.11	↑ ↑	Ribosome
	RPL30	8q22.2	↑ ↑	Ribosome
	PABPC1	8q22.3	↑ ↑	Translation Initiation
	TATDN1	8q24.13	↑ ↑	Hepatocarcinoma
	FBXO32[b]	8q24.13	↑ ↓	Phosphorylation Dependent Ubiquitination
	EXOSC8	13q13.3	↑ ↑	RNA processing
	C13orf7	13q31.1	↑ ↑	Unknown Function
	RANBP5	13q32.2	↑ ↑	Transport
	TPP2	13q33.1	↑ ↑	Proteolys
	SLC14A1-002	18q12.3	↑ ↑	Transport
Dukes D				
	IGFBP3 [a]	7p13	↑ ↑	Growth-IGF
	URG4 [a]	7p13	↑ ↓	Hepatocarcinoma
	Hs648110 [a]	7p15.2	↑ ↑	Unknown Function (Ribosome)
	C7orf46 [a]	7p15.3	↑ ↑	Unknown Function
	WDR67	8q24.13	↑ ↑	Unknown Function
	RFXAP	13q13.3	↑ ↑	Transcription Factor -Development
	ALG5	13q13.3	↑ ↑	Glycosylation
	NHLRC3	13q13.3	↑ ↑	Unknown Function
	KIAA1704	13q14.12	↑ ↑	Unknown Function
	CAB39L	13q14.2	↑ ↑	Unknown Function
	THSD1	13q14.3	↑ ↑	Extracellular Matrix
	AL831999	13q14.3	↑ ↑	Unknown Function
	SPRY2	13q31.1	↑ ↑	Signalling
	TGDS	13q32.1	↑ ↑	Unknown Function
	CLDN10	13q32.1	↑ ↑	Membrane-Tight Junction
	SLC10A2	13q33.1	↑ ↓	Transport -sodium/bile acid
	ANKRD10	13q34	↑ ↑	Unknown Function
	PCID2	13q34	↑ ↑	Unknown Function
	AF263545	18q12.3	↑ ↑	Unknown Function
	CD33L3	18q12.3	↑ ↑	Membrane
	ATP5A1	18q21.1	↓ ↓	Transport

Dukes A genes were significantly downregulated only in Dukes A and expression increases with increased Dukes stage. Dukes A-C genes were significantly upregulated only in Dukes A-C, but not in Dukes D. [a] Genes are located to significantly abberant segments without significant segment interaction between DNA and RNA. [b] Genes that are target for microRNAs identified in Paper II, Table 5.

stages (2 Dukes C and 1 Dukes D) and two variations, at 15q21.3 (2 Dukes A and 1 Dukes B) and 2p15 (2 Dukes A and 1 Dukes B) was present in early stages, (Table 1B in Paper IV).

DISCUSSION

General Summary

In the current thesis, colorectal tumors and normal colon tissue were used as models to investigate the role and order of genetic events in relation to transformation from early to late tumor stage. Initially, we focused on a single gene; *p53*- by no doubt implicated in neoplastic development, to relate its genetic defects to progression and survival in colorectal cancer (CRC) patients. The results confirmed the importance of p53 in tumorigenesis, but did not generate any clear evidence concerning progression or survival. Next, we broadened our spectra by applying genome-wide array-based CGH analysis to a number of colorectal tumors subdivided according to Dukes stage or long or short-term survival. We observed a large number of DNA aberrations related to carcinogenesis and progression that were in agreement with previous and simultaneous reports. However, due to the complexity of the identified aberrations, it was difficult to distinguish primary from secondary events in relation to colorectal tumorigenesis. The general DNA aberration patterns were therefore confirmed by a second genome-wide arrayCGH study and further combined with matching mRNA and microRNA profiles. Thereby, DNA aberrations without any correlated alterations in mRNA and microRNA were excluded. The combined aberration patterns generated 41 genes and 6 microRNAs with potential importance for colorectal progression. Finally, we investigated the occurrence of known and novel structural variations, CNVs, in DNA from normal colon mucosa from a cohort of CRC patients. Several known and novel CNVs appeared frequently in CRC patients but not in controls and were thus potentially associated to colorectal tumorigenesis.

Large-scale techniques such as arrayCGH enable simultaneous identification and analysis of unbalanced structural DNA aberrations although breakpoints, translocations and inversions are difficult to determine due to lack of resolution. Nevertheless, the development of microarray-based techniques has enhanced the collection of genetic data in the search for diagnostic and predictive markers associated to cancer. During the first years following commercialization of expression arrays in the beginning of this century, huge amounts of microarray based studies were published in the field of cancer. However, due to various ways to collect tissue biopsies, different RNA extraction protocols and diverse approaches for analysis of already complex specimens, interpretation and in particular comparisons between studies were initially close to impossible. Subsequently, rules for “Minimal Information About a Microarray Experiment” (MIAME) was released (Brazma, Hingamp et al. 2001) and in addition, implementation of standards for microarray handling and analysis convey a more solid and convenient technique. Although improvements have increased the ability to generate trustworthy data, consensus genes with differential expression in colorectal cancer are low. Many different genes and panels of genes have been presented as of potential importance for colorectal carcinogenesis or progression (Cardoso, Boer et al. 2007; Nannini, Pantaleo et al. 2009), but few if any have been proved reproducible enough for clinical applications.

One way to circumvent this problem was to evaluate expression data in relation to coordinated aberrations in DNA content and exclude differential expression without such correlations. However, combined analyses of this kind must be interpreted with caution, since functional consequences of DNA alterations are not always apparent. DNA alterations do not necessarily induce subsequent and intuitively expected changes in expression, as reported by others (Platzer, Upender et al. 2002; Tsafrir, Bacolod et al. 2006).

Which chromosomal aberrations distinguish early from late colorectal tumors?

Genomic instability reflects the accumulation of genetic alterations that are frequently identified in colorectal tumors and provides growth advantages leading to clonal expansion. There is a wide agreement that recurrent genetic events most likely encompass genes of importance that may also drive the neoplastic transformation from normal epithelial cells to malignant tumors (Vogelstein and Kinzler 2004; Tsafrir, Bacolod et al. 2006). In contrary, others claim that genomic instability rather is a bystander effect that progress along with tumorigenesis as a result of malfunctional cell cycle activities (Grady 2004; Tsafrir, Bacolod et al. 2006). Still, the origins of chromosomal abnormalities are elusive and a matter of debate.

