Università degli Studi di Catania
Scuola Superiore di Catania

International PhD

in Stem Cells

XXIV cycle

Novel genomic technologies and molecular diagnostics in Colorectal Cancer

Carmela Capizzi

Coordinator of PhD
Prof. Daniela F. Condorelli

Tutor
Prof.ssa Vincenza Barresi

a.a. 2008/2011

"Science is essentially a cultural activity.
It generates pure knowledge about ourselves and about the universe we live in, knowledge that continually reshapes our thinking"
John Sulston

To my family who helped me,
sometimes with a word,
sometimes with a simple glance.

Acknowledgments

This research reflects the dedication and support of many people for whom I am truly thankful.

I want especially to express my gratitude to my supervisors Prof. Daniele Condorelli and Prof. Vincenza Barresi for giving me the opportunity to combine my two passions of bioinformatics and research. I would also like to thank them for their advice, patience and support throughout the PhD.

My special thanks go to Nicolò Musso, my colleague, for helping me with the laboratory analyses and for supporting me during these years. If today I cross this line it is also thanks to him.

I also would like to thank the great collaboration of the "Casa Di Cura S.r.l. G.B. Morgagni", who provided tumor pieces on which we have worked.

I also would like to acknowledge my friends who have been very near and who have put up with my mood swings.

Lastly, I would like thank my parents and my brother Mario for their love and for always believing in me.

Abstract

Cancer is a disease of the genome that is characterized by substantial variability in the clinical course and response to therapies. *Colorectal cancer* (*CRC*) is a heterogeneous cancer and represents an ideal model to investigate and elucidate the genetic alterations involved in tumor onset and progression. In this study 51 CRC patients were subdivided into groups according to the presence of microsatellite instability (MSI) and chromosomal instability (CIN). Of the 51 CRCs, 13.73% were MSI and 86.27% were microsatellite stable (MSS). The frequency of KRAS mutations in MSI-H and in MSS cancer was 28.57% and 40.91%, respectively.

To identify and characterize genomic alteration associated with colorectal cancer (CRC) the samples were analyzed with the last generation of Affymetrix single nucleotide polymorphism/CNV microarrays (SNP Array 6.0) and two new tools were implemented, *Broad Cytogenetic Analysis* (*BroCyA)* and *Focal Cytogenetic Analysis* (*FoCyA)*, to identify broad (> ¼ chromosomal arm) and focal aberrations (< ¼ chromosomal arm). Broad copy number gains were noted on chromosomes 7, 8q, 9, 13q, 17q, 20 and broad copy number losses on chromosomes 4, 5q, 8p, 17p, 18, 19p, 20p and 22q. Moreover recurrent high level amplifications (HLAs) (copy number > 5.2) were located on chromosome 20, in regions containing known cancer pathway genes as STK4 and ID1, and homozygous deletions (HoD) containing potential new candidate tumor, suppressors such as BTG4 and D4S234E were located on chromosomes 11 and 4.

Recurrent somatic focal events (gains and losses) were identified in regions encompassing potential new candidate tumor suppressors and oncogenes, such as A2BP1 and PRDM16

Finally, several copy neutral-loss of heterozygosities (CN-LOHs) were detected, more frequently on chromosome 7p and 22q.

In conclusion, in this study some novel broad and focal copy number abnormalities (CNAs) and CN-LOHs were revealed in CRC. The precise and large-scale measurement of CNAs and CN-LOHs in the CRC genome provides a list of genes that might be involved in cancer development.

Contents

List of abbreviations

AJCC	American Joint Committee on Cancer
BroCyA	Broad Cytogenetic Analysis
CIN	Chromosomal INstability
CN-LOH	Copy Neutral Loss of Heterozygosity
CNAs	Copy Number Aberrations
CNVs	Copy Number Variations
CW	Contiguous Windows
CRC	Colorectal Cancer
FoCyA	Focal Cytogenetic Analysis
GTC	Genotyping Console Software
HeD	Heterozygous Deletions
HLA	High Level Amplification
HoD	Homozygous Deletion
MAPD	Median Absolute Pairwise Difference
MSI	MicroSatellite Instability
MSS	MicroSatellite Stability
Mu	Normal Mucosa sample
NK-AML	Normal Karyotype Acute Myeloid Leukemia
PSOS	Paired Segment Overlap
QC	Quality Control
RSA	Recurrent Segment Analysis
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism[I]
Tu	Tumor sample

[I]

1 Introduction

1.1 Biological background

The completion of human genome project has given new impetus to the study of human variation, showing that every individual is different from any other for only 0.5% of their DNA sequence. Responsible for this variable portion of the genome, in particular are the single nucleotide polymorphisms (SNPs - *Single Nucleotide Polymorphisms*), specific DNA bases that vary in individuals with a higher frequency than that found for point mutations. Other well known variations in the genome are copy number changes. Copy number change refers to the phenomena that the number of copies of a particular DNA segment varies among individuals. Copy number changes are either acquired by heredity (germline copy number changes) or postnatal development (somatic copy number changes). Germline and somatic copy number changes are usually referred to as copy number variations (CNVs) and copy number alterations (CNAs), respectively [110].

The study of variability represents a challenge for modern medicine, especially in the prospect of curing the sick by identifying the most effective treatment.

1.2 The International HapMap Project

The International HapMap Project (www.hapmap.org) was started in 2002 with the aim of identifying polymorphisms in the human genome and studying the distribution of these polymorphisms both within the genome of an individual, and across populations. Its completion in 2003 paved the way for studies to better understand and catalogue polymorphisms in the human genome.

The HapMap project has collected SNP data from 270 individuals belonging to four different populations/ethnicities. The 270 individuals were distributed among the following four populations:

1. Ninety Yoruba individuals from Ibadan, Nigeria (YRI). This dataset consists of 30 trios. Each trio consists of three related individuals - two parents and an adult child. All these individuals belong to a single community in Ibadan, Nigeria. All the individuals selected had four Yoruba grandparents.
2. Ninety individuals of European origin (CEU). This dataset consists of 30 trios from Utah with northern and western European ancestry. These samples were collected by the Centre d'Etude du Polymorphisme Humain (CEPH) in 1980.
3. Forty-five individuals from Tokyo, Japan (JPT). These are forty-five unrelated individuals from Tokyo. Each individual selected had all four Japanese grandparents.
4. Forty-five Han Chinese from Beijing (HCB). These are forty-five unrelated individuals living in the residential community of Beijing Normal University. These are all individuals who described themselves as having at least three out of four Han Chinese grandparents.

The phase I HapMap showed variation patterns for the four populations. SNPs were selected at 5kb intervals across the genome, with the requirement that the minor allele frequency (MAF) be >0.05, which is defined as “common” SNP. Approximately 1.3 million SNPs were genotyped in this phase of the project (InternationalHapMapConsortium). In phase II of the HapMap project, a further 2.1 million SNPs were genotyped for the same set of individuals. The resulting marker map had a SNP density of approximately one per

kilobase. In phase II, the marker selection criteria did not include a requirement for only common SNPs, so this HapMap contains more low frequency SNPs, with a better representation of rare SNPs. [59]

1.3 Single nucleotide polymorphisms

SNPs are single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some populations, wherein the least frequent allele has an abundance of at least 1% or greater.

Single nucleotide polymorphisms (SNPs) have an estimated density of one per every 1000 bp along the human genome, if two individuals are compared. The total density of SNPs among people all over the world is obviously much higher. Although SNPs occur both in coding and non-coding regions of the genome, they are distributed with unequal spacing [99]. Most SNPs do not have an effect on cell function, but some are believed, for example, to confer susceptibility or resistance to a disease or determine the severity or progression of disease. This depends largely on where a SNP occurs [111]. Non-synonymous (amino acid altering) SNPs in coding regions of a gene are of course believed to be important in causing genetic diseases because they may change the structure and function of the encoded protein. However, it has recently been identified that also synonymous SNPs (silent; coding SNPs that do not alter the amino acid), intergenic SNPs, and SNPs in introns and other non-coding regions may be functional. If these variations are located in promoters, splice junctions, or 5' and 3' untranslated regions, they may alter the structure, function, and expression of the gene product by affecting the regulation, splicing, and mRNA stability of a gene [21, 81].

Most SNPs seen in human populations are bi-allelic, i.e. there are two alleles seen in a population - the original nucleotide and the mutation. For a SNP to have three common alleles, a new mutation must happen at the same location in another individual and this mutation should also increase in frequency. The probability of observing this in human polymorphism data is low for two reasons:

1. single base pair mutation rates in the human genome are low (of the order of 10−9 per base pair per generation), and
2. human populations are relatively recent in origin.

1.4 Copy number change

1.4.1 Large-scale copy number variations

Large-scale copy number variations (CNVs), also called copy number polymorphisms (CNPs), constitute a large proportion of human genetic diversity [56, 105]. CNVs are defined as repetitive sequences, in which a repetitive unit is as large as 100 kb or greater. CNVs are widely distributed throughout the genome, also in coding regions, although obvious hot spots do exist. The average size of CNVs is estimated to be 300-460 kb and there are, on average, 11-12.4 CNVs between two individuals [56, 105]. However, these numbers are probably underestimates due to the small number of individuals assessed in these studies and the limited resolution of detection methods. It is still unclear what other genomic features, in addition to segmental duplications, enhance the occurrence of CNVs [43]. These CNVs are likely to complicate the genotyping of smaller variations such as SNPs and microsatellites.

1.4.2 Copy number abnormalities (CNAs)

Copy number abnormalities (CNAs) in genomic DNA have been associated with complex human diseases, including cancer [2, 41, 71, 75, 86, 105, 116]. In cancer, for instance, amplification of oncogenes is one possible mechanism for tumor activation [52-53]. Patient survival and metastasis development have been shown to be associated with certain CNAs [2, 41, 71, 75, 86, 105, 116] and, by relating patterns of CNAs with survival, gene expression, and disease status, studies about copy number changes have been instrumental for identifying relevant genes for cancer development and patient classification [71, 86-87].

CNAs found in cancer include whole-chromosome or regional alterations spanning part to whole arms of a chromosomes.

CNAs include:

1. Gains: a copy number gain represents an increase of one or a small number of copies of a DNA segment, typically spanning a large genomic region. If the gain consists of just one additional copy of a segment of DNA, it may be called a duplication.
2. High Level Amplifications: an high level amplification is a type of gain that can reach high copy numbers, at least 5. In the context of cancer biology, amplifications are gain-of-function mutations often seen in oncogenes.
3. Deletions: a deletion is the loss of genetic material, either heterozygous (if copy number is 1) or homozygous (when copy number is 0; also called nullisomy).

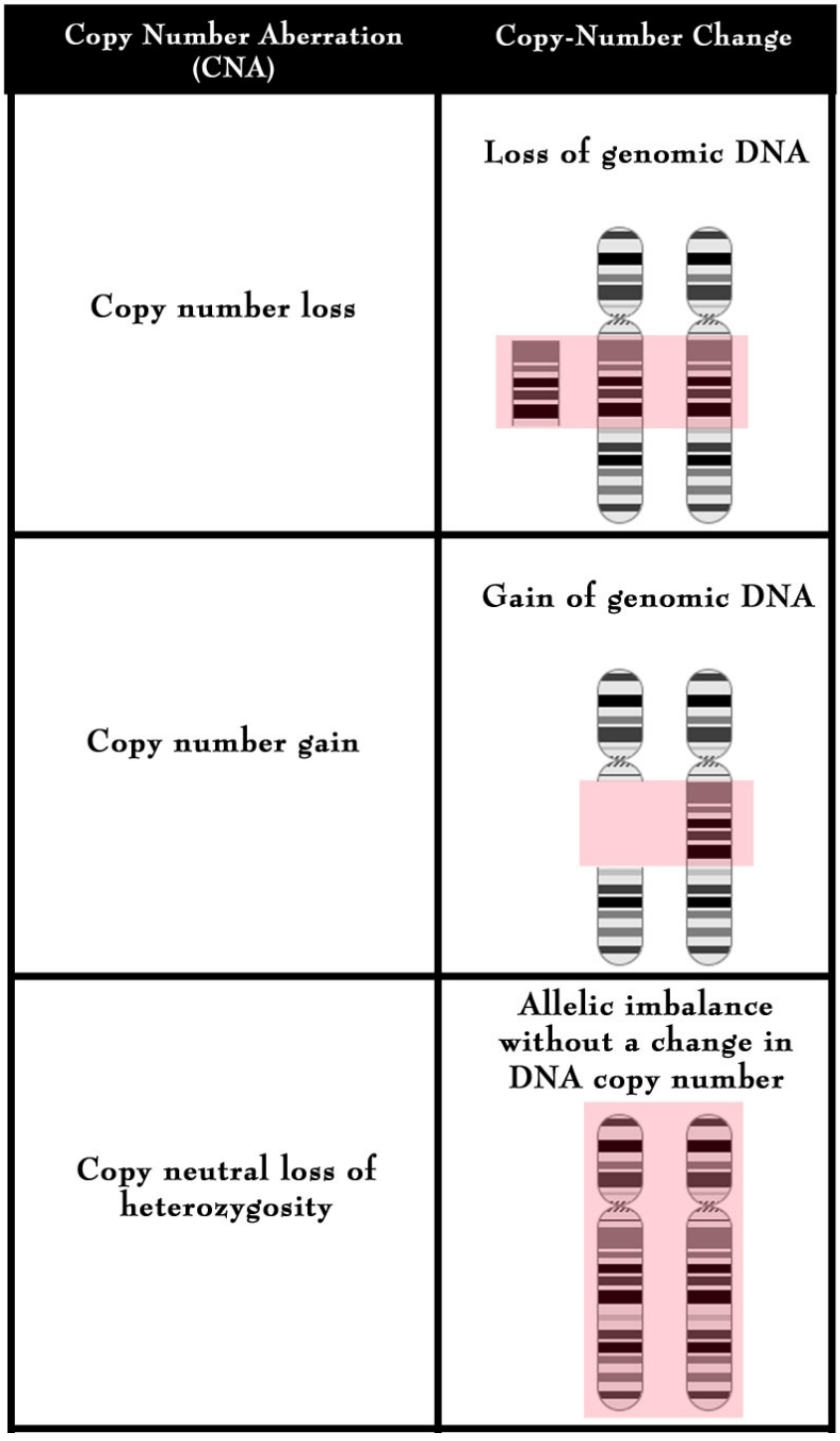

Figure 1
Genomic aberrations, CNA and CN-LOH.

1.4.3 Copy Neutral Loss of Heterozygosity

CN-LOH, also known as uniparental disomy (UPD), is observed in both hematological and solid tumors. In CN-LOH, one allele is

duplicated whilst there is loss of the other allele, hence a genomic event has taken place without a change in copy-number (Figure 1). This may lead to clinical conditions by producing either homozygosity for recessive mutations or else aberrant patterns of imprinting. Through the use of SNP-arrays, CN-LOH has been found to be a common event in many types of cancers and as such, serves as an alternative to a deletion, in terms of the 'second hit' in the Knudson two hit hypothesis of tumorigenesis.

1.5 SNP genotyping techniques

High-throughput genotyping technologies have been developing very rapidly during the past few years, and similar development is expected to continue. There are several high-throughput SNP genotyping techniques available today, with the capacity of several million genotypes per day. At present, there are at least two technologies that seem to be able to fulfill the needs of genome-wide association studies: genome-wide microarrays by Affymetrix and Illumina.

1.5.1 Genomic arrays

Array-based comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) arrays are two high resolution techniques that measure copy-number alterations (CNAs) and thus are important tools for studying genetic events in for instance cancer and developmental disorders.

CGH is a quantitative method based on the comparative hybridization of two samples (patient/tumor and a reference sample), labeled with different fluorescent colours, to metaphase spreads from a

healthy control. Large scale genomic alterations can then be detected based on the fluorescence ratio of the hybridized DNA samples from the patient compared to the reference.

CGH-arrays are a further development of the same principle. Here, hybridization of labeled DNA is performed on microarray slides that contain probes, each representing a unique DNA sequence. This method allows a higher resolution and provides the exact positions of the chromosomal aberration compared to conventional CGH.

SNP-arrays are at present predominantly used in research for genotyping and screening of genomic aberrations, yet are also applied for diagnostic and prognostic purposes. SNP-arrays offer high resolution power and can detect small changes in copy number (~10-25kb) but can also detect copy neutral loss of heterozygosity (CN-LOH). For the work presented within this thesis, Affymetrix SNP-arrays were applied for whole-genome screening, which is described below.