However, tumors with chromosomal instability (CIN) encompass diverse genetic changes and show substantial intratumor variability considering the patterns of genetic rearrangement. Since a subset of these changes is most often shared by all neoplastic cells within a tumor, a step-wise accumulation of genetic abnormalities is likely to appear during tumor growth (Heim, Mandahl et al. 1988). This is evident, since the amount of genetic alterations in terms of entire chromosomes as well as subchromosomal aberrant lesions is recurrent and apparently increases with tumor progression (Ried, Heselmeyer-Haddad et al. 1999; Hermsen, Postma et al. 2002).

In Paper II and III, genome-wide DNA alterations were characterized and correlated to colorectal tumor progression defined by Dukes A-D staging system (Lagerstedt, Staaf et al. 2007; Lagerstedt, Kristiansson et al. 2009). The DNA patterns in general, as well as in both early and late tumor stages (defined as Dukes A and B versus Dukes C and D) were in agreement with previously published data (Ried, Heselmeyer-Haddad et al. 1999; Hermsen, Postma et al. 2002; Douglas, Fiegler et al. 2004; Nakao, Mehta et al. 2004; Jones, Douglas et al. 2005; Mehta, Nakao et al. 2005). Frequent early stage DNA changes include gain in parts of chromosomes 7p and 13q and whole chromosome 20 and loss in parts of chromosome 18q, Late stage alterations include gain of 7p, 7q, 8q, 13q and loss of 8p, 18p and 21q. Furthermore, our observations also indicate a to our knowledge novel association between late stage and poor prognosis tumors and loss of a 5Mb DNA segment on 21q21.1-2. Further, we have also found that the loss of 8p clearly and exclusively appear in Dukes D tumors as well as in tumors from patients with short-term survival.

In relation to Dukes A, B and D, Dukes C tumors display higher proportions of aberrant DNA distributed in an increased number of chromosomes (Figure 4d). We identified this particular pattern in two independent cohorts of patients by arrayCGH analyses on two different platforms (BAC and oligonucleotide) based on diverse principles. Dukes D tumors involved less aberrant DNA than Dukes C, which may indicate that Dukes D tumor clones contain "just right" genetic defects (Fodde, Smits et al. 2001) that convey the ability to give rise to distant metastases. Hence, the rate of aberrant genetic events should be high enough to hit necessary target genes, but still moderate enough to avoid cell death caused by excess genomic instability (Cahill, Kinzler et al. 1999). The exclusive properties of Dukes D clones, bearing obtained advantageous abilities, may surpass less adapted clones and thereby creating a more homogenous genetic composition within the tumor.

Dukes D patients are diagnosed at an earlier age than the other patients in the present cohort (10 years earlier) which may be explained by two separate scenarios. Either a number of Dukes D tumors results from an earlier onset or alternatively, an increased development rate that does not follow the proposed stepwise progression model (Fearon and Vogelstein 1990). The latter scenario has actually been proposed by others although not specified as a phenomenon typical for Dukes D tumors (Weinberg 2007). The earlier onset may involve familial genetic events and consequently predispose for malignant disease by specific mutations or structural variation present in each and every cell. There are similarities to FAP and HNPCC hereditary variants, since these patients are diagnosed 10 or 20 years earlier than patients with sporadic disease. In case of the inherited or familiar scenario, the Dukes D genotype would thus contain genetic events predisposing for CRC and just await the right external tumor-favoring stimuli to trigger progression into malignant transformation.

A central and apparent observation in analyses based on genomic DNA alterations is the increased properties of aberrant DNA in relation to early to late tumor progression (Ried, Heselmeyer-Haddad et al. 1999). Interestingly, this particular trend was also illustrated by combination of aberrant genetic events (Figure 4a-c). In Paper III, data from CGH microarrays were combined with data from mRNA and microRNA expression microarrays to identify tumor specific genetic differences in relation to colorectal tumor progression. DNA and RNA interactions significantly increased from early to late Dukes stages at the chromosomal, subchromosomal and gene level (Figure 4a and Supplementary List 2 and 3) and the interaction between DNA and microRNA significantly increased at the gene level (Figure 4b). Dukes C and D together showed an increased number of aberrant DNA segments compared to Dukes A and B. Interestingly the mRNA and microRNA expression patterns were not concordant with the genomic DNA alterations since the number of differentially expressed mRNA and microRNA transcripts decreased. This trend was negative in relation to tumor progression (illustrated in Figure 4a-b) which was in contrast to the positive trend generated from interaction of DNA segments and mRNA/microRNA expression. Dukes D showed the least number of regulated genes compared to all the other Dukes stages, including Dukes A. Thus, the total number of chromosomal gains represents the only variable that alone constitutes a genome-wide predictor of progression (Figure 4c). Nevertheless, the number of

combined Dukes D DNA and RNA/microRNA genes represents a higher or equal amount to that of Dukes C (Figure 4a-c). This suggests that despite lower amounts of possible RNAs or microRNAs, their combined effect illustrates a tumor with highly successful and devastating properties. This is in line with the hypothesis that some tumors acquire highly favorable changes and hence the capacity to metastasize to distantly located organs as discussed previously.

Interactions in all Dukes A-D stages were only found in chromosome 20, which makes this region as a whole most likely related to colorectal carcinogenesis (Habermann, Paulsen et al. 2007), Alterations in chromosome 20 have previously been studied thoroughly by others and correlation between gains and up-regulation has been observed (Carvalho, Postma et al. 2009). Moreover, gains of large parts of chromosome 20 are present already in adenomas (Hermsen, Postma et al. 2002). In conclusion, the amount of aberrant DNA content as well as the number of correlated alterations in DNA and mRNA/microRNA, increases with progression from early to late tumor stage according to Dukes classification. Further, our results suggest that Dukes D comprise tumors with a high degree of aberrant DNA that to some extent affect expression. Dukes D tumors display lower amounts of regulated mRNAs and microRNAs, which may reflect that only selective, highly essential pathways are activated.

Does combined p53 mutation and loss of heterozygocity influence colorectal progression and survival?

In Paper I, combined allelic loss and *p53* mutation were considered equal to functional inactivation of the protein according to the "two hit-hypothesis" and as previously suggested considering p53 inactivation in CRC (Baker, Fearon et al. 1989; Forslund, Lonnroth et al. 2001).

Since patients with LOH downstream of the *p53* gene showed decreased survival compared to patients with retained heterozygocity at this locus ($p<0.05$, log-rank test) this marker (D17S720) was the only variable that displayed difference predictive of outcome. Hence, ablated p53 function, according to our definition, was not associated to survival or any particular stage during the progression from early to late colorectal cancer.