1.5.2 Single nucleotide polymorphism arrays

The Affymetrix GeneChip SNP-array consists of a square glass substrate mounted in a plastic cartridge where the glass contains an array of oligonucleotides each 25bps in length. For each SNP, different oligonucleotides of 25bp are tiled, all with a slight variation in perfect matches, mismatches, and flanking sequences surrounding the SNP. Tumor DNA is fragmented using the restriction enzyme Nsp1, and adapters are then added to allow PCR amplification of the fragments. Following the amplification, fragments are again fragmented and labeled. These labeled fragments are then hybridized to the microarray chip, the non-bound material is washed off, and the signals emitted from the fluorescent probes are detected, measured and stored for analysis

(www.affymetrix.com). Affymetrix currently manufactures SNP-array chips incorporating 10K to 2.7M markers, thereby providing great resolution to enable the detection of both known and novel aberrations throughout the entire genome. Consequently, SNP-array technology allows the alignment of SNPs in chromosomal order and the identification of chromosomal alterations such as CNAs and CN-LOH (Figure 1).

The advantage of using SNP-arrays is that they can detect both copy number, LOH (i.e. a deletion) and CN-LOH in comparison to FISH and array-CGH, which can only detect copy-loss LOH.

2 Colorectal Cancer

2.1 Cancer

Cancer is one of the most important health problems of the current era and also a leading cause of death among populations. In order to be successful in the treatment of cancer, early diagnosis, before the tumor spreads to the surrounding tissues or distant organs, is mandatory. It is now known that cancer originates through a multistep process. In this model, the first stage, the initiation, is caused by the acquisition in a cell of a mutation that can provide a growth advantage and/or irreversible alterations in cellular homeostasis and differentiation. The next step, the promotion, can be a potentially reversible or interruptible clonal expansion of the initiated cell by a combination of growth stimulation and inhibition of apoptosis. Further progression steps occur upon clonal expansion of the initial cells and accumulation of a sufficient number of mutations and epigenetic alterations to acquire growth stimulus-independency and resistance to growth inhibitors and apoptosis, ultimately leading to an unlimited replicative potential. The acquisition of the ability to invade the surrounding tissue defines the malignant character of cancer cells, while the process through which cells can migrate to distal organs and acquire the potential to form metastasis represents the achievement of a full malignant cancerous phenotype.

Colorectal cancer represents an ideal model to investigate and elucidate the genetic alterations involved in tumor onset and progression, mainly because it arises and progresses through a series of well-defined histopatological changes, the so-called adenoma-carcinoma sequence.

2.2 General background on Colorectal Cancer

Colorectal cancer (CRC) starts in the colon or the rectum. CRC is a disease primarily observed in longstanding developed nations. However, in recent years, high CRC rates have been reported also in newly developed countries.

This cancer is rare before age 40, appearing more frequently around 60 years. Incidence and mortality from colorectal cancer are similar in both men and women. Screening can reduce the mortality associated with the disease, but the participation rates are still suboptimal [23].

2.3 Anatomy

The colon is the last segment of the human digestive system. After food is digested in the stomach, it enters the small intestine where the nutrients are absorbed through digestion. The indigestible part is then passed to the large intestine and eventually expelled from the body through the rectum using the specialized muscles and nerves in the anus, which acts as a valve. See Figure 2 below (from the individual's vantage) :

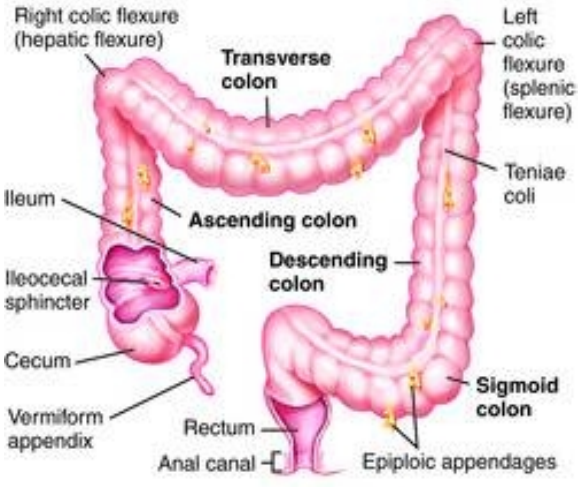

Figure 2
A diagram of the colon.

The large intestine consists of the colon and the rectum (called the terminal extraperitoneal segment). It is called "*large*" because its diameter is roughly 5 to 6.5 cm in diameter in the cecum and right colon (although it narrows to about 2.5 cm at the end of the rectum). In general, the purpose of the colon is to absorb the water and mineral salts from undigested food, with the residue passing as faeces towards the rectum where it is to be excreted. The length of the colon is typically about 1.5 m and is composed of five sections: ascending colon, transverse colon, descending colon, sigmoid colon, and rectum. Unlike the small intestine, which is almost sterile, the colon has significant bacteria (which have some very beneficial effects). In fact, about 90% of the dry weight of the stools is bacteria and not undigested food.

The colon is somewhat like a corrugated tube. It is elastic, flexible and can expand and move. There are several named junctures in the colon, however the two main "*kinks*" are the right colic flexure (between the ascending and transverse colon) and the left colic flexure (between the transverse and descending colon, and also called the splenic flexure). The sigmoid maintains more of an "*s*" shape.

Although the colon wall contains several layers of tissue, the inner lining or epithelium is of greatest interest relative to colon cancer because that is where most colon cancers begin. The colonic epithelium has a glandular appearance from the inside and acts principally to absorb water and secrete mucus. It is characterized by the long, thin pits called crypts, which contain special cells.

2.4 The stage

The stage describes the extent of the cancer in the body. It is based on how far the cancer has grown into the wall of the intestine, whether or not it has reached nearby structures, and whether or not it has spread to the lymph nodes or distant organs. The stage of a cancer is one of the most important factors in determining prognosis and treatment options.
There are actually 2 types of staging for colorectal cancer.

- The *clinical stage*, based on the results of the physical exam, biopsy, and any imaging tests.
- The *pathologic stage*, which is based on the same factors as the clinical stage, plus what is found as a result of the surgery.

The most commonly used staging system for colorectal cancer is that of the *American Joint Committee on Cancer* (*AJCC*). The stage is expressed in Roman numerals from stage I (the least advanced) to stage IV (the most advanced).

- In *stage 0*, abnormal cells are found in the mucosa (innermost layer) of the colon wall. These abnormal cells may become cancer and spread. Stage 0 is also called carcinoma in situ.
- In *stage I*, cancer has formed in the mucosa (innermost layer) of the colon wall and has spread to the submucosa (layer of tissue

under the mucosa). Cancer may have spread to the muscle layer of the colon wall.

- *Stage II* is subdivided in *stage IIA*, *stage IIB*, and *stage IIC*.
    - A: cancer has spread through the muscle layer of the colon wall to the serosa (outermost layer) of the colon wall.
    - B: cancer has spread through the serosa (outermost layer) of the colon wall but has not spread to nearby organs.
    - C: cancer has spread through the serosa (outermost layer) of the colon wall to nearby organs.

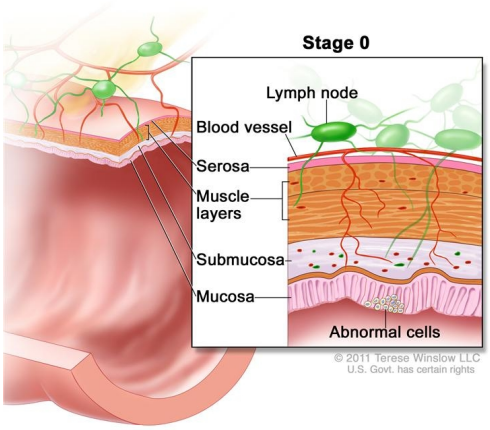

Figure 3
Stage 0 (colon carcinoma in situ).
Abnormal cells are shown in the mucosa of the colon wall.

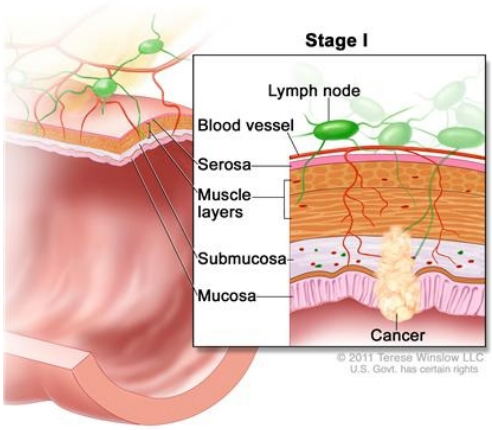

Figure 4
Stage I colon cancer.
Cancer has spread from the mucosa of the colon wall to the submucosa.

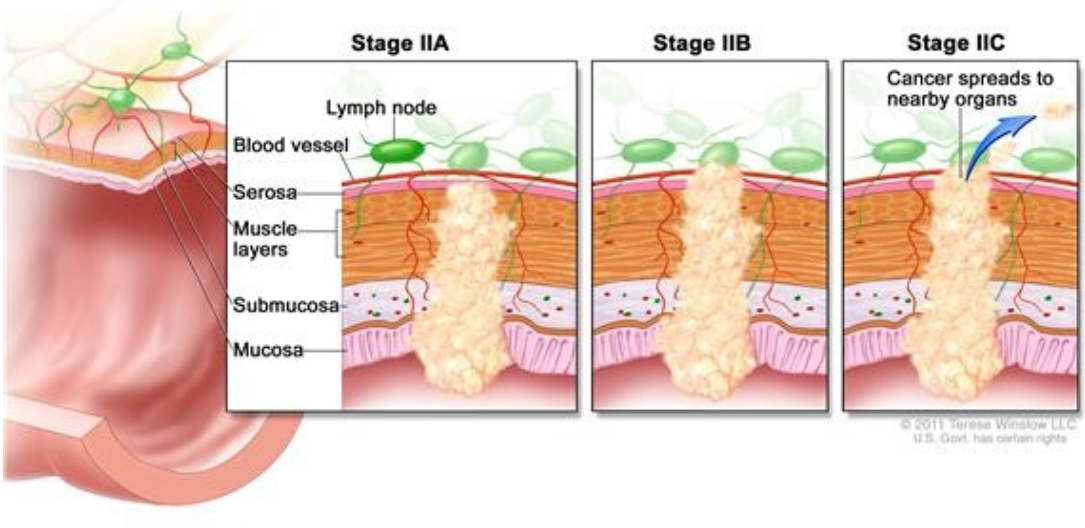

Figure 5
Stage II colon cancer.
In *stage IIA*, cancer has spread through the muscle layer of the colon wall to the serosa. In *stage IIB*, cancer has spread through the serosa but has not spread to nearby organs. In *stage IIC*, cancer has spread through the serosa to nearby organs.

- *Stage III* colon cancer is divided into *stage IIIA*, *stage IIIB*, and *stage IIIC*.
    - In *stage IIIA*:

▪ Cancer may have spread through the mucosa (innermost layer) of the colon wall to the submucosa (layer of tissue under the mucosa) and may have spread to the muscle layer of the colon wall. Cancer has spread to at least one but not more than 3 nearby lymph nodes or cancer cells have formed in tissues near the lymph nodes; or

- Cancer has spread through the mucosa (innermost layer) of the colon wall to the submucosa (layer of tissue under the mucosa). Cancer has spread to at least 4 but not more than 6 nearby lymph nodes.

- In *stage IIIB*:
    - Cancer has spread through the muscle layer of the colon wall to the serosa (outermost layer) of the colon

wall or has spread through the serosa but not to nearby organs. Cancer has spread to at least one but not more than 3 nearby lymph nodes or cancer cells have formed in tissues near the lymph nodes; or

- Cancer has spread to the muscle layer of the colon wall or to the serosa (outermost layer) of the colon wall. Cancer has spread to at least 4 but not more than 6 nearby lymph nodes; or
- Cancer has spread through the mucosa (innermost layer) of the colon wall to the submucosa (layer of tissue under the mucosa) and may have spread to the muscle layer of the colon wall. Cancer has spread to 7 or more nearby lymph nodes.

- In *stage IIIC*:
  - Cancer has spread through the serosa (outermost layer) of the colon wall but has not spread to nearby organs. Cancer has spread to at least 4 but not more than 6 nearby lymph nodes; or
  - Cancer has spread through the muscle layer of the colon wall to the serosa (outermost layer) of the colon wall or has spread through the serosa but has not spread to nearby organs. Cancer has spread to 7 or more nearby lymph nodes; or
  - Cancer has spread through the serosa (outermost layer) of the colon wall and has spread to nearby organs. Cancer has spread to one or more nearby lymph nodes or cancer cells have formed in tissues near the lymph nodes.

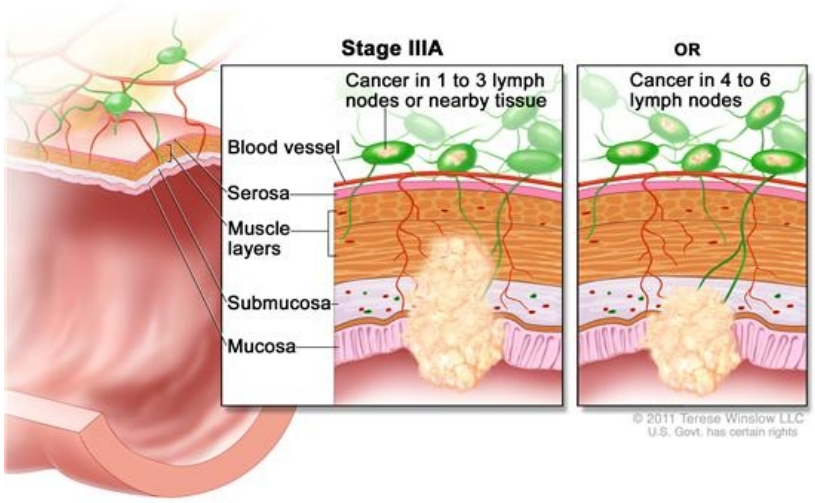

Figure 6
Stage IIIA colon cancer.
Cancer may have spread through the mucosa of the colon wall to the submucosa and muscle layer, and has spread to one to three nearby lymph nodes or tissues near the lymph nodes. OR, cancer has spread through the mucosa to the submucosa and four to six nearby lymph nodes.

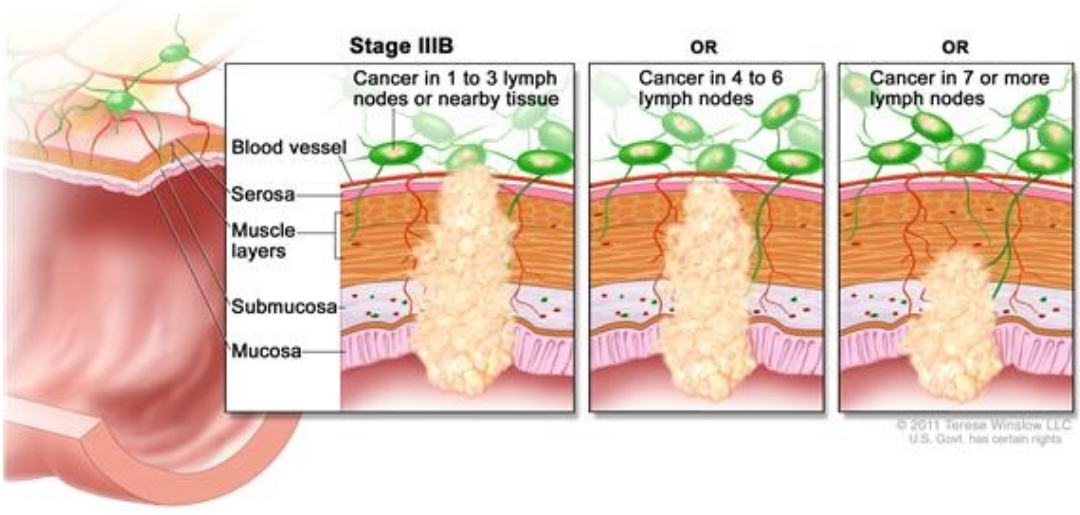

Figure 7
Stage IIIB colon cancer.
Cancer has spread through the muscle layer of the colon wall to the serosa or has spread through the serosa but not to nearby organs; cancer has spread to one to three nearby lymph nodes or to tissues near the lymph nodes. OR, cancer has spread to the muscle layer or to the serosa, and to four to six nearby lymph nodes. OR, cancer has spread through the mucosa to the submucosa and may have spread to the muscle layer; cancer has spread to seven or more nearby lymph nodes.