Nevertheless, 76 % of the analyzed tumors showed LOH or mutation in one or both alleles which confirms that p53 gene alterations are common events in colorectal tumors. Accordingly, structural p53 alterations probably occur at an early stage during neoplastic cell transformation from late adenoma to early carcinoma, as suggested in the Vogelstein model (Fearon and Vogelstein 1990). However, at the time of analysis, it was only recently reported that MDM2 was the main regulator of p53 activity and that some tumor suppressor genes, evidently even *p53*, were discovered to obtain haploinsufficient properties. Today, these genetic events are defined and characterized to a much greater extent. First, overexpression of *MDM2* due to gene amplification has been identified in almost 10% of CRCs (Momand, Jung

et al. 1998; Forslund, Zeng et al. 2008) though there is evidence that *MDM2* overexpression does not always correlate to amplification (Forslund, Zeng et al. 2008). Abnormal activation of *MDM2* can arise due to a polymorphism located in its promoter (Bond, Hu et al. 2004), which thereby attenuate p53 activity. Second, a single mutant p53 protein has the ability to disable the tetrameric structure that otherwise constitutes the functional p53 oligomer, which was proposed as a possible explanation for haploinsufficiency (Santarosa and Ashworth 2004). In conclusion, we found no association between tumor progression and functional p53 inactivation, although *p53* alterations were frequently observed in CRC tumors. However, deletions that occur downstream of the *p53* gene may be indicative for increased disease specific mortality.

Which genetic events are of importance for colorectal tumor progression?

In Paper III, candidate regions from interaction analyses were identified at chromosomal arms 7p (Dukes D), 8p (Dukes D), 8q (Dukes C and D), 13q (Dukes D) and 18q (Dukes C and D) that may encompass genetic events with impact on tumor progression. Gene expression patterns for genes within the candidate regions were evaluated followed by selection of regulated genes with significant difference between Dukes A, B and C, D or A, B, C and D. Fourty-one genes were located in these regions that were regulated in a way that discriminate early from late or metastatic tumor stage (Table 5). We identified two kind of candidate genes; 36 genes that only were significantly regulated in late stages of the disease and thus potential biomarkers and five genes that were regulated in Dukes A and B (or A, B and C) but not in C and D. The latter may be important for biological reasons, but are not interesting for predictive purposes (as biomarkers), since they are not measurable in Dukes C or D. Moreover, five of the potential biomarker genes have a reversed relationship between DNA alteration and the corresponding regulation (i.e. gain and down-regulation or loss and up-regulation) which makes them less valid since they lack the structurally aberrant DNA basis. As described above, we found that differential gene expression present in early stage tumors sometimes vanish in the later stage probably as a consequence of the corresponding gain or loss that appear in these tumors. In regions with interaction between aberrant DNA segments and mRNA differential expression in late Dukes stage tumors (Dukes D), this situation occurred for five genes on chromosome 8p and 13q (Table 3). *ESCO2, GTF2E2* and *PBK* are located to chromosome 8p and even though they have not previously been related to CRC, they are all implicated in processes in close relation to cancer events such as sister chromatid cohesion (*ESCO2*), transcription (*GTF2E2*) and *p53* destabilization (*PBK*). Two genes on chromosome 13q were significantly down-regulated in Dukes A and expression gradually increases with increased tumor stage. Only one of the two genes at 13q (*ITM2B*) is known to encode a protein which is an integral membrane protein that may modulate cell survival (Lee, Jeong et al. 2007) (Table 3).

Almost half of the candidate genes identified in Paper III (19/41) were located at chromosome 13, which implies that regions with DNA and RNA interaction on chromosome 13 may be important for colorectal tumor progression. The composition of structural chromosome 13

alterations is complex and varies within different Dukes stages (Figure 6) (Gisselsson, Jonson et al. 2001). Furthermore, we found that Dukes D tumors showed significantly more regulated genes in chromosome 13 compared to the other Dukes stages, which represents a unique pattern in this data set since no other chromosome showed such differences in any group. In contrast, recurrent gain as well as interaction of DNA segments and mRNA differential expression in chromosome 20 was observed in all Dukes stages. Actually, sixteen genes located on chromosome 20 showed combined gain and upregulation in all Dukes A-D stages and no such combined events were found elsewhere in the genome. Among these genes were *AURKA* (Figure 7) and *CSE1L*, which were recently related to colorectal cancer (Bertucci, Salas et al. 2004; Camps, Grade et al. 2008). Loss of 18q is a frequent genetic event that is proposed to occur during progression from intermediate to late adenoma in the CRC progression model (Fearon and Vogelstein 1990). *SMAD7* (Figure 7), have been presented as a key target gene, since it maps to a CRC susceptibility locus at 18q21 (Broderick, Carvajal-Carmona et al. 2007; Pittman, Naranjo et al. 2009). In accordance, the expression patterns in Dukes A-D show a downregulation trend in the Paper III data set. Moreover, *PTP4A3* (Figure 7), also referred to as *PRL-3,* have been associated to metastatic CRC tumors and is located to a recurrently amplified region on 8q24 (Saha, Bardelli et al. 2001). In Paper II and III, this amplified region was observed in short-term survivors as well as late stage tumors and correspondingly, *PTP4A3* showed a trend versus increased expression although below the cutoff for significantly altered expression (data not shown).

The majority of the 36 candidate genes with alterations in Dukes C and D (Paper III) code for proteins implicated in transcriptional/translational processes, transport, membrane proteins and post-translational modification. Differential expression of *TATDN1* and *URG4* genes are related to hepatocellular carcinoma, which makes them interesting as the liver is the most common primary location for distant metastasis in CRC. *NAT2* is involved in the metabolism of carcinogens and may affect risk of colorectal cancer (Bell, Stephens et al. 1995; Hein 2002) (1995). Finally, *STX1A* (7q11.23) and *FBXO32* (8q24.13) were target genes for suggested candidate microRNAs miR-602 and miR-144 (miRBase Target Database, Wellcome Trust Sanger Institute; http://microrna.sanger.ac.uk/targets/v5/) (Griffiths-Jones, Saini et al. 2008).