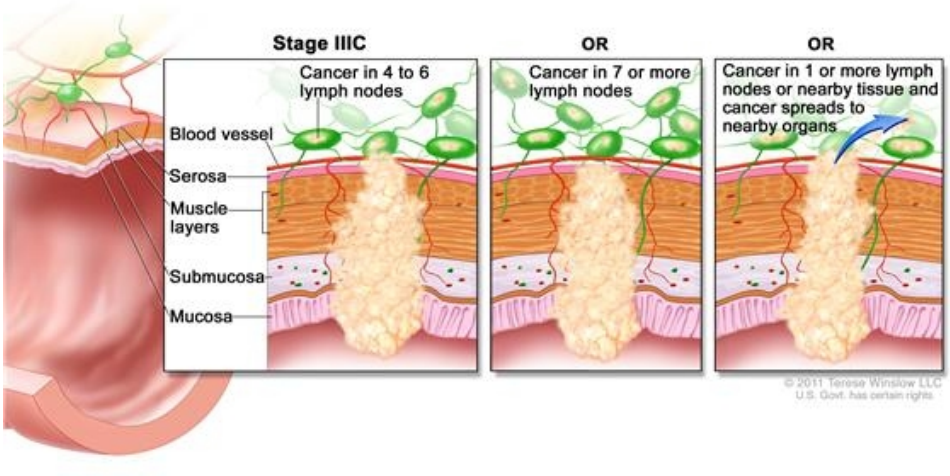

Figure 8
Stage IIIC colon cancer.
Cancer has spread through the serosa of the colon wall but not to nearby organs; cancer has spread to four to six nearby lymph nodes. OR, cancer has spread through the muscle layer to the serosa or has spread through the serosa but not to nearby organs; cancer has spread to seven or more nearby lymph nodes. OR, cancer has spread through the serosa to nearby organs and to one or more nearby lymph nodes or to tissues near the lymph nodes.

- *Stage IV* colon cancer is divided into stage IVA and stage IVB.
  - *Stage IVA*: Cancer may have spread through the colon wall and may have spread to nearby organs or lymph nodes. Cancer has spread to one organ that is not near the colon, such as the liver, lung, or ovary, or to a distant lymph node.
  - *Stage IVB*: Cancer may have spread through the colon wall and may have spread to nearby organs or lymph nodes. Cancer has spread to more than one organ that is not near the colon or into the lining of the abdominal wall.

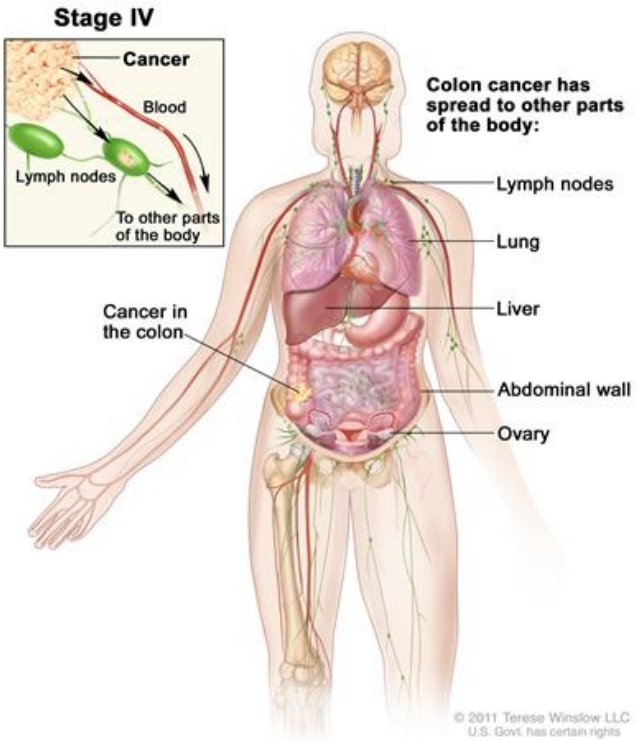

Figure 9
Stage IV colon cancer.
The cancer has spread through the blood and lymph nodes to other parts of the body, such as the lung, liver, abdominal wall, or ovary.

Another staging system is the *TNM system*. The *TNM system* describes 3 key pieces of information:

- **T** describes how far the main (primary) tumor has grown into the wall of the intestine and whether it has grown into nearby areas.
- **N** describes the extent of spread to nearby (regional) lymph nodes. Lymph nodes are small bean-shaped collections of immune system cells that are important in fighting infections.
- **M** indicates whether the cancer has spread (metastasized) to other organs of the body. (CRC can spread almost anywhere in the body, but the most common sites of spread are the liver and lungs.)

Numbers or letters appear after T, N, and M to provide more details about each of these factors. The numbers 0 through 4 indicate increasing severity. The letter X means "cannot be assessed because the information is not available."

Older staging systems for colorectal cancer, such as *the Dukes and Astler-Coller systems,* are mentioned briefly below for comparison.

Dukes classification changed by Astler-Coller	**American Joint Committee on Cancer (AJCC) stage**	**TNM stage**	**TNM stage criteria**
A	Stage 0	T_{is}, N_0, M_0	T_{is}: Tumor is in mucosa; *carcer-in-situ*
B1	Stage I	T_1, N_0, M_0	T_1: tumor invades submucosa
B1	Stage I	T_2, N_0, M_0	T_2: tumor invades muscularis propria
B2	Stage IIA	T_3, N_0, M_0	T_3: tumor invades subserosa or beyond (without other organsinvolved)
B2	Stage IIB	T_4, N_0, M_0	T_4: tumor invades adjacent organs or perforates the visceral peritoneum
C	Stage IIIA	T_{1-2}, N_1, M_0	T_1 o T_2. N_1: metastasis to 1 to 3 regional lymph nodes
C	Stage IIIB	T_{3-4} N_1 M_0	T_3 o T_4. N_1: metastasis to 1 to 3 regional lymph nodes
C	Stage IIIC	T_{1-4}, N_2 M_0	Any T. N_2: metastasis to 4 or more regional lymph nodes
D	Stage IV	T_{1-4}, any N, M_1	Any T and N. M_1: distant metastases present

Table 1
Comparison of AJCC, TNM, and Dukes stages.

2.5 Risk Factor in colorectal cancer

The cause of colorectal cancer is still relatively unknown, although researchers have accumulated a considerable amount of information on the factors, which may increase one's risk of developing the disease. Today, the disease, like all other forms of cancer, is considered to be the end result of many factors, both environmental and hereditary.

Colorectal cancer is a disease that affecting individuals above 40 years of age and 90% of cases occur in persons over the age of 50. Individuals with a family history of colorectal cancer are at an increased risk of developing the disease. The degree of risk depends upon the type of relative affected. [82]

CRC occurs both in women and in men: men tend to get colorectal cancer at an earlier age than women, but women live longer so they 'catch up' with men and thus the total number of cases in men and women is equal. (Fig. 10)

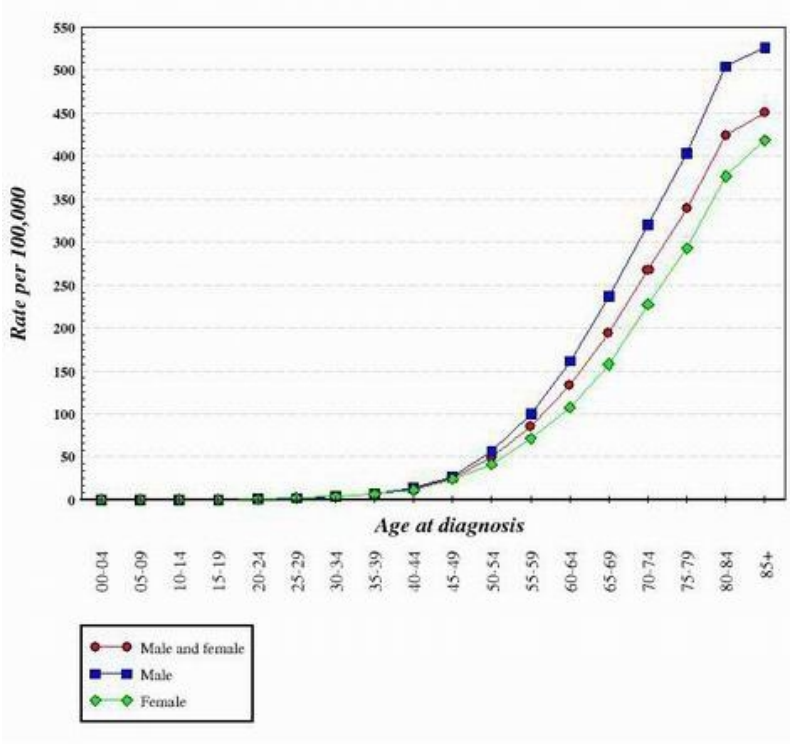

Figure 10
The graph below shows colon cancer rates in the United States as a function of age. (Public domain images via http://www.cancerquest.org/colon-rectal-cancer-risks.html)

Different studies have demonstrated that incidence of colon cancer correlates greatly with certain lifestyle factors, including diet. It is very difficult, however, to identify dietary items that cause a particular cancer. Studies show correlations between chronic heavy alcohol consumption and low folate intake and an increased risk of colorectal cancer[106]. On the other hand, some dietary factors are associated with a decreased risk of colorectal cancer. Research suggests that a diet rich in fruits and vegetables may provide a protective effect against the disease [93]. Nonetheless, the influence of these dietary factors on colorectal cancer risk is a topic still under debate.

2.6 Treatments

Treatment depends partly on the stage of the cancer. In general, CRC patients with receive post-operative chemotherapy if the lymph nodes are positive. Treatment is also determined by the patient's age, medical history, overall health, and tolerance for specific medications and therapies.

In general, treatments may include:

- Surgery (most often a colectomy) to remove cancer cells
- Chemotherapy to kill cancer cells
- Radiation therapy to destroy cancerous tissue

2.6.1 Surgery

Surgery is the primary treatment for CRC and over the past decades the ratio of CRC patients with a potentially curable disease has increased due to improved surgical techniques. This procedure aims to remove the affected bowel section and its lymphatic system.

2.6.2 Chemotherapy

Adjuvant chemotherapy has been developed to reduce the incidence of relapse. Its role is still a subject of debate in stage II CRC as approximately 75% of the patients are cured by surgery alone and adjuvant chemotherapy would only cure an additional 1-6% [15]. Instead, it is systematically used as adjuvant therapy for tumors that have reached the third stage of progression (invasion of serosa and lymph node involvement), with the aim of reducing recurrence risk. Moreover, it represents the first-line treatment in metastatic patients and aims to prolong survival and improve quality of life. Unfortunately, drug

treatments generally produce only a partial and short-termed clinical response [27].

The most widely used anti-cancer drugs in CRC treatment are 5-fluorouracil (5-FU), oxaliplatin, irinotecan and their possible combinations.

5-FU was developed in 1957 by Heidelberger and colleagues. The main anti-tumoral effect of 5-FU is a competitive inhibition of thymidylate synthase (TS), a rate-limiting enzyme involved in DNA synthesis and repair. 5-Fu can also exert its anticancer effects through incorporation of its metabolites into RNA and DNA.

As a single agent, 5-FU shows little activity against the most advanced forms of cancer [27, 47, 102]. Although the initial response rate is improved by combining 5-FU with leucovorin (LV), a chemically synthesized reduced folate, also referred to as folinic acid [79], thanks to its ability to inhibit thymidylate synthase [78], there is not a significant increase in survival rate [84].

This led to the development of new drugs with an analogous mechanism of action such as capecitabine, which used alone or in combination with LV, induces a better response rate with a lower toxicity profile [22].

Oxaliplatin was developed as an analogue of cisplatin in order to achieve greater therapeutic efficacy. Like all alkylating compounds, oxaliplatin is able to form guanine-guanine or adenine-guanine adducts between DNA complementary strands. These adducts hinder DNA polymerase progression during replication, thus interfering with normal cell division processes. Clinical studies have proven oxaliplatin effectiveness either as CRC first-line therapy or as a secondary treatment of 5-FU refractory cancers [25, 73].

Irinotecan is a semisynthetic inhibitor of topoisomerase I, a nuclear enzyme important in DNA uncoiling for replication and transcription [54, 60]. Clinical studies have shown that in patients insensible to 5-FU treatment, irinotecan produced a response rate of 13.5% and tumor stabilization in 44% of cases with a median survival of 45 weeks [118]. This has led to irinotecan acquisition as secondary treatment of 5-FU insensible CRC patients. Studies investigating oxaliplatin/5-FU/leucovorin (FOLFOX) combination benefits started after the observation of a synergistic effect of these drugs in vitro and in mice models [94]. In an international multicenter study (MOSAIC) patients have shown a significant increase in 3 years disease-free survival, with a 23% reduction in recurrence risk, compared to control, and a moderate toxicity profile [7]. Some studies on metastatic patients [17, 34, 100] evaluated the irinotecan/5-FU/leucovorin combination effectiveness (FOLFIRI). Compared to irinotecan alone, the results have shown a 21% to 39% increase in response rate, a 4.3 to 7 months increase in progression free survival and a 12.6 to 14.8 months increase of median survival. In cases of unresponsiveness to first-line therapy, it is possible to proceed with palliative chemo/radiotherapy treatment to reduce bleeding and pain.

Liver is the most common CRC metastatization site. Liver metastases tend to appear within two years after surgical removal of primary tumor in 70-80% of cases [19, 96]. The main therapeutic approach in the treatment of liver metastases involves the surgical removal of liver affected tissue, followed by chemotherapy. The 5-year survival does not exceed 40-50% [1, 38, 45, 93].

2.6.3 Radiation

Although radiation therapy is occasionally used in patients with colon cancer, there is no standard protocol for it and the system is determined by the patient's general condition and severity of symptoms. It is considered to be more appropriate for managing tumors of the rectum, rather than tumors of the colon and has often been used in these cases either neoadjuvantly, to reduce the size of the tumor prior to surgery, adjuvantly, to help prevent local recurrence, and palliatively, to relieve symptoms such as pain and bleeding.

2.7 Prognostic and Predictive Biomarkers in Colorectal Cancer

Genetics has a key role in colorectal cancer and, in the recent years, many genetic alterations observed in CRC have been proposed as biomarkers to estimate CRC prognosis.

2.7.1 Genetic Phenotypes

The molecular alterations are considered as alternative mechanism driving of colorectal cancer and they are categorized according to two main specific phenotypes of genomic instability. The most common forms of genetic instability, known in colorectal cancer, are:

1. ***Chromosomal INstability*** (CIN), that is observed in 80-85% of sporadic colorectal cancers and is characterized by large structural and numerical chromosomal abnormalities (aneuploidy) [93]. CIN seems to develop from errors in the DNA replication checkpoint and mitotic-spindle checkpoint [46], but the mechanism of CIN is not clear.

It is characterized by the mutation of APC or β-catenin that causes excessive proliferation of the epithelium and aberrant cell-cell interaction [64, 93]. Following the loss of APC, it is can accumulate a series of mutations of different genes; these events constitute a '*multistep process*', where in each molecular modification there is a different pathological and clinical entity. The chromosomal abnormalities of CIN group are not distributed randomly among different chromosomes, but are repeated with regularity at the expense of some chromosomes. Among the most often are observed trisomies of chromosomes 7, 8q, 13 and 20 and deletions of chromosomes 8p, 17p (where is located p53, a tumor-suppressor gene) and 18q. Several retrospective studies have demonstrated the prognostic importance of certain cytogenetic abnormalities (of (18q), del (8p), del (4), del (14), del (15q)), particularly when associated with one another [3-4, 13, 107].

2. ***MicroSatellite Instability*** (MSI), that is characterized by small insertion and deletions in repetitive DNA tracts (microsatellite). Microsatellites are repeats of short nucleotide sequences distributed around the genome. The sequences can consist of one (mono), two (di) or up to six nucleotides. During DNA-replication, if the mismatch repair (MMR) system is malfunctioning, these mutations within microsatellite sequences result in genomic. The genes involved in maintaining the integrity of post-mitotic DNA are called MMR genes and include hMLH1, hMSH2, hMSH6, and PMS2. MSI is associated to a normal karyotype (euploidy or near-euploidy). MSI patients tend to have good prognosis compared to MSS patients [31, 49-50, 88, 101, 113, 120], but do not show an overall benefit from adjuvant

treatment with 5-FU (5-flurouracil), evidently a possible predictive marker for chemoresistance [31, 88, 95].]