Surprisingly, as few as 6 microRNAs (including miR-602) was altered in a way that discriminate between Dukes A and B and Dukes C and D. In fact, 95 % of the microRNAs with interaction in C and D were also regulated in Dukes A and B, compared to 75 % in mRNA. Hence, there are few differences in microRNA regulation between early and late Dukes stages in CRC tumors, which suggest that microRNA may constitute regulators whose absence rather than presence promote malignancy in colorectal tumors. Accordingly, colorectal as well as other tumors recurrently display down regulation of microRNA in relation to normal cells (Blenkiron and Miska 2007). Nevertheless, we identified two microRNAs, miR-602 and miR-373 (Figure 7), which possibly impacts distantly metastatic tumors and four microRNAs of potential biological importance (Table 5 in Paper III). MiR-373 was recently suggested as a promoter of metastasis in breast cancer cells (Huang,

Gumireddy et al. 2008; Negrini and Calin 2008) but not previously related to metastasis in colorectal tumors. Up-regulation of miR-21 has been correlated to poor outcome in colorectal cancer patients (Schetter, Leung et al. 2008), but we did not find any clear evidence for that in the present data set. However, novel microRNAs are still rapidly arising and the number of microRNAs have almost doubled compared to those included in our analyses (718 today compared to 474 in 2007), implying that data should be interpreted with caution.

In conclusion, 41 candidate genes and 6 microRNAs may be involved in the progression from early to late colorectal tumor stage and among these genes 36 represent candidate biomarkers. Moreover, *SMAD7* and *AURKA* may represent genes that are of importance for colorectal tumorigenesis. Taken together, changes at the structural DNA level as basis for selection of biomarkers represent a robust rationale for characterization of tumor progression since the expected change in regulation has a stable intercourse with the clinical process.

Can nonpolymorphic structural variation contribute to detection of familial colorectal cancer (Paper II and IV)?

In recent years it has become evident that the human genome is more complex than previously appreciated, due to the presence of genetic variation between humans in terms of polymorphisms and structural variation (Iafrate, Feuk et al. 2004). These differences convey the possibility to identify specific variation associated to disease and subsequently provide diagnostic markers. In colorectal tumors, more than ten susceptibility loci (described in the introduction part) have already been discovered by SNP-based array technology. However, the power of these analyses has been questioned, since the odds ratios obtained within these studies are low (ranging from 1.1-1.5) (Easton and Eeles 2008; Houlston, Webb et al. 2008; Cheah 2009). Hence, identification of additional risk loci is required for predictive purposes next to colonoscopic examination and available markers for hereditary CRC.

Since structural variation range from single nucleotides to several megabases and involves variable conformations/arrangements, it is essential to cross-examine both SNPs and all types of CNVs for exploration of phenotypic variation within tumors (Stranger, Forrest et al. 2007). In relation to SNPs, structural variation is likely to have even more impact on phenotypic variation and may thus serve to add further risk sites (Iafrate, Feuk et al. 2004; Redon, Ishikawa et al. 2006; Korbel, Urban et al. 2007; Stranger, Forrest et al. 2007). It is therefore desirable, not only to determine novel predictory biomarkers, but also nonpolymorphic, structurally variant markers aside from the existing SNP markers uncovered so far. Hence, sufficient numbers of diagnostic markers based on structural genetic aberrations and polymorphic sites would considerably benefit CRC patients in terms of early diagnosis and thereby improved survival.

In situ synthesised oligonucleotide microarray technology is an efficient tool for CNV detection (de Smith, Tsalenko et al. 2007). Moreover, oligonucleotide arrays provide genome-wide sequence coverage which can not be ministered by SNP-arrays, since SNPs are not

evenly distributed throughout the genome (Carter 2007). Accordingly, high resolution oligonucleotide microarrays were used in Paper IV, to perform genome-wide scans for common DNA aberrations in visibly normal colon mucosa from colorectal cancer (CRC) patients. 118 known and 87 novel potential copy number variations (CNV) were observed in DNA pools grouped according to Dukes A-D system. The relation between identified known and novel CNVs are in agreement with other oligobased microarray studies (de Smith, Tsalenko et al. 2007; Camps, Grade et al. 2008). Among these, 29 known and 8 novel CNVs were found in more than 50 % and 8 known and 5 novel CNVs were found in more than 75 % of the DNA pools, indicating that these potential CNVs are either frequent loci in Northern Europeans or CRC specific structural variants.

The novel CNVs included three novel potential CNV regions that may discriminate early (Dukes A and B) from late (Dukes C and D) tumor stage, albeit present in as few as 25-30% of the analyzed samples. 58 out of the 87 novel CNVs, contained at least one gene. With few exceptions, genes within regions of structural variation do not contain known tumor genes, according to the majority of the GWAS performed to date (Easton 08). However, *SMAD7*, involved in TGFβ mediated activities (Boulay, Mild et al. 2003; Broderick, Carvajal-Carmona et al. 2007) (18q21.1), constitute an exception. In our data *CHRNB4* (15q25.1) encodes a cholinergic receptor located within a region recently associated to lung cancer (Amos, Wu et al. 2008; Liu, Vikis et al. 2008) and represent the only tumor related gene present in CNVs in more than 50% of analyzed specimens. No genes were found in the regions of potentially novel CNVs that discriminate early from late tumor stage.

Nine verified CNV regions were common between the total amount of 205 (5 %) CNVs found by oligo microarrays and 28 (36 %) tiling BAC arrays, used in Paper II. The overlap is satisfying, considering the notable differences between analysis platforms and strategies. Interestingly, one verified 1.5 Mb locus at 5q13.2 was identified in 12 out of 13 analyzed specimens including Paper II and IV samples as well as in DNA from blood and thereby indicate association to colorectal tumors. In conclusion, several known and novel nonpolymorphic structural variants were identified in normal DNA from CRC patients and may be predictive of familial colorectal cancer incidence. Clearly, however, these findings need further validation.

CONCLUSIONS

Taken together, the results in terms of the genetic events presented within this thesis is summarized by their potential relevance within the different stages that defines the stepwise progression proposed to transform normal cells into malignant colorectal tumors (Figure 7). First, gain in parts of chromosome 20, encompassing cell-cycle regulating kinase *AURKA,* as well as *p53* alterations and differential expression of microRNAs, may be involved in the development from adenoma to carcinoma. Second, several genomic DNA alterations appear to rise during progression defined as early (Dukes A and B) to late (Dukes C and D), namely loss of 18q and gain of 8q that harbor *SMAD7* and *PTP4A3*, respectively. Third, distant metastatic potential may be associated to loss of 8p and increased expression of miR-373. Fourth, putative structural variants observed in normal colon mucosa may predispose for the onset of malignant transformation in familial sporadic CRC. Finally, there is a clear relationship between the increased properties of aberrant DNA content as well as the number of combined genetic events and tumor progression. Nevertheless, Dukes D tumors possibly develop in a way that does not fit into the stepwise progression model, as illustrated by earlier onset and less genetic aberrations.