2.7.2 KRAS

The V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) gene encodes a 21-kDa small protein that is activated transiently as a response to extracellular stimuli or signals such as growth factors, cytokines, and hormones via cell surface receptors [34, 74]. On its activation, the KRAS protein is also capable of turning off the signalling pathway by catalyzing hydrolysis of guanosine triphosphates (GTP) to guanosine diphosphates. KRAS mutations can be detected in approximately 30-40% of all patients with CRC and are associated with proliferation and decreased apoptosis. The most common KRAS mutations in codons 12 and 13 are activation mutations, leading to continuous activation of downstream pathways [34, 74].

To date, KRAS is not considered as a prognostic marker. There is no association between KRAS mutations and tumor location or stage, patients' geographic origin; rather, it has a predictive role. Multiple studies have shown that patients with KRAS mutations in codons 12 or 13 (Gly12Asp, Gly12Ala, Gly12Val, Gly12Ser, Gly12Arg, Gly12Cys and Gly13 Asp) do not benefit from anti-epidermal growth factor receptor (EGFR) therapy with cetuximab or panitumumab, two monoclonal antibodies that have clearly demonstrated efficacy in the treatment of metastatic CRC (mCRC). In contrast, about 40% of patients with metastatic colorectal cancer unresponsive to other therapies, and who lack a KRAS mutation, show a partial response with these agents.

These findings suggest that only patients without KRAS mutations should be eligible to receive these therapies [69].

3 Aim of Thesis

Comprehensive knowledge of the mutational events responsible for cancer is a critical foundation for future diagnostics, prognostics, and targeted therapeutics. With recent advances in genomic technology, researchers are trying to study large collections of tumors to characterize the alterations that have occurred in human genomes. DNA arrays containing probes for hundreds of thousands of genetic loci have made possible to detect regional amplifications and deletions with high resolution.

The aim of the present thesis was to determine the pattern of broad and focal chromosomal aberrations in colorectal cancer and to set up some novel bioinformatics tools useful in translating SNP array technology to clinical practice.

In order to achieve these objectives the frequency and type of the following genetic events were evaluated in a series of 51 colorectal samples:

- Microsatellite Instability by capillary electrophoresis
- Chromosomal Instability, broad and focal copy number abnormalities, Homozygous Deletion (*HoD*), High Level Amplification (*HLA*) and Copy Neutral-Loss Of Heterozigosity (*CN-LOH*) by SNP array 6.0

Moreover, two new tools were implemented in Python, in order to distinguish somatic and germ-line aberrations by analysis of matched tumor/normal mucosa couples and to evaluate the recurrency of such abnormalities in a series of patients. These bioinformatics tools have been also devised in order to extract relevant information from SNP array data and prepare reports used in the routine clinical setting.

4 Materials and methods

4.1 Patients

A total of 51 patients, underwent surgical resection for primary invasive colorectal cancer at the "*Casa Di Cura S.r.l. G.B. Morgagni*" in Catania, were studied. Patients' age ranged from 37 to 93 years, with an average (± SD) of 70.35 (± 14.48) years (median value, 73 years). In particular, the patients have been subdivided according to the sex and their characteristics are summarised in Table 2.

For each sample, two pieces of tumor were taken at distance proportional to dimension of tumor. One piece of normal colonic mucosa for each sample was taken at distance of 3 cm by tumor. Although, twenty-eight normal colonic mucosae of correspondent samples were analyzed. The CRC specimens were frozen and stored at -80°C before DNA extraction.

Genomic DNA from CRC samples was prepared in Catania, at the Complex Systems Laboratory, Scuola Superiore di Catania. Microsatellite instability test and KRAS Mutation Analysis were performed at the Complex Systems Laboratory, Scuola Superiore di Catania while Microarray experiments were performed at CIRES laboratory, Ragusa, where an Affymetrix instrument was available.

Sex	**N. patients**	**Age**	**Stage**			
			I	**II**	**III**	**IV**
F	22 (43.14%)	70.35 (± 14.48)	2 (3.92%)	6 (11.77%)	11 (21.57%)	3 (5.88%)
M	29 (56.86%)	70.86 (± 15.85)	1 (1.96%)	12 (23.53%)	11 (21.57%)	5 (9.80%)

Table 2
Characteristics of patients

4.2 Genomic DNA extraction.

Genomic DNA (gDNA) was extracted from tissue using the QIAamp DNA Mini Kit according to the manufacturer's instructions. The concentration and the quality of the DNA were determined using a ND-1000 spectrophotometer (NanoDrop, Thermo Scientific, USA).

The same sample was analysed with three different biological methods:

- High-resolution genome-wide DNA copy number and SNP genotyping analysis;
- Microsatellite Instability Test
- KRAS Mutation Analysis by direct sequencing with 310 Genetic Analyzer

4.3 High-resolution genome-wide DNA copy number and SNP genotyping analysis.

High-resolution genome-wide DNA copy number and SNP genotyping analysis was performed according to the protocol supplied by the manufacturer (Affymetrix, Inc., Santa Clara, CA, USA) for Affymetrix SNP 6.0 arrays. The protocol (scheme in Fig. 11) has been designed to improve the signal-to-noise ratio: DNA was first digested with restriction enzymes, then ligated to adapters and amplified. During the PCR amplication only the smaller restriction fragments (up to about 1.2 kbp, 200-1100bp size range) were amplified, reducing the complexity (but also the representation) of DNA.

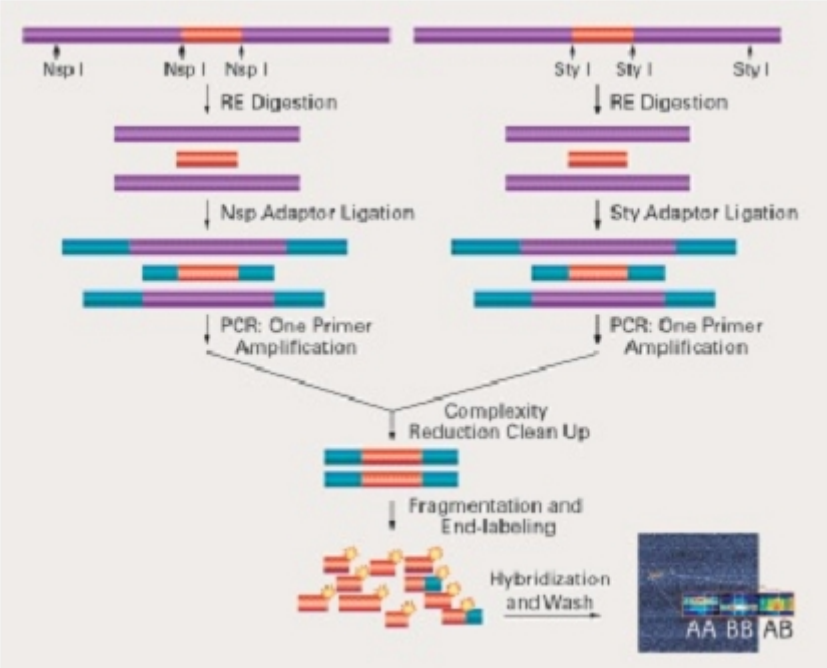

Figure 11
Genotyping Mapping Assay Overview: probes are chosen from restriction digestion fragments selected in order to reduce target complexity before labeling and hybridization

The main stages are:

1. *DNA extraction*: Genomic DNA was extracted using the DNA Blood Mini kit (Qiagen, Valencia, CA, USA) from mononuclear cells isolated from bone marrow aspirate samples by Ficoll gradient centrifugation. DNA quantity and quality were assessed using the NanoDrop 1000 Spectrophotometer (Thermo Scientific,) and for selected cases by agarose gel electrophoresis. The high quality DNA is a critical step, since PCR inhibitors, including high concentrations of heme (from blood) or chelating agents (i.e., EDTA) or salts, used to precipitate DNA during extraction, can interfere with the restriction enzymes of following steps;
2. *Digestion:* two aliquots of each DNA sample (5.00 uL at concentration of 50 ng/uL, totally 500 ng) were digested by two restriction enzymes, Nsp I and Sty I, capable of recognising

specific sequences (5' RCATG Y 3' and 5'C/CWWGG 3' respectively) to obtain fragments in the 200 to 1,100 bp size range, compatible with the distribution of SNPs on the array in the human genome. With this approach, it is possible to obtain a set of fragments, which guarantee a large coverage of the whole genome (70-80%), except telomers and centromers, characterized by a low SNPs rate;

3. *Ligation*: all fragments resulting from restriction enzyme digestion, regardless of size, were ligated to two annealed oligonucleotides (which act as adaptators), capable of recognizing the cohesive 4 bp overhangs derived from Nsp l or Sty l restriction site. The key enzyme is the T4-DNA ligase, typical of bacteria infected by T4 phagus, which uses ATP as cofactor and can ligate only sharp extremities. This step is needed for the next PCR, so that a generic primer could recognize the adaptor sequence to amplify adaptor-ligated DNA fragments;
4. *PCR*: was performed to select and amplify fragments, using a single primer with the TITANIUMTM DNA Amplication Kit (Clontech Laboratories, Inc. Mountain View, CA). The reaction conditions were optimized to preferentially amplify fragments in the 200 to 1,100 bp size range, which guarantee the genome coverage from 30% to 50%. The fragments size range was confirmed by an agarose gel running;
5. *Purification*: PCR products were purified with Agencourt AMPure Magnetic Beads (Agencourt Bioscience Corporation, Beverly, MA) and the amplicons were quantified using a NanoDrop-1000 spectrophotometer.

6. *Labeling*: fragmented PCR amplicons were end-labeled with a specific proprietary biotin-labelled reagent by a Terminal Deoxynucleotidyl Transferase for 4 hours and finally hybridized at 500C overnight (16-18 hours) in a GeneChip Hybridization Oven 640 (Affymetrix, Inc.);
7. *Washing*: chips were washed for several minutes in decreasing salt buffers, stained and dried in a GeneChip Fluidics Station 450 (Affymetrix, Inc.).

Arrays were scanned with a GeneChip Scanner 3000 7G (Affymetrix, Inc.) and a raw file was generated for each of them. Data coming from scansion was analysed using Affymetrix *Genotyping Console* (GTC) version 4.0, which implements the novel genotype calling algorithm Birdseed, which performs a multiple-chip analysis to estimate signal intensity for each allele of each SNP, fitting probe-specific effects to increase precision. Moreover, the reduced complexity of the hybridization brings with it the possibility of amplification bias of different regions of the genome and detection of changes reflecting differences in restriction digestion patterns between individuals rather than in true copy number.

The following algorithms were used: 1) *SNP 6.0 Birdseed v2* algorithm for genotyping, 2) *BRLMM-P-Plus* algorithm and Hidden Markov Model with regional GC correction for copy number analysis, 3) the *LOH algorithm*. As a quality control of the genotyping and copy number results "*Contrast QC value*" and "*Median Absolute Pairwise Difference*" (MAPD) were calculated as implemented in the GTC 4.0 software.

4.4 Microsatellite Instability Test

MSI analysis was performed on paired tumor-normal tissue DNA samples. DNA extracts (2 μl) were applied in the Multiplex-PCR approach according to the manufacturer's instructions (AB ANALITICA). Fluorescent dye-labeled PCR amplification was performed using the Bethesda panel of microsatellite markers (D2S123, D5S346, D17S250, BAT25 and BAT26).

This panel was extended with the additional markers BAT40, NR21, NR24, D18S58, TGFβRII, TPOX and TH01. these two last markers are used in order to confirm the perfect match between tumor sample and corresponding normal mucosa. Fluorescent dye-labeled and unlabelled primers were obtained; the 5' oligonucleotide was end-labeled with FAM (TGFβRII, NR24, D2S123, D5S346, D17S250, BAT26), HEX (BAT40, D18S58, BAT25, TH01, TPOX), or TAMRA (NR21) fluorescent dyes. Finally the microsatellite instability was analyzed on an *ABI PRISM 310 Genetic Analyzer* using *GeneScan Analysis Software* (Applied Biosystems Japan Ltd.).

International criteria for the determination of MSI in CRC were used to differentiate high instability (MSI-H) from low instability (MSI-L) or microsatellite stability. MSI-H tumors were defined as having instability in four or more markers of the 12 markers tested.

4.5 KRAS Mutation Analysis by direct sequencing with 310 Genetic Analyzer

Mutation analysis of KRAS at codon 12 and 13 has been performed on gDNA extracted from fresh tissue as previously reported. The PCR reaction has been carried out in a total mix reaction of 50 μl

containing 25 pmoles of each primer, PCR buffer 10x (Invitrogen), 10 μM dNTP (Invitrogen), MgSO4 50 mM (Invitrogen), Enhancer Solution 10x and 1.5 unit of Taq polymerase. PCR conditions are as follows: 1 cycle at 95°C for 5 minutes; 35 cycles at 95°C for 1 minute, 60°C for 2 minutes, and 72°C for 3 minutes; followed by 1 cycle at 72°C for 5 minutes. The PCR products have been purified using an HiYIELD *Gel/PCR DNA Fragment Extraction Kit* (Real Genomics). The PCR products have been direct sequenced with *Big Dye V1.1 Terminator Kit* (Applied Biosystems, Foster City, CA, USA) and, after purification with the *Centri-Sep Spin Columns* (Applied Biosystems), they have been analyzed by capillary electrophoresis (ABI PRISM® 310 Genetic Analyzer, Applied Biosystems, Foster City, CA, USA).

5 Results

5.1 Microsatellite instability status

Fragment analysis by capillary electrophoresis of the ten MSI-markers showed that 13.73% (7/51) of the tumors had unstable microsatellites (MSI-H). One specimen was MSI-L with only one mutated marker and was included among MSS. The remaining MSS tumors (n = 43) revealed no aberrant pattern in any of the markers. All corresponding normal mucosa showed characteristic wild type patterns. Patients' characteristics are summarised in Table 3. Patients have been subdivided according to the microsatellite instability in MSS and MSI.

	N. patients (%)	**Age**		**Sex**	
		F	**M**	**F**	**M**
MSS	44 (86.27)	70.95 $\pm$ 17	71.35 $\pm$ 12	21	23
MSI	7 (13.73)	69	64.67 $\pm$ 20.27	1	6

Table 3
Patients subdivided according to the microsatellite instability test (MSS versus MSI)

Figure 12 shows the differences between wild type pattern and aberrant pattern for the different MSI markers. In MSI tumors there are two regions of amplification in microsatellite markers: one is the amplification of non-mutated allele and the other is of the unstable allele (this is indicated with a red arrow in figure); while for stability there is only one region of amplification of non-mutated allele.

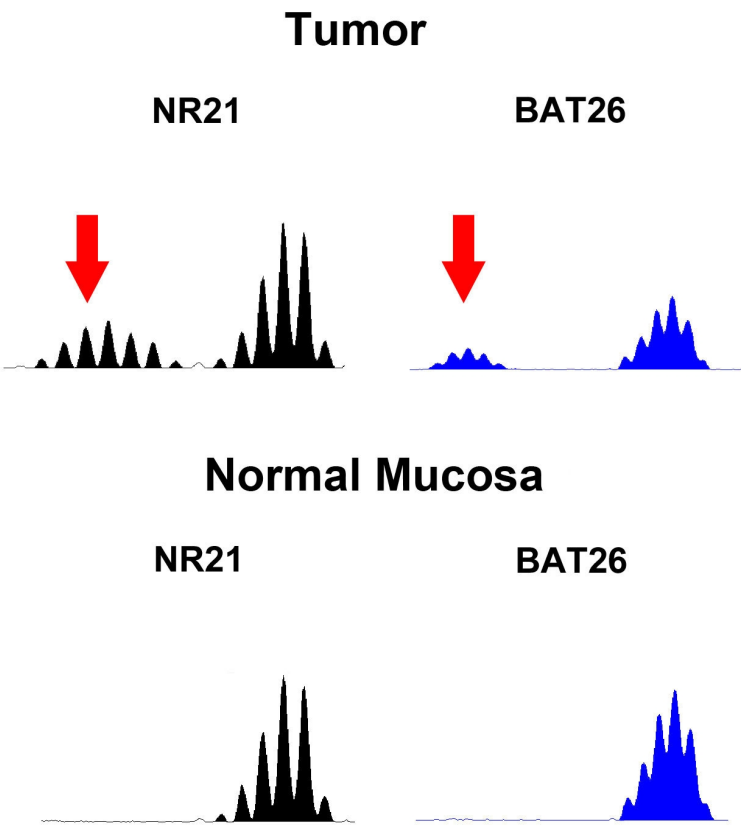

Figure 12
Electropherograms of two MSI markers: NR21 and BAT26.
The top electropherograms are from tumor showing aberrant MSI-pattern, while the lower ones are obtained from normal mucosa and illustrates the characteristic wild type pattern in the microsatellites. The tumor and mucosa illustrated in each marker are from the same patient.

5.2 KRAS mutation detection

The detection of KRAS mutation has been performed by direct sequencing.

The results are summarized in table 4. KRAS mutation has been detected in 39.22% of patients in accordance with literature where KRAS mutation frequency ranges between 30% and 40% [69]. Moreover, a higher frequency of KRAS mutations is observed in females: 10/21 female MSS (47.62%) and 1/1 female MSI versus 8/23

male MSS (34.78%) and 1/6 male MSI. In particular, KRAS mutations have been detected more frequently on codon 12 (Gly12Val).

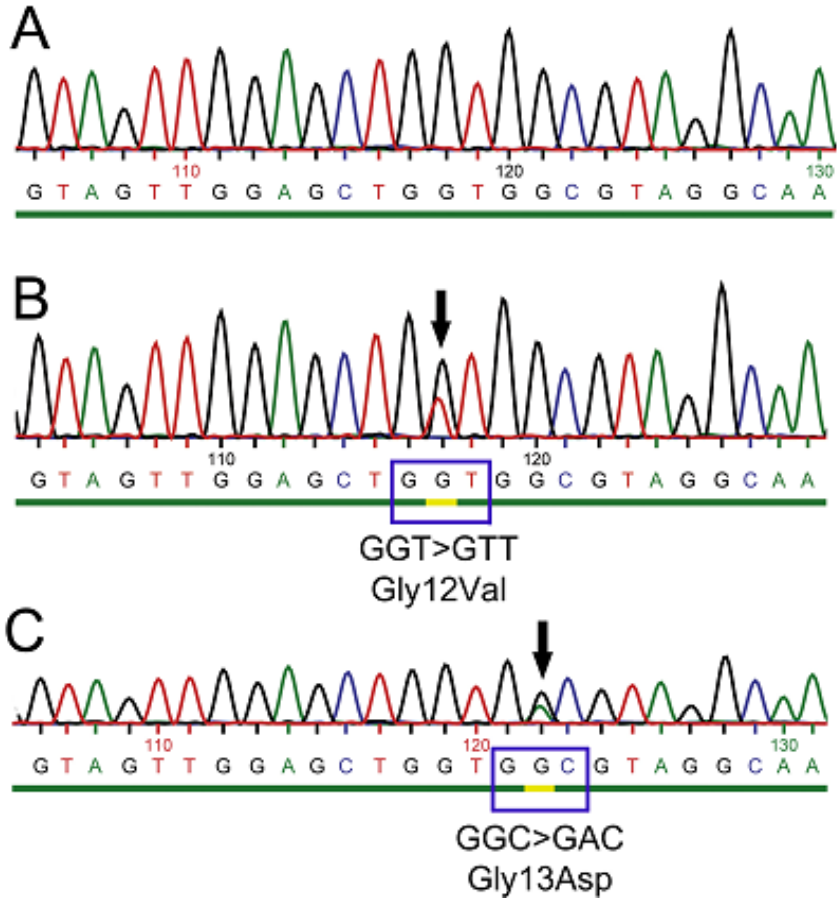

Figure 13
Sequences of codons 12 and 13 of KRAS gene.
In A a sequence without mutation is shown, in B the sequence has a mutation G/T in codon 12, G<u>G</u>T>G<u>T</u>T. It causes the replacement of glycine with valine. In C the sequence has a mutation G/A in codon 13, G<u>G</u>C>G<u>A</u>C. It causes the substitution of glycine with aspartic acid.

In addition, the mutated KRAS has been found more frequently in stage III and IV CRC as shown Fig. 14, while wild-type KRAS tumors are found more frequently in stage II.

		Tot.	MSS			MSI		
			Tot.	F	M	Tot.	F	M
N. patients with KRAS Mutation (%)	Wild Type	31 (60.78%)	26 (59.09%)	11	15	5	0	5
	Gly12Ala	1 (1.96%)	1 (2.27%)	0	1	0	0	0
	Gly12Asp	4 (7.84%)	3 (6.82%)	1	2	1	0	1
	Gly12Arg	0	0	0	0	0	0	0
	Gly12Cys	1 (1.96%)	1 (2.27%)	1	0	0	0	0
	Gly12Ser	1 (1.96%)	1 (2.27%)	0	1	0	0	0
	Gly12Val	8 (15.69)	7 (15.91%)	4	3	1	1	0
	Gly13Asp	5 (9.80)	5 (11.36%)	4	1	0	0	0

Table 4
KRAS mutation detected by direct sequencing in 51 CRC patients

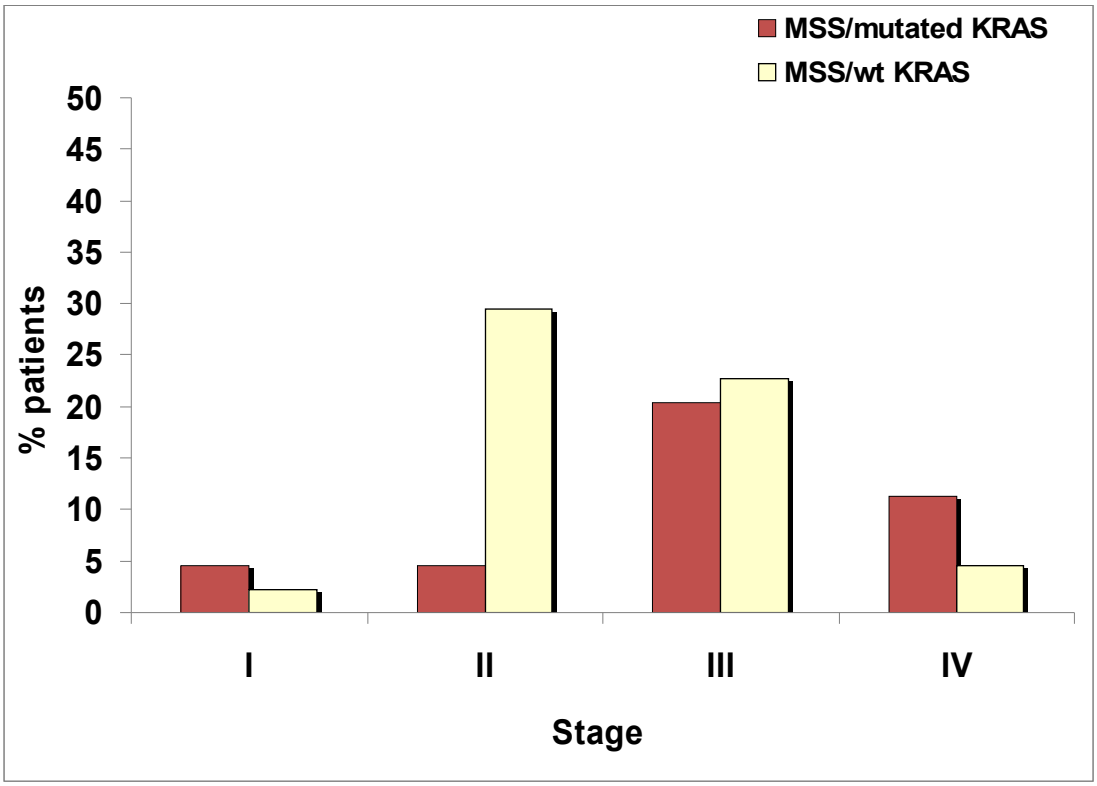

Figure 14
Frequency of KRAS mutation at each stage according to the MSS status

5.3 Implementation of bioinformatics tools

In this thesis, two bioinformatic tools have been implemented for identifying regions of aberration that are more likely to be involved in CRC.

5.3.1 Broad Cytogenetic Analysis: tool to identify broad aberrations

Broad Cytogenetic Analysis (*BroCyA*) was implemented to identify and quantify broad chromosomal aberrations (structural aberrations that involve more than 25% of a chromosomal arms or numerical aberrations involving whole chromosomes).

BroCyA tool, in order to define segments as gain or loss, uses Log_2ratio value calculated by Affymetrix *Genotyping Console* (*GTC*). This value is the ratio between signal for each marker in each sample and the corresponding median value in a reference group (270 HapMap individuals) and provides an estimate of copy number. Log_2Ratio of each chromosome is calibrated to a copy number value equal 2 in a diploid genotype.

5.3.1.1 Thresholds

In order to define the gain and loss thresholds, 13 samples were analyzed: 11 remission of Normal Karyotype Acute Myeloid Leukemia (NK-AML) patients and 2 normal samples. For each sample the average Log_2Ratio was calculated for each chromosome. Both in remission and normal samples the average was 1.98 ($\pm$ 0.12). On the basis of this, all values greater than 1.98+2SD (2.21) have been considered as gains and less than 1.98-2SD (1.74) as losses.

In SNP Array 6.0 there is a median inter-marker distance of 696 base pairs (bp). In order to find broad aberrations, it was required to consider all segments formed by at least 50 contiguous markers, obtaining segments larger than 0.035 Mb. These short segments have been joined in larger segments according to their distance measured in number of intersegment markers. If the intersegment distance was lower than an established threshold the short segments (formed by at least contiguous 50 markers showing the same type of alteration) were joined. In order to determine at what threshold value the short segments can be joined, the number of CNAs per sample was calculated with different threshold values: from 100 to 1500 intersegment markers. As shown in Fig. 15 a plateau value is reached with intersegmental distances higher than 700 markers, and such value was chosen as threshold value for the analysis.

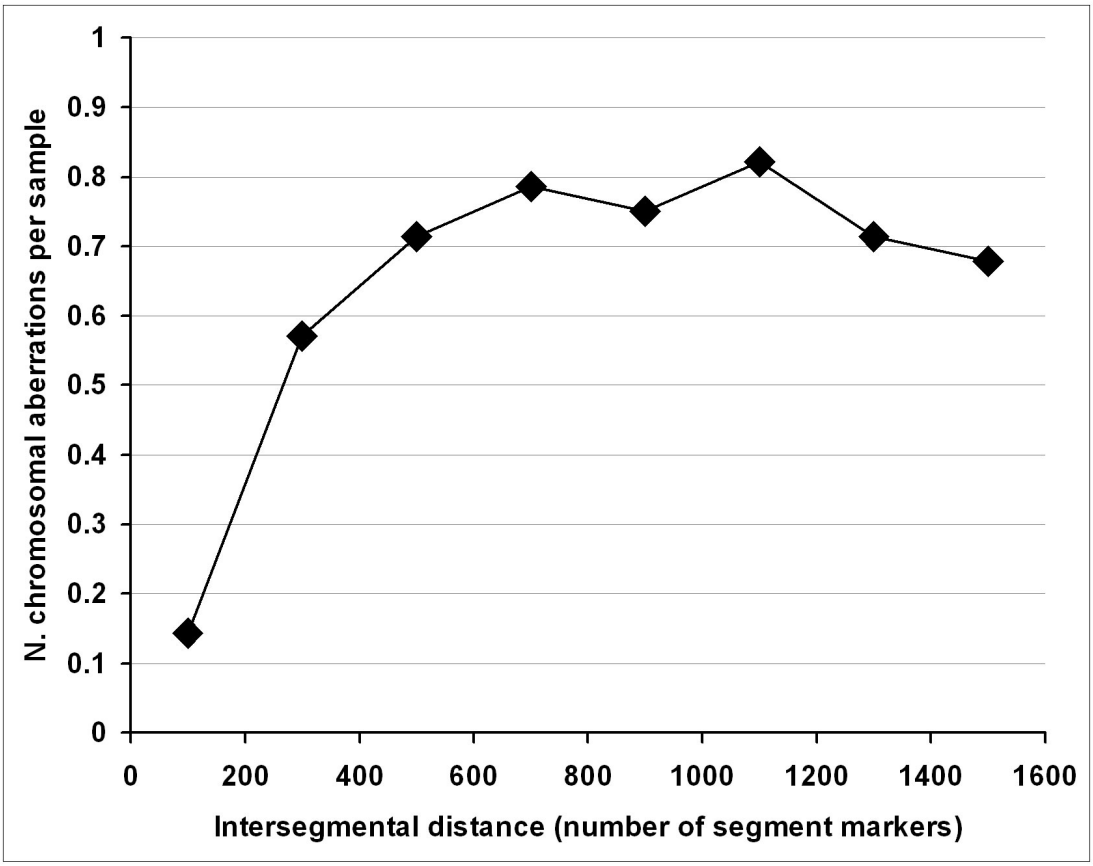

Figure 15
Number of CNAs per sample calculated with different threshold values

5.3.1.2 Characteristics of bioinformatics tools

BroCyA is characterized by two levels (Fig. 16-17). In a first step *BroCyA* generates large aberrations: all markers with values less than 1.74 are joined as loss segments, while the markers with value greater than 2.21 are joined as gain segments (Fig. 18). Then all segments composed by at least 50 consecutive markers are considered (*short segments*). Adjacent short segments, with intersegment distance lower than 700 markers, are joined in a single large segment. In a second step *BroCyA* calculates the mean Log_2Ratio of large segments obtained in the previous step and removes all gain large segments with mean values less than 2.21 and loss large segments with mean values higher than 1.74. All large segments that respond to criteria of this second step of *BroCyA* and

with a physical size (measured in base pairs) higher than 25% of a chromosomal arm are reported in a list (Fig. 19) Finally, in this step *BroCyA* creates a table with frequency of chromosomal aberrations. If there are more aberrations of the same type, (gains or losses) in the same chromosome, in the second step of the algorithm the size of each aberration (loss or gain) is summed up separately for chromosomal arms and expressed as percentage of the p and q arm involved in each type of aberration. The results are reported in a table according to the following rules:

- if the sum of the percentage of the p arm and the q arm are greater than 70% of the total size of the chromosome the aberration is reported as “whole chromosome”;
- if the sum of the percentage of the p arm and the q arm are less than 70% one aberration is reported for each arm.

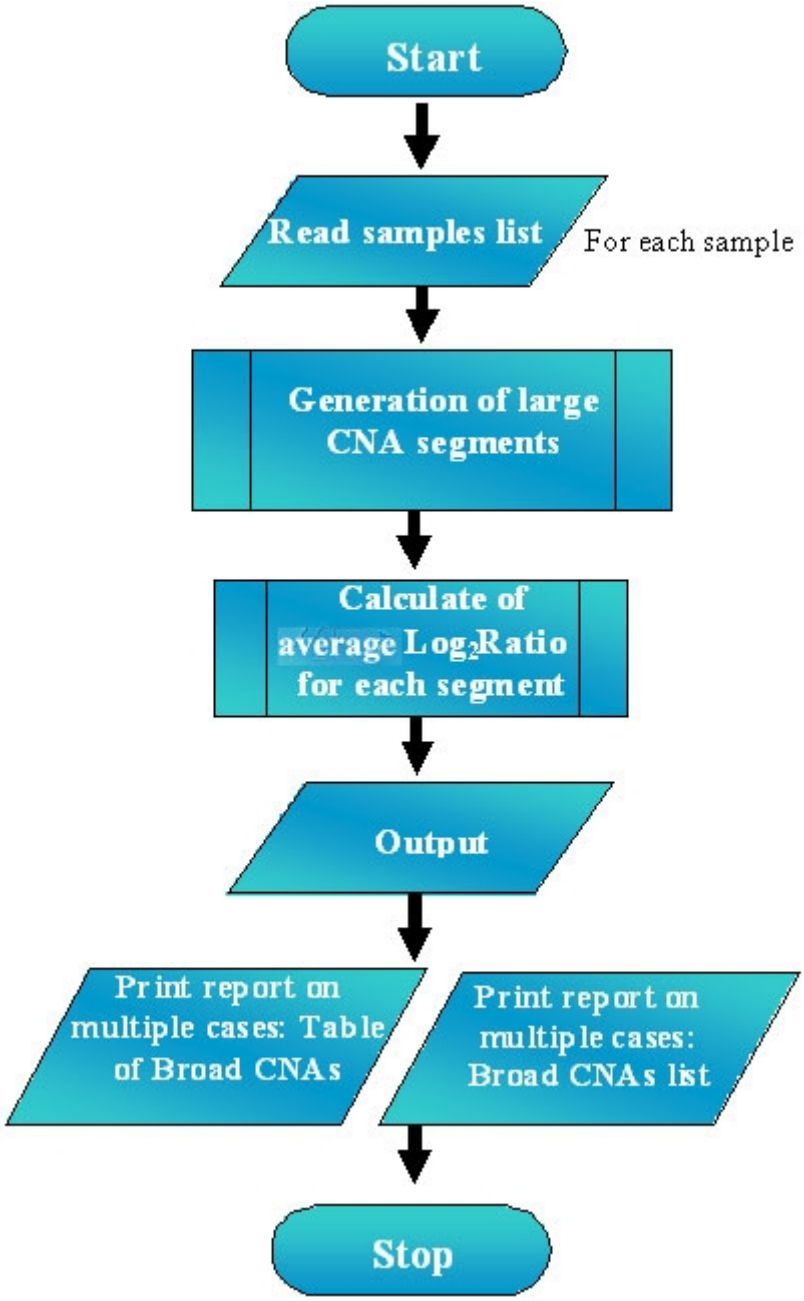

Figure 16
Flowchart of *BroCyA* tool: it identifies aberrations with size higher than 25% of the chromosomal arm. In this case a list of sample is considered as input and the output is: 1) a summary table with all broad aberrations detected in the samples; 2) a list of broad aberrations for each sample **average Log_2Ratio)**