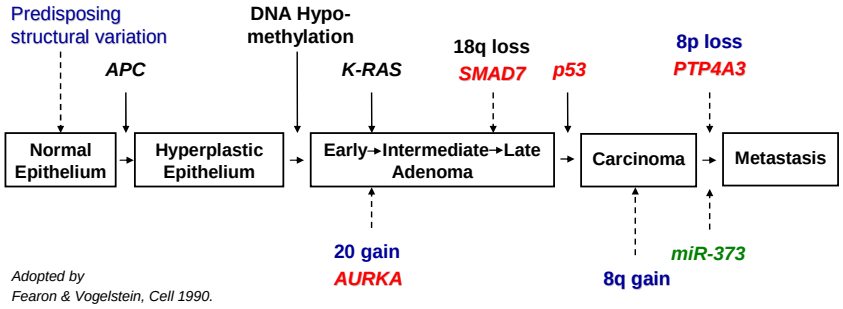


Figure 7. This modified "*Vogelgram*" is supplied with additional genetic events of putative importance for carcinogenesis or tumor progression, compared to the original model. Each event was observed in Paper I-IV and has also been presented by others, as described in the discussion part. Blue indicates genomic DNA aberrations, red refer to altered genes, either at the DNA, expression or both levels and green refer to microRNA transcripts with differential expression. Each added genetic event is also indicated by a dashed arrow.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all who have contributed to and supported me on my way to complete this thesis. Particularly, I would like to thank:

My supervisor **Fredrik Enlund,** first; by providing the opportunity to accomplish the completion of this thesis by your excellent guidance, great knowledge and never failing support, second; skilful help to navigate through the field of tumor biology, third; for your huge encouragement and inspiration through sometimes challenging situations. We did it!

Christina Lönnroth, my co-supervisor and dear friend, for sharing your deep knowledge with me, your great sense of humor, empathy and endless support. Thanks for many laughs and fruitful discussions and for being such a fantastic lab bench picnic partner!

Erik Kristiansson, my co-supervisor and idol ☺ for skilful guidance, endless knowledge about absolutely everything, always having a spare minute and for your unique ability to communicate science, in particular statistical issues. You make microarrays talk!

Kent Lundholm, for introducing me into the research field, being my tutor for many years and for financial support.

My colleagues and dear friends at Kir Metabollab; **Annika Gustafsson**, for sharing everything – from science and congress plans to daycare pickup schedules… You are a true friend! **Britt-Marie Iresjö** my roommate, constant adviser and friend - thanks for many great discussions, technical "BM and Stina" solutions and for always being there. ESC next year? **Marianne Andersson** for your great laboratory skills and knowledge as well as many joyful lab hours, included sometimes weird situations (!) followed by endless laughter! You're the best! **Lena Edberg** for helpfulness and friendship. **Anita F**, **Ann N**, **Annette**, **Christian**, **Hans**, **Lena E**, **Lena H**, **Lilian**, **Marita**, **Monica**, **Rigmor**, **Ulrika** and **Wenhua** for a nice atmosphere and stimulating company!

Anita Ohlsson, Ulla Körner, Agneta Sären and **Ingela Stave** for all help and for being so kind and supportive!

Dick Delbro for your help and support and your sense of humor! **Ann Forslund** for introducing me to p53.

Åke Borg and coworkers at SCIBLU, Lund; for nice collaboration and excellent guidance. **Johan Staaf**, for invaluable help and discussions concerning microarrays, and also **Johan Vallon-Christersson, Johan Enell** and **Göran Jönsson**.

Elisabeth Hansson and **Ulf Kressner** at Uddevalla Hospital for nice collaborations and clinical advice and **Ingela Apelman** for help with data handling. Lars Lindblom and Svante Nordgren for nice collaboration.

At Gastlab; **Sören Lundberg** for trustworthy and fast technical support and assistance, and **Emma Spak** for being supportive and for friendship.

Everyone at LLCR; especially **Carola, Ella, Gabriella, Marta, Ywonne** and **Yvonne** for sharing your knowledge and for always making me feel welcome!

Everyone at **Genanalys** for providing a nice and friendly athmospere!

Lina, my new friend, for kind help with editing my thesis.

My new roommates, especially **Haris** for PC assistance and many laughs and **Johan** for your kind concern.

Eva Mattsson, friend and tutor, and **Gunnar Eliasson**, my uncle, for care, good advice and very professional support.

All my friends in Göteborg, especially **Jenny**, my very best friend- you take such good care of me! For empathy, wisdom and many, many great times! And for sometimes being my mom ☺ **and Jörgen** for many stimulating discussions and several crazy events, **Susanne** for your encouragement and being such a good friend, **Matilda and Henrik** for relaxing "Sunday dinners" during weekdays, **Annika and Stefan**, **Birgitta and Björn.**

All my friends in Stockholm, especially **Jessica**, **Johanna** and **Viktoria** for long-lasting friendship and always believing in me.

Ulla and **Janne,** my parents-in-law, for unfailing care and support.

Aunt Anna and **Affe,** my old relatives, for your love and support!

Fredde, my brother, for your sympathy and great (best!) sense of humor and for always understanding my thoughts. My very dear sister-in-law, **Nadine**, for your compassion and big heart. For empathy and invaluable support during tough times.

Mamma Kerstin for so many things.. Your never failing love, support, concern, and for always being there for me, no matter what. For sharing your broad knowledge and for being a great grandma.

Mats, my husband, for endless love, your strength, care about our family and for being the best dad (and an excellent chef ☺). I love you!

Karin, my daughter and love and joy, you bring me constant happiness and make every day special!

Karins sibling, still in my belly, for constantly keeping me company, which I especially appreciate during late hours in front of the computer.

Hej, hej **pappa**!

This work was performed at the Department of Surgery, Institute of Clinical Sciences and has been supported by grants from the Swedish Cancer Society, the Swedish Research Council, Assar Gabrielsson Foundation (AB Volvo), Jubileumskliniken Foundation, IngaBritt & Arne Lundberg Research Foundation, Swedish and Gothenburg Medical Societies, the Medical Faculty, University of Gothenburg and Sahlgrenska University Hospital Foundation.

REFERENCES

ACS, A. C. S. (2007). Global Cancer Facts & Figures 2007, Estimated Number of New Cancer Saces by World Area.

Autio, R., S. Hautaniemi, et al. (2003). "CGH-Plotter: MATLAB toolbox for CGH-data analysis." Bioinformatics **19**(13): 1714-5.

Baker, S. J., E. R. Fearon, et al. (1989). "Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas." Science **244**(4901): 217-21.

Bell, D. A., E. A. Stephens, et al. (1995). "Polyadenylation polymorphism in the acetyltransferase 1 gene (NAT1) increases risk of colorectal cancer." Cancer Res **55**(16): 3537-42.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society **57**(1): 289-300.

Bertucci, F., S. Salas, et al. (2004). "Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters." Oncogene **23**(7): 1377-91.

Blenkiron, C. and E. A. Miska (2007). "miRNAs in cancer: approaches, aetiology, diagnostics and therapy." Hum Mol Genet **16 Spec No 1**: R106-13.

Boland, C. R., S. N. Thibodeau, et al. (1998). "A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer." Cancer Res **58**(22): 5248-57.

Bolstad, B. M., R. A. Irizarry, et al. (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." Bioinformatics **19**(2): 185-93.

Boman, B. M. and E. Huang (2008). "Human colon cancer stem cells: a new paradigm in gastrointestinal oncology." J Clin Oncol **26**(17): 2828-38.

Bond, G. L., W. Hu, et al. (2004). "A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans." Cell **119**(5): 591-602.

Boveri, T. (1914). Zur Frage der Entstehung Maligner Tumoren.

Brazma, A., P. Hingamp, et al. (2001). "Minimum information about a microarray experiment (MIAME)-toward standards for microarray data." Nat Genet **29**(4): 365-71.

Broderick, P., L. Carvajal-Carmona, et al. (2007). "A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk." Nat Genet **39**(11): 1315-7.

Cahill, D. P., K. W. Kinzler, et al. (1999). "Genetic instability and darwinian selection in tumours." Trends Cell Biol **9**(12): M57-60.

Cairns, J. (1975). "Mutation selection and the natural history of cancer." Nature **255**(5505): 197-200.

Camps, J., M. Grade, et al. (2008). "Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome." Cancer Res **68**(5): 1284-95.

Cancerfonden (2008). Cancerfondsrapporten 2008.

Cardoso, J., J. Boer, et al. (2007). "Expression and genomic profiling of colorectal cancer." Biochim Biophys Acta **1775**(1): 103-37.

Carter, N. P. (2007). "Methods and strategies for analyzing copy number variation using DNA microarrays." Nat Genet **39**(7 Suppl): S16-21.

Carvalho, B., C. Postma, et al. (2009). "Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression." Gut **58**(1): 79-89.

Cheah, P. Y. (2009). "Recent advances in colorectal cancer genetics and diagnostics." Crit Rev Oncol Hematol **69**(1): 45-55.

Douglas, E. J., H. Fiegler, et al. (2004). "Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas." Cancer Res **64**(14): 4817-25.

Dukes, C. E. and H. J. Bussey (1958). "The spread of rectal cancer and its effect on prognosis." Br J Cancer **12**(3): 309-20.

Easton, D. F. and R. A. Eeles (2008). "Genome-wide association studies in cancer." Hum Mol Genet **17**(R2): R109-15.

Epidemiologiskt Centrum, S. (2007). Cancer Incidence in Sweden 2006.

Fearon, E. R. and B. Vogelstein (1990). "A genetic model for colorectal tumorigenesis." Cell **61**(5): 759-67.

Fero, M. L., E. Randel, et al. (1998). "The murine gene p27Kip1 is haplo-insufficient for tumour suppression." Nature **396**(6707): 177-80.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-11.

Firestein, R., A. J. Bass, et al. (2008). "CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity." Nature **455**(7212): 547-51.

Fischer, S. G. and L. S. Lerman (1983). "DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory." Proc Natl Acad Sci U S A **80**(6): 1579-83.

Fodde, R., J. Kuipers, et al. (2001). "Mutations in the APC tumour suppressor gene cause chromosomal instability." Nat Cell Biol **3**(4): 433-8.

Fodde, R., R. Smits, et al. (2001). "APC, signal transduction and genetic instability in colorectal cancer." Nat Rev Cancer **1**(1): 55-67.

Forrester, K., C. Almoguera, et al. (1987). "Detection of high incidence of K-ras oncogenes during human colon tumorigenesis." Nature **327**(6120): 298-303.

Forslund, A., C. Lonnroth, et al. (2001). "Mutations and allelic loss of p53 in primary tumor DNA from potentially cured patients with colorectal carcinoma." J Clin Oncol **19**(11): 2829-36.

Forslund, A., Z. Zeng, et al. (2008). "MDM2 gene amplification is correlated to tumor progression but not to the presence of SNP309 or TP53 mutational status in primary colorectal cancers." Mol Cancer Res **6**(2): 205-11.

Foulds, L. (1954). "The experimental study of tumor progression: a review." Cancer Res **14**(5): 327-39.

Foulds, L. (1957). "Tumor progression." Cancer Res **17**(5): 355-6.

Frohling, S. and H. Dohner (2008). "Chromosomal abnormalities in cancer." N Engl J Med **359**(7): 722-34.

Ghadimi, B. M., M. Grade, et al. (2003). "Gain of chromosome 8q23-24 is a predictive marker for lymph node positivity in colorectal cancer." Clin Cancer Res **9**(5): 1808-14.

Gisselsson, D., T. Jonson, et al. (2001). "Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors." Proc Natl Acad Sci U S A **98**(22): 12683-8.

Goldgar, D. E., D. F. Easton, et al. (1994). "Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands." J Natl Cancer Inst **86**(21): 1600-8.

Goldstein, D. B., A. Ruiz Linares, et al. (1995). "An evaluation of genetic distances for use with microsatellite loci." Genetics **139**(1): 463-71.

Grady, W. M. (2004). "Genomic instability and colon cancer." Cancer Metastasis Rev **23**(1-2): 11-27.

Griffiths-Jones, S., H. K. Saini, et al. (2008). "miRBase: tools for microRNA genomics." Nucleic Acids Res **36**(Database issue): D154-8.

Habermann, J. K., U. Paulsen, et al. (2007). "Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis." Genes Chromosomes Cancer **46**(1): 10-26.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Harley, C. B. (2008). "Telomerase and cancer therapeutics." Nat Rev Cancer **8**(3): 167-79.

Harris, C. C. (1993). "p53: at the crossroads of molecular carcinogenesis and risk assessment." Science **262**(5142): 1980-1.