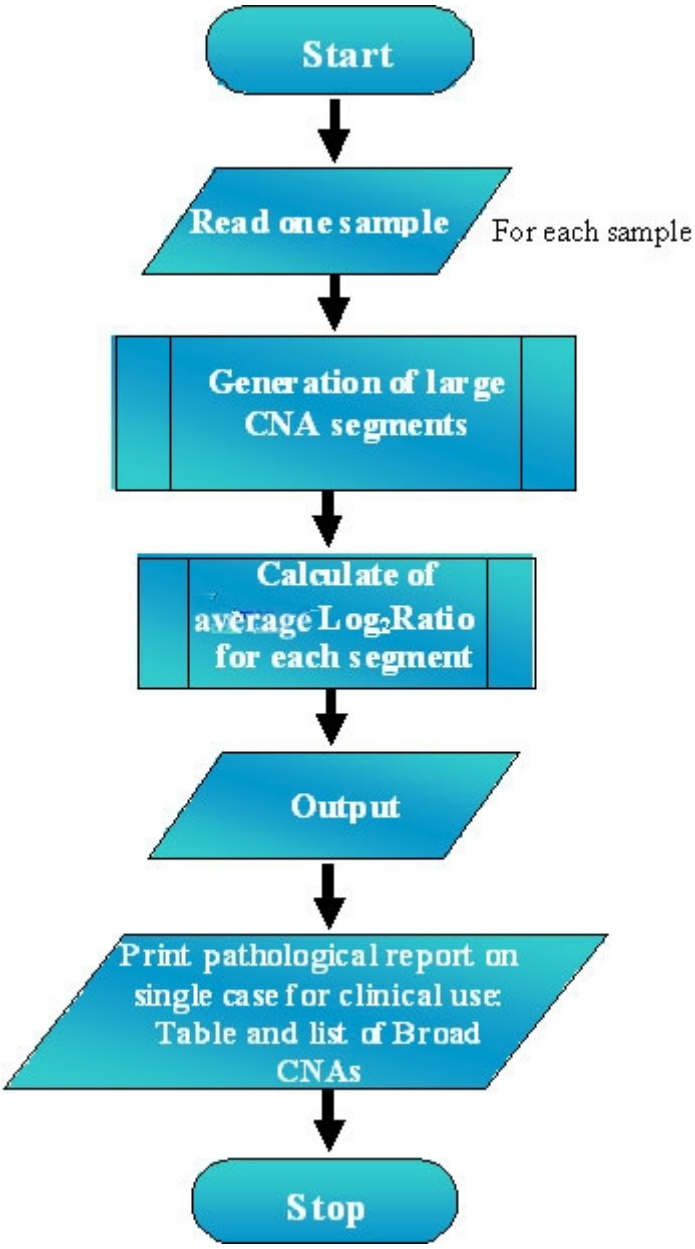

Figure 17
Flowchart of *BroCyA* tool: it identifies aberrations with size higher than 25% of the relative arm. In this case one sample is considered as input and the output is a report on its cytogenetics

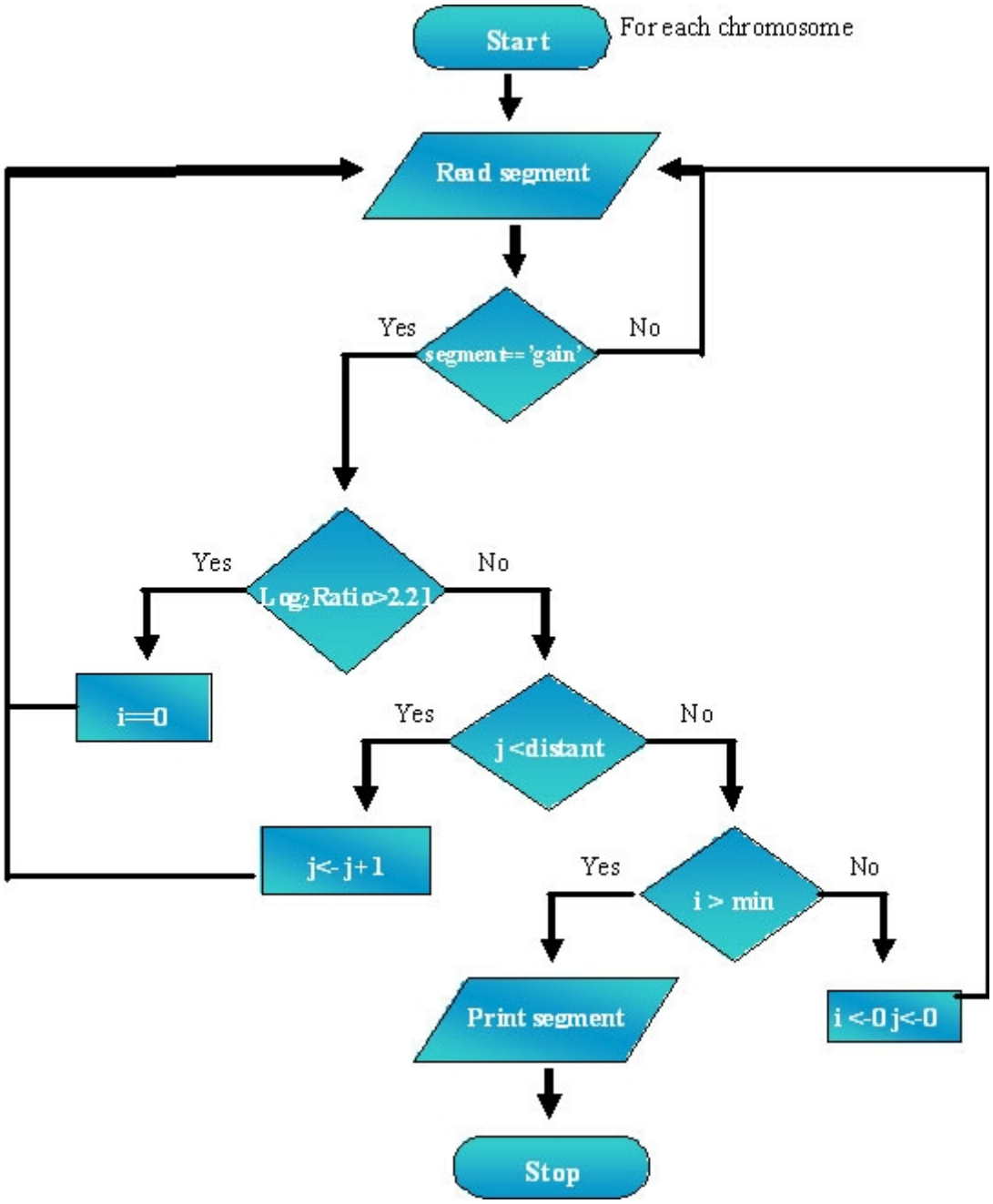

Figure 18
Flowchart of tool to generate large CNA segments. The same process is used to generate broad loss segments. In that case the threshold is equal to 1.74

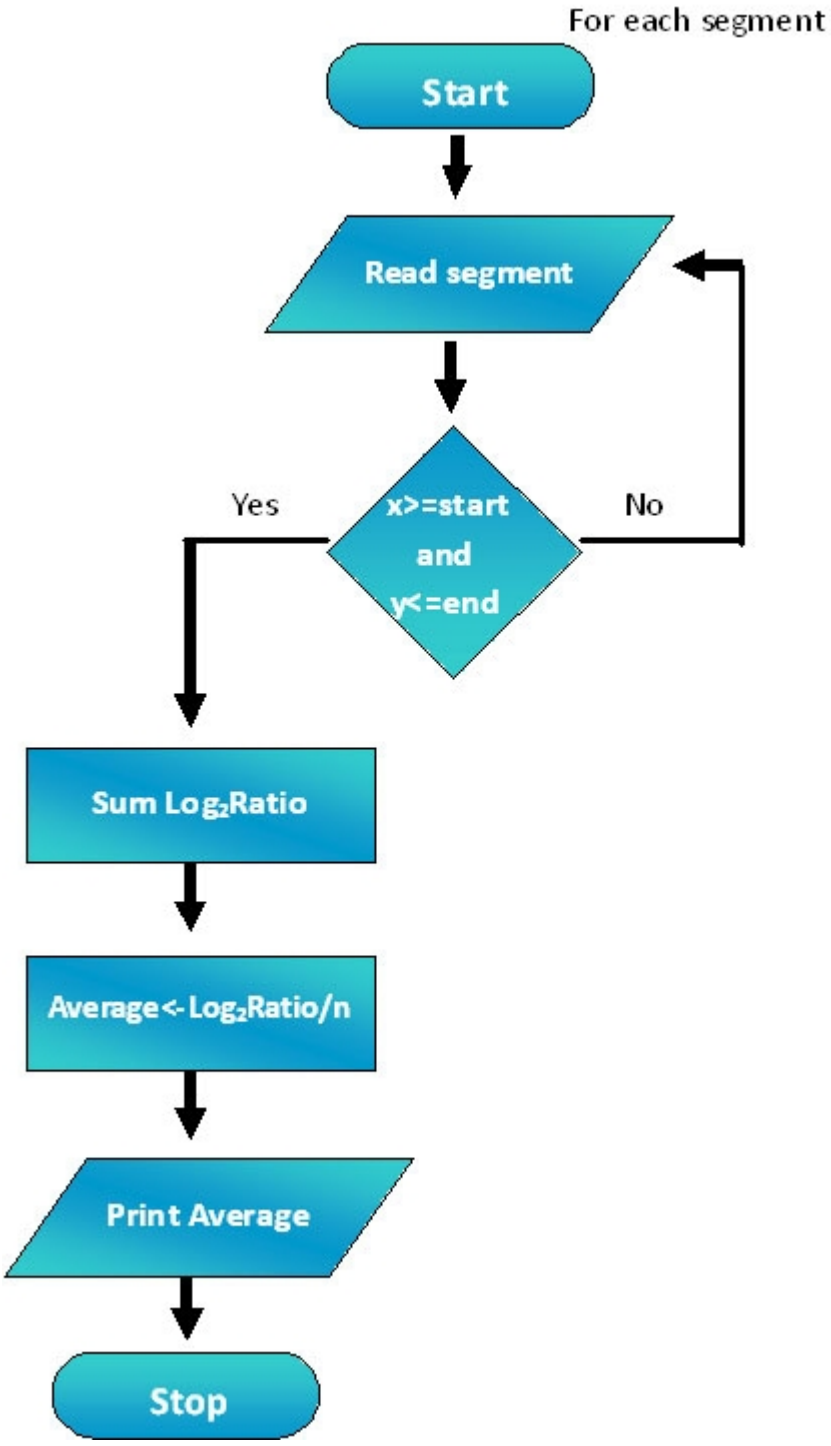

Figure 19
Flowchart of script to calculate mean Log_2Ratio for each large segment found in the previous step of *BroCyA* tool

5.3.1.3 Broad CNAs

BroCyA has been used to analyze the SNP-array results obtained from 51 CRC samples (7 MSI and 44 MSS tumors) and 29 normal coupled mucosae.

5.3.1.4 Number of Broad CNAs

The frequency distribution of chromosomal aberrations in the sample population is reported in Fig. 20. This graph demonstrates the percentage of samples (ordinate) showing a chromosomal aberrations number within the intervals reported in abscissa (ranging from 0 to 21-22). The following groups have been compared: normal mucosae samples (Mu), microsatellite instability tumor samples (MSI) and microsatellite stability tumor samples (MSS) (Fig. 20). All normal mucosae and around 75% of MSI tumors show a very low number of broad chromosomal aberrations (from 0 to 3-4 aberrations) and have been identified as negative to chromosomal instability (CIN-). A low percentage of MSI tumors (25%) are characterized by 5-6 chromosomal aberrations (grey zone in the graph of Fig. 22 also indicated as CIN+/-), suggesting that a moderate form of chromosomal instability can be observed also in this group of tumors.

The majority of MSS tumors (82% of MSS tumors) have from 7 to 22 aberrations (CIN+). A low percentage of MSS tumors (14% of MSS tumors) show a low number of broad chromosomal aberrations, comparable to those observed in normal mucosae (0-4 aberrations), and has been classified as CIN-. Only 2 MSS tumors (4%) show an intermediate values of 5-6 broad aberrations (grey zone in the graph of Fig. 20, also indicated as CIN+/-). In conclusion four different groups

can be identified: MSS/CIN+ (70% of total samples), MSS/CIN- (12%), MSS/CIN+/- (4%) MSI, CIN- (10% of total samples), MSI/CIN+/- (4% of total samples). Therefore in the majority of cases chromosomal instability and microsatellite instability are mutually exclusive. Fig. 21 and 22 show SNP array data (in the form of a virtual karyogram) and microsatellite test results for representative examples of a MSI/CIN-, tumor (Fig. 21) and MSS/CIN+ (Fig. 22).

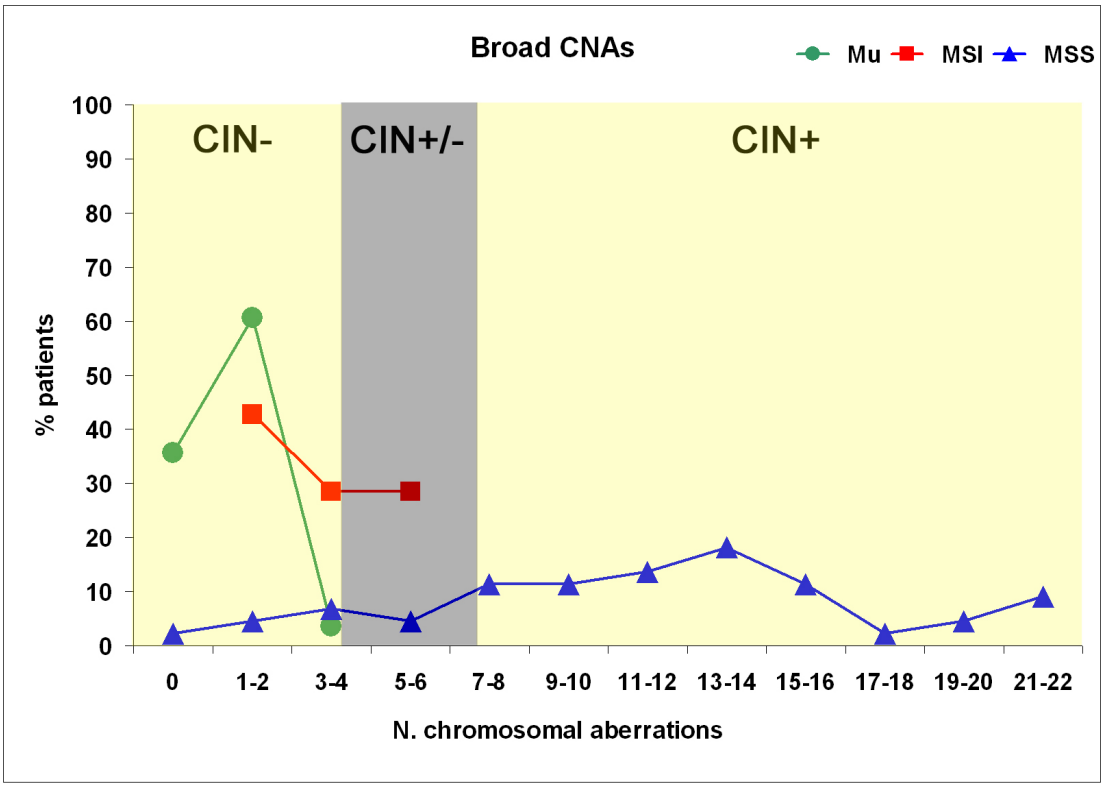

Figure 20
Frequency distribution of chromosomal aberrations in the sample population. Data obtained in normal mucosae samples (Mu), in microsatellite instability tumor samples (MSI) and in microsatellite stability tumor samples (MSS) are shown.