Hayflick, L. (1965). "The Limited in Vitro Lifetime of Human Diploid Cell Strains." Exp Cell Res **37**: 614-36.

He, T. C., A. B. Sparks, et al. (1998). "Identification of c-MYC as a target of the APC pathway." Science **281**(5382): 1509-12.

Heim, S., N. Mandahl, et al. (1988). "Genetic convergence and divergence in tumor progression." Cancer Res **48**(21): 5911-6.

Hein, D. W. (2002). "Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis." Mutat Res **506-507**: 65-77.

Herman, J. G. and S. B. Baylin (2003). "Gene silencing in cancer in association with promoter hypermethylation." N Engl J Med **349**(21): 2042-54.

Herman, J. G., A. Umar, et al. (1998). "Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma." Proc Natl Acad Sci U S A **95**(12): 6870-5.

Hermsen, M., C. Postma, et al. (2002). "Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability." Gastroenterology **123**(4): 1109-19.

Hinds, D. A., L. L. Stuve, et al. (2005). "Whole-genome patterns of common DNA variation in three human populations." Science **307**(5712): 1072-9.

Houlston, R. S., E. Webb, et al. (2008). "Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer." Nat Genet **40**(12): 1426-35.

Huang, Q., K. Gumireddy, et al. (2008). "The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis." Nat Cell Biol **10**(2): 202-10.

Iafrate, A. J., L. Feuk, et al. (2004). "Detection of large-scale variation in the human genome." Nat Genet **36**(9): 949-51.

Jaeger, E., E. Webb, et al. (2008). "Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk." Nat Genet **40**(1): 26-8.

Jones, A. M., E. J. Douglas, et al. (2005). "Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma." Oncogene **24**(1): 118-29.

Jones, P. A. and S. B. Baylin (2002). "The fundamental role of epigenetic events in cancer." Nat Rev Genet **3**(6): 415-28.

Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science **258**(5083): 818-21.

Knudson, A. G., Jr. (1971). "Mutation and cancer: statistical study of retinoblastoma." Proc Natl Acad Sci U S A **68**(4): 820-3.

Korbel, J. O., A. E. Urban, et al. (2007). "Paired-end mapping reveals extensive structural variation in the human genome." Science **318**(5849): 420-6.

Kristiansson, E. (2007). Statstical analysis of gene expression data. Department of Mathematical Sciences. Göteborg, University of Gotheburg**:** 39.

Kubbutat, M. H., S. N. Jones, et al. (1997). "Regulation of p53 stability by Mdm2." Nature **387**(6630): 299-303.

Lagerkvist, A., J. Stewart, et al. (1994). "Manifold sequencing: efficient processing of large sets of sequencing reactions." Proc Natl Acad Sci U S A **91**(6): 2245-9.

Lagerstedt, K. K., U. Kressner, et al. (2005). "The role of combined allelic imbalance and mutations of p53 in tumor progression and survival following surgery for colorectal carcinoma." Int J Oncol **27**(6): 1707-15.

Lagerstedt, K. K., E. Kristiansson, et al. (2009). Genes with relevance for early to late progression of colon carcinoma suggested from results in combined microarray analyses.

Lagerstedt, K. K., J. Staaf, et al. (2007) "Tumor Genome Wide DNA Alterations Assessed by Array CGH in Patients with Poor and Excellent Survival Following Operation for Colorectal Cancer." Cancer Informatics **Volume**, 341-355 DOI:

Lane, D. P. (1992). "Cancer. p53, guardian of the genome." Nature **358**(6381): 15-6.

Lee, S., J. Jeong, et al. (2007). "Forerunner genes contiguous to RB1 contribute to the development of in situ neoplasia." Proc Natl Acad Sci U S A **104**(34): 13732-7.

Lengauer, C., K. W. Kinzler, et al. (1997). "Genetic instability in colorectal cancers." Nature **386**(6625): 623-7.

Levine, A. J., W. Hu, et al. (2006). "The P53 pathway: what questions remain to be explored?" Cell Death Differ **13**(6): 1027-36.

Lichtenstein, P., N. V. Holm, et al. (2000). "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland." N Engl J Med **343**(2): 78-85.

Lipson, D., Y. Aumann, et al. (2006). "Efficient calculation of interval scores for DNA copy number data analysis." J Comput Biol **13**(2): 215-28.

Lynch, H. T. and A. de la Chapelle (2003). "Hereditary colorectal cancer." N Engl J Med **348**(10): 919-32.

McCarroll, S. A. and D. M. Altshuler (2007). "Copy-number variation and association studies of human disease." Nat Genet **39**(7 Suppl): S37-42.

McCarroll, S. A., T. N. Hadnott, et al. (2006). "Common deletion polymorphisms in the human genome." Nat Genet **38**(1): 86-92.

McCarroll, S. A., F. G. Kuruvilla, et al. (2008). "Integrated detection and population-genetic analysis of SNPs and copy number variation." Nat Genet **40**(10): 1166-74.

McLure, K. G. and P. W. Lee (1998). "How p53 binds DNA as a tetramer." EMBO J **17**(12): 3342-50.

Mehlen, P. and A. Puisieux (2006). "Metastasis: a question of life or death." Nat Rev Cancer **6**(6): 449-58.

Mehta, K. R., K. Nakao, et al. (2005). "Fractional genomic alteration detected by array-based comparative genomic hybridization independently predicts survival after hepatic resection for metastatic colorectal cancer." Clin Cancer Res **11**(5): 1791-7.

Mitelman, F., B. Johansson, et al. (2004). "Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer." Nat Genet **36**(4): 331-4.

Momand, J., D. Jung, et al. (1998). "The MDM2 gene amplification database." Nucleic Acids Res **26**(15): 3453-9.

Nagase, H. and Y. Nakamura (1993). "Mutations of the APC (adenomatous polyposis coli) gene." Hum Mutat **2**(6): 425-34.

Nakao, K., K. R. Mehta, et al. (2004). "High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization." Carcinogenesis **25**(8): 1345-57.

Nannini, M., M. A. Pantaleo, et al. (2009). "Gene expression profiling in colorectal cancer using microarray technologies: results and perspectives." Cancer Treat Rev **35**(3): 201-9.

Nathanson, D. R., A. T. t. Culliford, et al. (2003). "HER 2/neu expression and gene amplification in colon cancer." Int J Cancer **105**(6): 796-802.

NCI, N. C. I. (2009, 20090415). "Genetics of Colorectal Cancer." from www.cancer.gov.