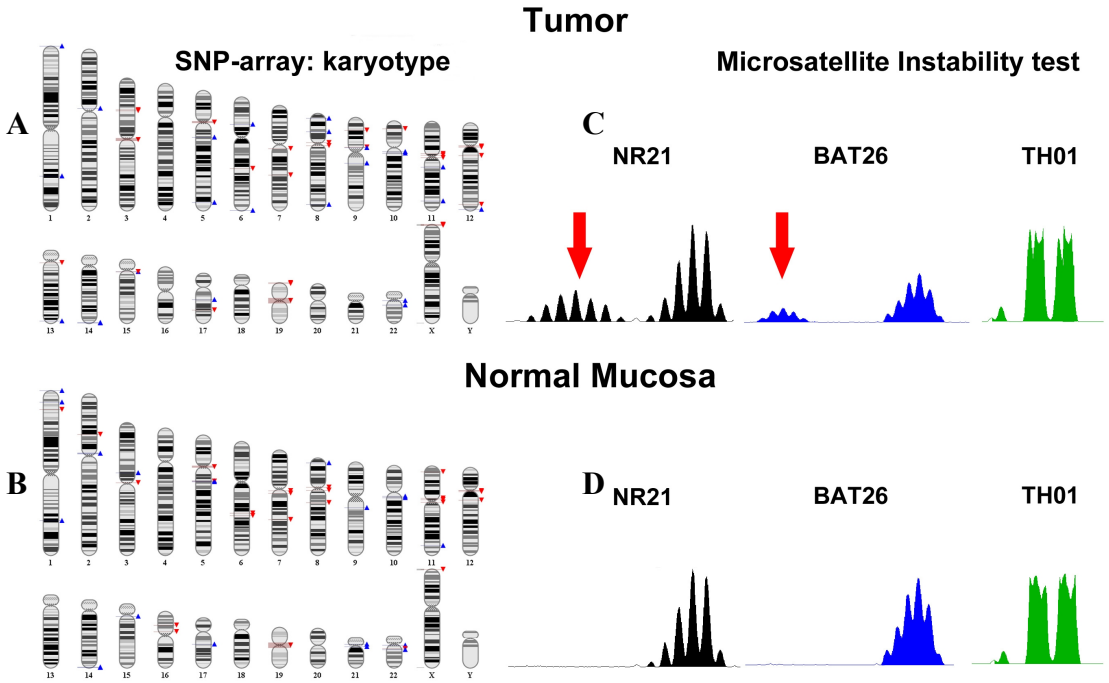

Figure 21
Molecular Cytogenetic through SNP-arrays and microsatellite analysis in CRC characterized by MicroSatellite Instability (MSI).
A and B show the molecular karyotype obtained by SNP-array in tumor and normal mucosae of the same CRC sample. Both karyotypes show no chromosomal instability; blue triangles indicate small amplifications, while red triangles represent small deletions.
The figures in C and D are three microsatellite markers analyzed, NR21, BAT26 and TH01. The instability analysis was done on CRC tumor (C) and on the normal mucosae of the same patient (D). In CRC tumor, for NR21 AND BAT26 microsatellite markers, there are two regions of amplification: one is the amplification of non-mutated allele and the other is that of the unstable allele (this is indicated with a red arrow); while for TH01 microsatellite markers there is a region of amplification.

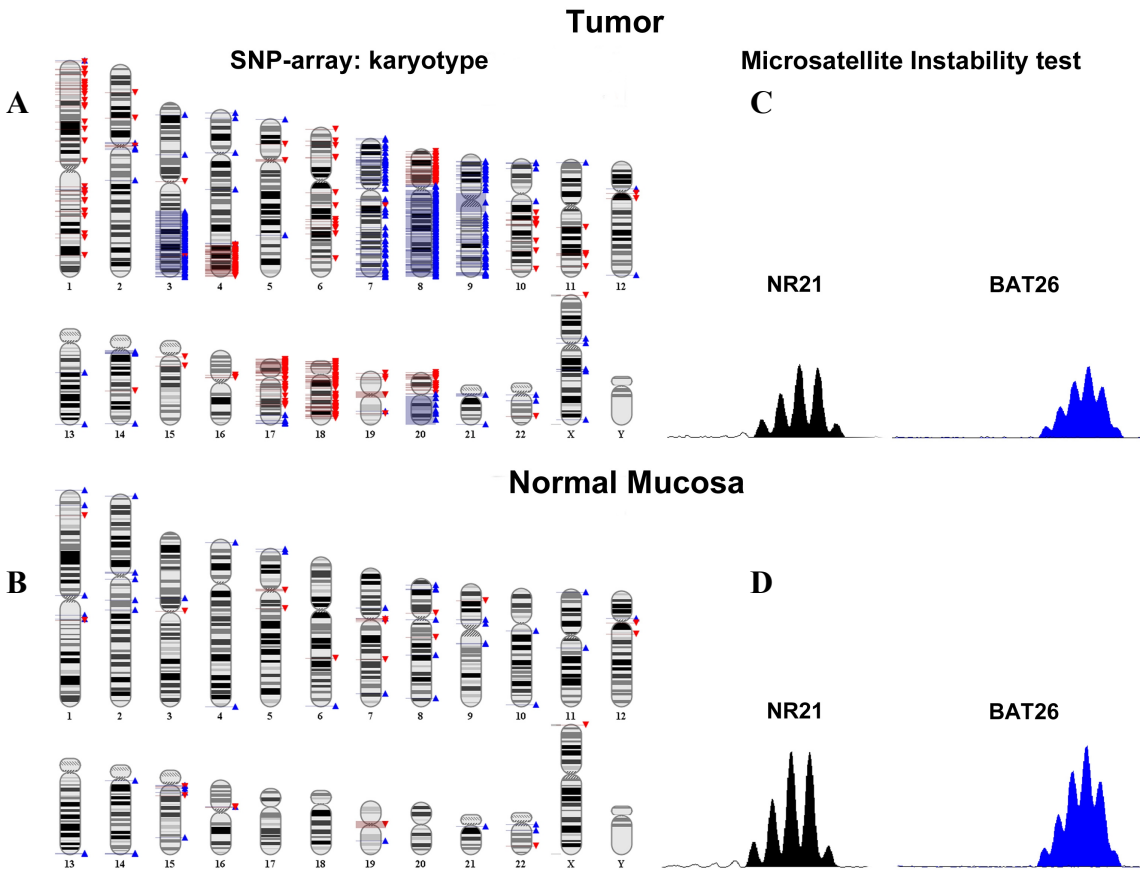

Figure 22
Molecular Cytogenetic through SNP-arrays and microsatellite analysis in CRC characterized by Chromosomal Instability (CIN).
A and B show the molecular karyotype obtained by SNP-array in tumor and normal mucosae of the same CRC sample. Both karyotypes show an evident chromosomal instability, characterized by deletions (red triangles) of half the arm of the 4q chromosome, arms 8p, 17p, 18p, 18q and 20p. The amplifications, showed in blue triangles, are present in whole chromosome 7, 8q, whole 9 and 20q.The microsatellite analysis of NR21 and BAT26, performed on tumor and normal mucosa of same sample, highlights stability.

5.3.1.5 Frequency of Broad CNAs

The frequency of broad chromosomal aberrations in MSS and MSI tumors have been reported in Fig. 23-24. *BroCyA* has revealed frequent (>10%) broad gains on chromosomes 7, 8q, 9, 13, 16, 17q, 20 and X and frequent (>10%) broad losses on chromosomes 1p, 4, 5q, 8p, 17p, 18, 20p.

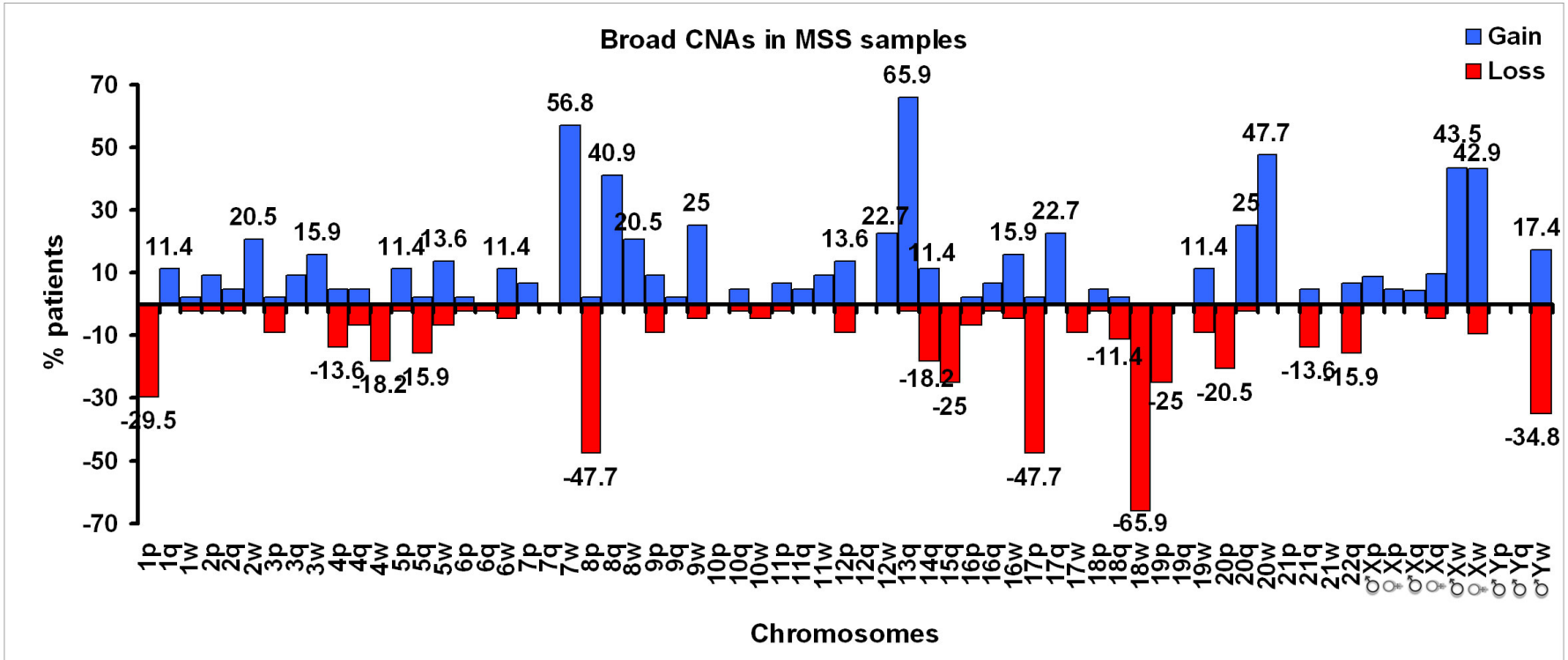

Figure 23
Percentage of MSS tumors bearing specific broad chromosomal aberrations (broad gains in blue; broad losses in red).

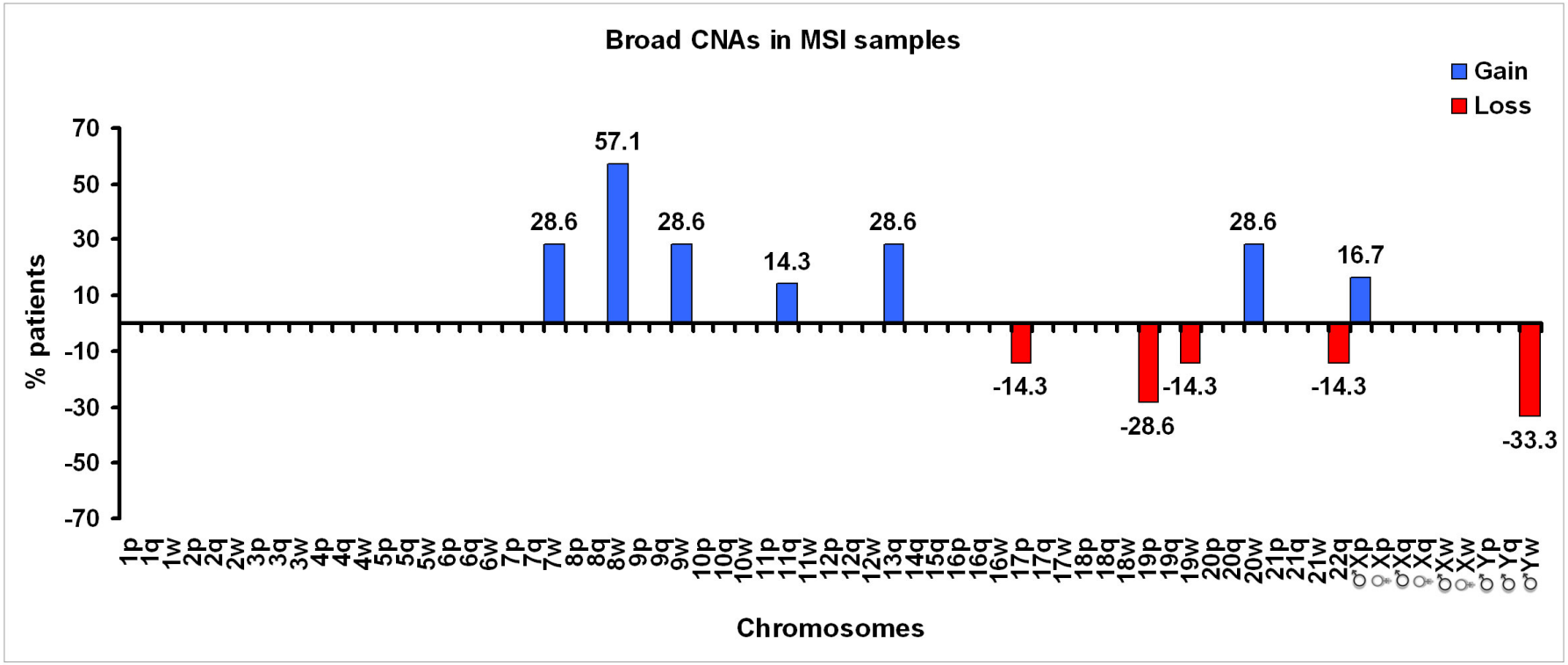

Figure 24
Percentage of MSI tumors bearing specific broad chromosomal aberrations (broad gains in blue; broad losses in red)..

5.3.1.6 Intratumor heterogeneity by double-sampling data

The correlation between data obtained by *BroCyA* tools in two samples distant at least 1 cm (average 1.8 cm) in the same tumor (double-sampling pairs) has been calculated.

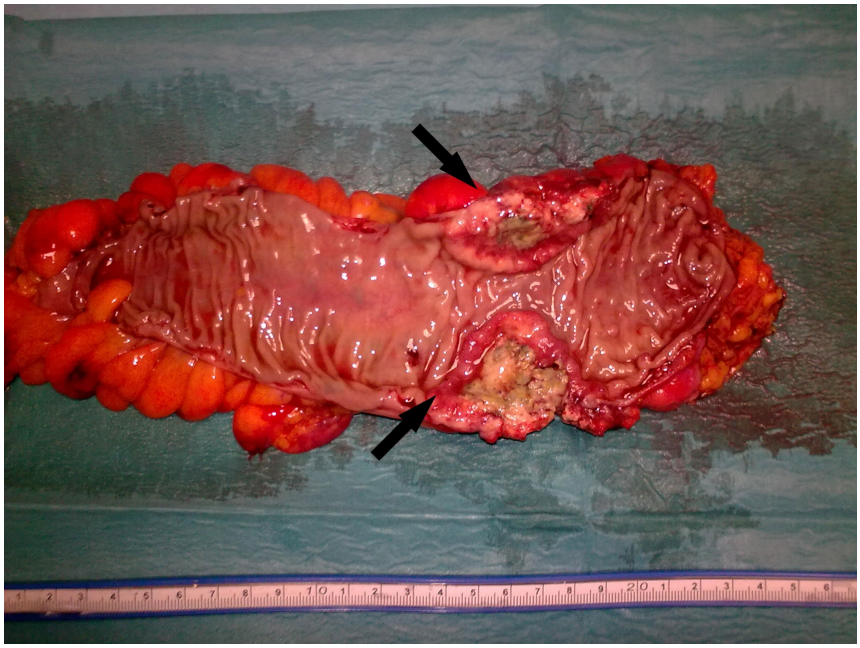

Figure 25
Example of double-sampling pairs. The points where double-sampling pairs have been taken have been indicated with black arrows

Fifteen double-sampling pairs have been considered and the number of aberrations have been calculated. Five samples are from normal karyotype tumors and 10 are from CIN+ tumors. Pearson's coefficient has been calculated within pairs and between pairs, see table 5.