Negrini, M. and G. A. Calin (2008). "Breast cancer metastasis: a microRNA story." Breast Cancer Res **10**(2): 203.

Nowak, M. A., N. L. Komarova, et al. (2002). "The role of chromosomal instability in tumor initiation." Proc Natl Acad Sci U S A **99**(25): 16226-31.

Nowell, P. C. (1976). "The clonal evolution of tumor cell populations." Science **194**(4260): 23-8.

Nowell, P. C. (1986). "Mechanisms of tumor progression." Cancer Res **46**(5): 2203-7.

Olshen, A. B., E. S. Venkatraman, et al. (2004). "Circular binary segmentation for the analysis of array-based DNA copy number data." Biostatistics **5**(4): 557-72.

Peltomaki, P. and H. F. Vasen (1997). "Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer." Gastroenterology **113**(4): 1146-58.

Pillai, R. S., S. N. Bhattacharyya, et al. (2007). "Repression of protein synthesis by miRNAs: how many mechanisms?" Trends Cell Biol **17**(3): 118-26.

Pittman, A. M., S. Naranjo, et al. (2009). "The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression." Genome Res.

Platzer, P., M. B. Upender, et al. (2002). "Silence of chromosomal amplifications in colon cancer." Cancer Res **62**(4): 1134-8.

R Development Core Team (2009). R: A Language and Enviroment for Statistical Computing.

Rabbitts, T. H. (1994). "Chromosomal translocations in human cancer." Nature **372**(6502): 143-9.

Redon, R., S. Ishikawa, et al. (2006). "Global variation in copy number in the human genome." Nature **444**(7118): 444-54.

Ried, T., K. Heselmeyer-Haddad, et al. (1999). "Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation." Genes Chromosomes Cancer **25**(3): 195-204.

Romond, E. H., E. A. Perez, et al. (2005). "Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer." N Engl J Med **353**(16): 1673-84.

Saal, L. H., C. Troein, et al. (2002). "BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data." Genome Biol **3**(8): SOFTWARE0003.

Saha, S., A. Bardelli, et al. (2001). "A phosphatase associated with metastasis of colorectal cancer." Science **294**(5545): 1343-6.

Santarosa, M. and A. Ashworth (2004). "Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way." Biochim Biophys Acta **1654**(2): 105-22.

Schena, M., D. Shalon, et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." Science **270**(5235): 467-70.

Scherer, S. W., C. Lee, et al. (2007). "Challenges and standards in integrating surveys of structural variation." Nat Genet **39**(7 Suppl): S7-15.

Schetter, A. J., S. Y. Leung, et al. (2008). "MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma." JAMA **299**(4): 425-36.

Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-8.

Sharp, A. J., D. P. Locke, et al. (2005). "Segmental duplications and copy-number variation in the human genome." Am J Hum Genet **77**(1): 78-88.

Sjoblom, T., S. Jones, et al. (2006). "The consensus coding sequences of human breast and colorectal cancers." Science **314**(5797): 268-74.

Skotheim, R. I., C. B. Diep, et al. (2001). "Evaluation of loss of heterozygosity/allelic imbalance scoring in tumor DNA." Cancer Genet Cytogenet **127**(1): 64-70.

Skotheim, R. I., G. E. Lind, et al. (2005). "Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development." Cancer Res **65**(13): 5588-98.

Smyth, G. K. (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments." Stat Appl Genet Mol Biol **3**: Article3.

Smyth, G. K. and T. Speed (2003). "Normalization of cDNA microarray data." Methods **31**(4): 265-73.

Sobin, L. H. and I. D. Fleming (1997). "TNM Classification of Malignant Tumors, fifth edition (1997). Union Internationale Contre le Cancer and the American Joint Committee on Cancer." Cancer **80**(9): 1803-4.

Stehelin, D., H. E. Varmus, et al. (1976). "DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA." Nature **260**(5547): 170-3.

Stranger, B. E., M. S. Forrest, et al. (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes." Science **315**(5813): 848-53.

Tenesa, A., S. M. Farrington, et al. (2008). "Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21." Nat Genet **40**(5): 631-7.

Thibodeau, S. N., G. Bren, et al. (1993). "Microsatellite instability in cancer of the proximal colon." Science **260**(5109): 816-9.

Tomlinson, I., E. Webb, et al. (2007). "A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21." Nat Genet **39**(8): 984-8.

Tomlinson, I. P., E. Webb, et al. (2008). "A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3." Nat Genet **40**(5): 623-30.

Tsafrir, D., M. Bacolod, et al. (2006). "Relationship of gene expression and chromosomal abnormalities in colorectal cancer." Cancer Res **66**(4): 2129-37.

Tuzun, E., A. J. Sharp, et al. (2005). "Fine-scale structural variation of the human genome." Nat Genet **37**(7): 727-32.

Urquhart, A., C. P. Kimpton, et al. (1994). "Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers." Int J Legal Med **107**(1): 13-20.

Wang, H., R. A. Ach, et al. (2007). "Direct and sensitive miRNA profiling from low-input total RNA." RNA **13**(1): 151-9.

Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-43.

Weinberg, R. A. (2007). The Biology of Cancer.

Weissenbach, J., G. Gyapay, et al. (1992). "A second-generation linkage map of the human genome." Nature **359**(6398): 794-801.

Venkatachalam, S., Y. P. Shi, et al. (1998). "Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation." EMBO J **17**(16): 4657-67.

WHO, W. H. O. (2002). National Cancer Control Programmes: policies and managerial guidelines. **2nd edition**.

Winawer, S. J., A. G. Zauber, et al. (1993). "Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup." N Engl J Med **329**(27): 1977-81.

Vogelstein, B. and K. W. Kinzler (2004). "Cancer genes and the pathways they control." Nat Med **10**(8): 789-99.

Wong, K. K., R. J. deLeeuw, et al. (2007). "A comprehensive analysis of common copy-number variations in the human genome." Am J Hum Genet **80**(1): 91-104.

Vousden, K. H. and X. Lu (2002). "Live or let die: the cell's response to p53." Nat Rev Cancer **2**(8): 594-604.

Vousden, K. H. and C. Prives (2005). "P53 and prognosis: new insights and further complexity." Cell **120**(1): 7-10.

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

Zanke, B. W., C. M. Greenwood, et al. (2007). "Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24." Nat Genet **39**(8): 989-94.

Zhang, J., L. Feuk, et al. (2006). "Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome." Cytogenet Genome Res **115**(3-4): 205-14.