Mean of correlation coefficient	
Within pairs	0.70
Between pairs	0.31

Table 5

5.3.1.7 HLA and HoD

The *BroCyA* analysis was extended using a different configuration. This approach uncovered additional *High Level Amplifications* (HLAs) and *Homozygous Deletions* (HoD).

HLAs were defined the amplifications where the copy number was greater than 5.2, and HoD the deletions with a copy number value less than 0.77. The value of HLA was arbitrarily chosen while the HoD value was assigned by calculating the average Log2Ratio of chromosome Y in 11 remission samples of NK-AML patients and 2 normal samples and considering Log2Ratio -2SD (1.03 - 0.22). In order to reduce false positive segments, HLAs and HoDs found in NK-AML remission samples and in the two normal samples were removed.

In Fig. 26 the distribution of HLAs in chromosomes of MSS tumors is reported expressed as the percentage of patients that show HLAs in a specific chromosome. A high number of HLAs were observed in chromosome 20. No HLAs were detected in MSI tumors and in 29 normal mucosae samples. HoDs were identified both in tumors and in normal mucosae samples. HoDs in normal mucosae represent germ-line variations, while somatic and germ-line HoDs can be distinguished in tumor samples. These results have been summarized in Fig. 27.

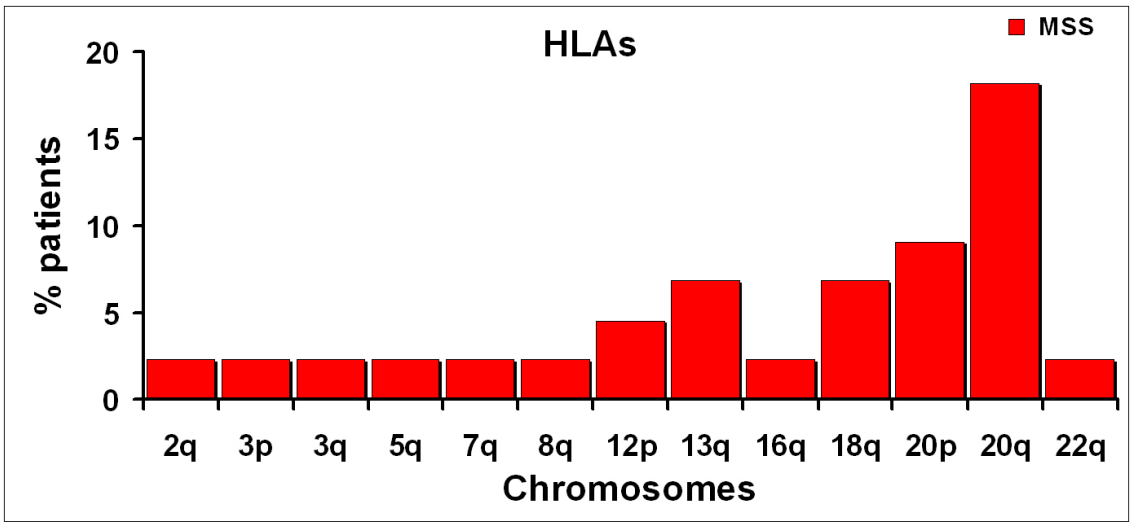

Figure 26
Frequency of HLAs in MSS status

Furthermore, recurrent HLA and HoD segments were determined in MSS and MSI tumor/mucosa couples, and the somatic or germ-line origin of these aberrations was established by comparison between the tumor and the corresponding normal mucosa. In addition, segment with copy number equal to 1 ($0.82<Log_2Ratio<1.74$), called *Heterozygous Deletions* (HeD) have been also detected.

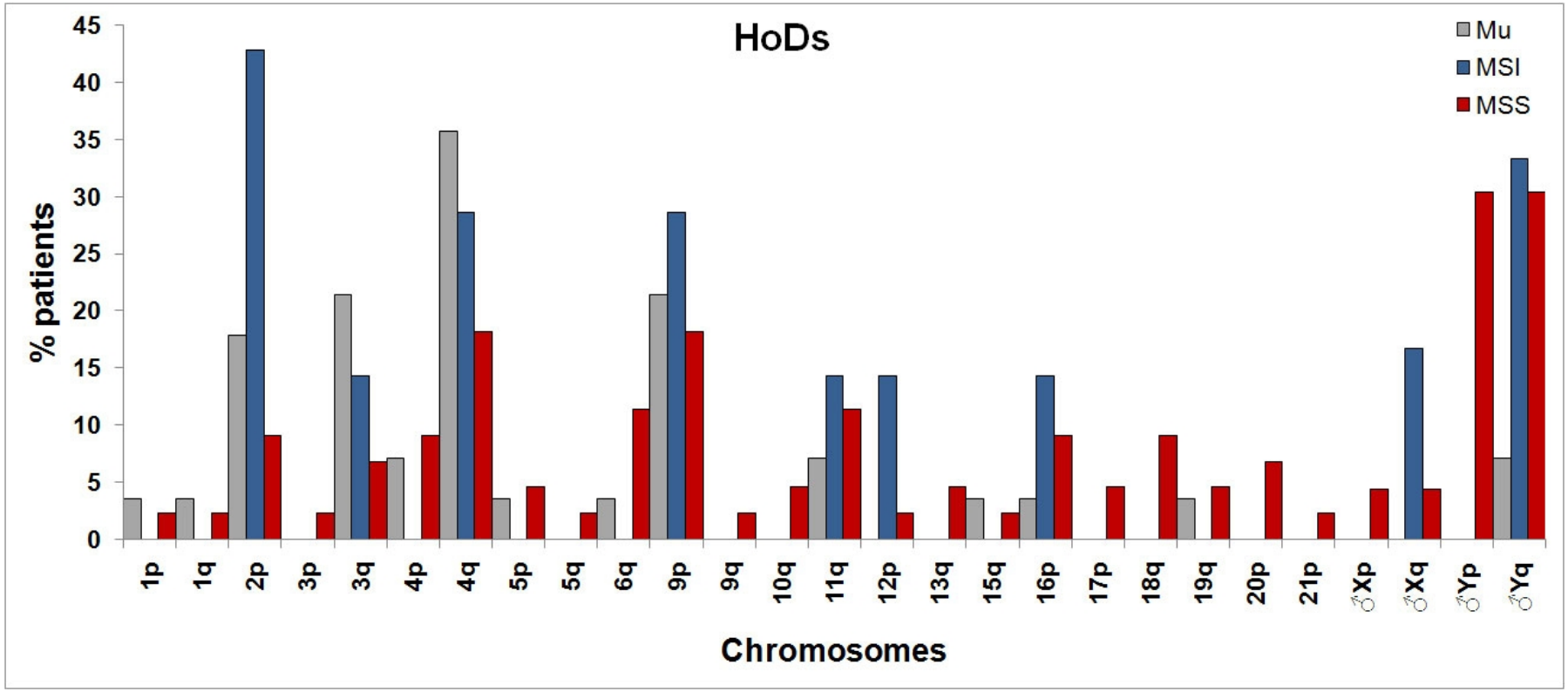

Figure 27
Frequency of HoDs in MSS, MSI, Mu status

Each HLA and HoD segment contains interesting candidate cancer genes, which can have a driver role in CRC cancer. For this reason, using the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), created by Genome Bioinformatics Group of UC Santa Cruz, I have searched for genes contained in the various copy number abnormal regions, considering as source the February 2009 human reference sequence (GRCh37) produced by the Genome Reference Consortium.

Chr	Start (kb)	End (kb)	Size (kb)	N. Samples	Somatic	Genes contained in the Genomic Region
20	58468	58492	24	3	3	SYCP2
20	34846	34893	47	3	3	C20orf4
20	43637	43681	44	3	3	STK4
20	43742	43820	78	3	3	WFDC5 – WFDC12 – PI3
20	22211	22348	137	2	2	No coding regions
20	29836	29852	16	2	2	DEFB115
20	29996	30205	209	2	2	DEFB123 – DEFB124 – REM1 - HM13 - ID1
20	42963	43146	183	2	2	R3HDML - HNF4 – HNF4A – C20orf62 – TTPAL - SERINC3
20	51429	51933	504	2	2	TSHZ2
20	56109	56260	151	2	2	PCK1– PMEPA1
20	56520	56671	151	2	2	No coding regions
20	56745	56971	226	2	2	RAB22A- VAPB

Table 6
Somatic HLAs ($Log_2R>5.2$) in MSS tumors (24 matched tumor/normal mucosa couples)

Chr	Start (kb)	End (kb)	Size (kb)	N. Samples		Somatic			Germline		Genes contained in the Genomic Region
						HoD		HeD	HoD	HeD	
				HoD	HeD	From Wt	From HeD				
2	34711	34722	11	3	0	0	0	0	3	0	No coding regions
4	69392	69506	114	2	0	0	0	0	2	0	UGT2B15
3	98930	98966	36	1	0	0	0	0	1	0	No coding regions
8	39247	39376	129	1	1	0	0	0	1	1	tMDC – tMDC III
11	110980	111117	137	1	0	1	0	0	0	0	No coding regions
11	111307	111533	226	1	0	1	0	0	0	0	BTG4 – C11orf88 – LAYN – SIK2

Table 7
Somatic HoDs in MSI tumors (5 matched tumor/normal mucosa couples)

Chr	Start (kb)	End (kb)	Size (kb)	N. Samples		Somatic			Germline		Genes contained in the Genomic Region
						HoD		HeD	HoD	HeD	
				HoD	HeD	From Wt	From HeD				
8	39257	39343	86	7	5	0	3	3	4	2	tMDCII, tMDC
9	44728	44821	93	6	4	0	1	0	5	4	No coding regions
4	69372	69506	134	4	4	0	0	4	4	0	UGT2B15
4	70149	70211	62	4	6	0	0	6	4	0	UGT2B28 – UGT2B11
20	14870	14975	105	3	7	1	2	7	0	0	MACROD2
2	52765	52791	26	2	0	0	0	0	2	0	No coding regions
6	162751	162792	41	2	1	2	0	1	0	0	PARK2
11	55362	55454	92	2	1	0	1	0	1	1	OR4P4 – OR4S2 – OR4C6
19	43661	43731	70	2	3	0	1	0	1	3	PSG4 – PSG - PSG5
1	103419	103816	397	1	2	1	0	2	0	0	COL11A1
1	248698	248819	121	1	0	0	0	0	1	0	OR2T29 – OR2T34 – OR2T10 – OR2T11
2	77917	78024	107	1	0	1	0	0	0	0	No coding regions
3	98917	98976	59	1	1	0	1	0	0	1	No coding regions
3	162540	162588	48	1	0	0	0	0	1	0	No coding

											regions
4	4250	4393	143	1	10	0	1	7	0	3	LYAR – ZNF509 – D4S234E – NSG1
4	34755	34853	98	1	6	0	0	6	1	0	No coding regions
5	46247	46335	88	1	9	0	1	3	0	5	No coding regions
6	78945	79061	116	1	0	0	0	0	1	0	No coding regions
9	12005	12114	109	1	1	0	1	0	0	1	No coding regions
9	119393	121439	2046	1	0	1	0	0	0	0	ASTN2 – TRIM32 – TLR4
13	114140	114170	30	1	0	0	1	0	0	0	DCUN1D2
15	20551	20647	96	1	9	0	0	4	1	5	No coding regions
16	6222	6524	302	1	4	1	0	4	0	0	A2BP1
16	6551	6589	38	1	5	1	0	5	0	0	A2BP1
16	6636	6749	113	1	5	1	0	5	0	0	A2BP1
16	6849	7014	165	1	6	1	0	5	0	1	A2BP1
17	7899	8025	126	1	13	0	1	12	0	1	No coding regions
18	66386	66450	64	1	18	1	0	18	0	0	CCDC102B

Table 8.
Somatic HoDs in MSS tumors (24 matched tumor/normal mucosa couples)

5.3.2 Focal Cytogenetic Analysis: tool to identify focal aberrations

In order to identify gains or losses involving less than 25% of a chromosomal arm, defined focal chromosomal aberrations, another tool, called *Focal Cytogenetic Analysis (FoCyA),* was implemented (Fig. 28).
For each sample two files were used:

- The "*segment file*" (.tsv) that contains a list of all gains or losses (DNA segments spanning multiple consecutive markers showing an increase or a decrease in copy number in comparison to a reference normal population), obtained by algorithms implemented in the GTC

software. Several criteria were used in order to decrease the number of false positive segments during the analysis:

- a)segments must include at least five consecutive monoallelic or biallelic markers showing the same type of copy number change,
- b)the average distance between segmental markers must be lower than 5 kb,
- c) size of segments must be larger than a predefined ‘‘*lower threshold*’’, 100 kb.

- the list of broad aberrations obtained by *BroCyA* tool.

The tool *FoCyA* removes all broad chromosomal aberrations from the segments contained in the “*segment file*”, providing a list of focal aberrations present in the sample.

In order to reveal recurring somatic CNA (copy number abnormalities) segments in tumors samples by comparison with corresponding normal colon mucosa,. analysis in matched tumour/mucosa couples was performed by different scripts (Fig. 29). The first step was to calculate the percentage of overlap between tumor CNAs and normal mucosa ones through the “*Paired Segment Overlap”* (*PSOS*) script. *PSOS* performs a selection of the segments that do not overlap or whose overlap percentage is below a user-defined threshold. Selected adjacent CNA segments with intersegment distance lower than 10Mb are joined. Therefore, *PSOS* provides a list of putative somatic CNAs and this list is further analyzed by the *Contiguous Windows* (*CW*) script. *CW* subdivides each chromosome in contiguous windows of 1 kb and marks them only if they belong to putative somatic CNA segments previously identified by *PSOS*. Lastly, *Recurrent Segment Analysis* (*RSA*) script calculates, for each chromosomal 1kb-window, the number of samples that show a marked

window (recurrence number) and prepares a list of segments by joining adjacent windows that show a recurrence number higher or equal to a predefined threshold (usually displaying different lists according to the threshold recurrence number: from the minimum observed recurrence number to the maximum observed recurrence number). In conclusion, these lists contain the so-called "*recurrent somatic CNA segments*" subdivided according to the recurrence number (Fig. 29).

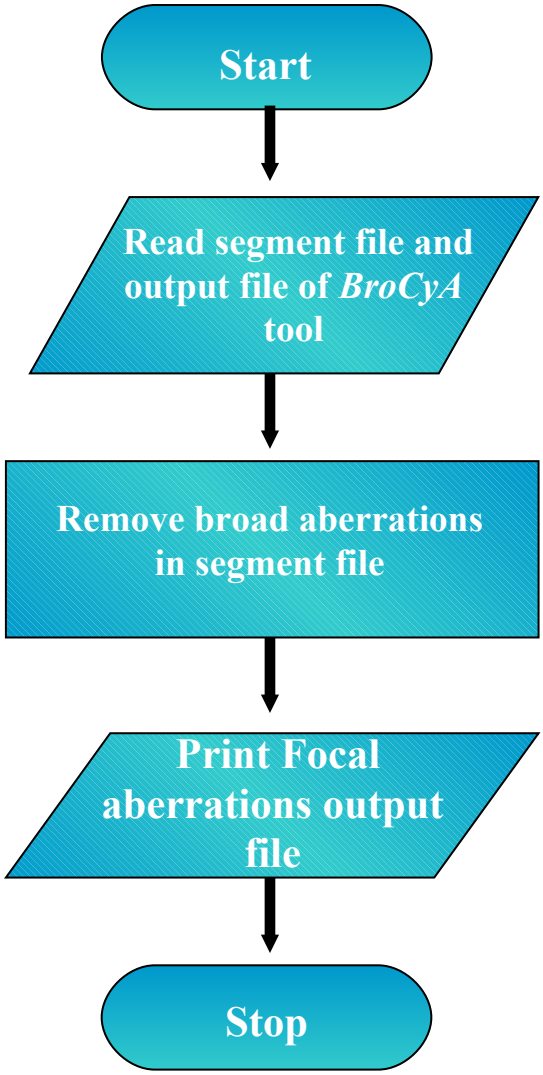

Figure 28
Flowchart of Focal Cytogenetic Analysis (*FoCyA*): tool that identifies focal aberrations

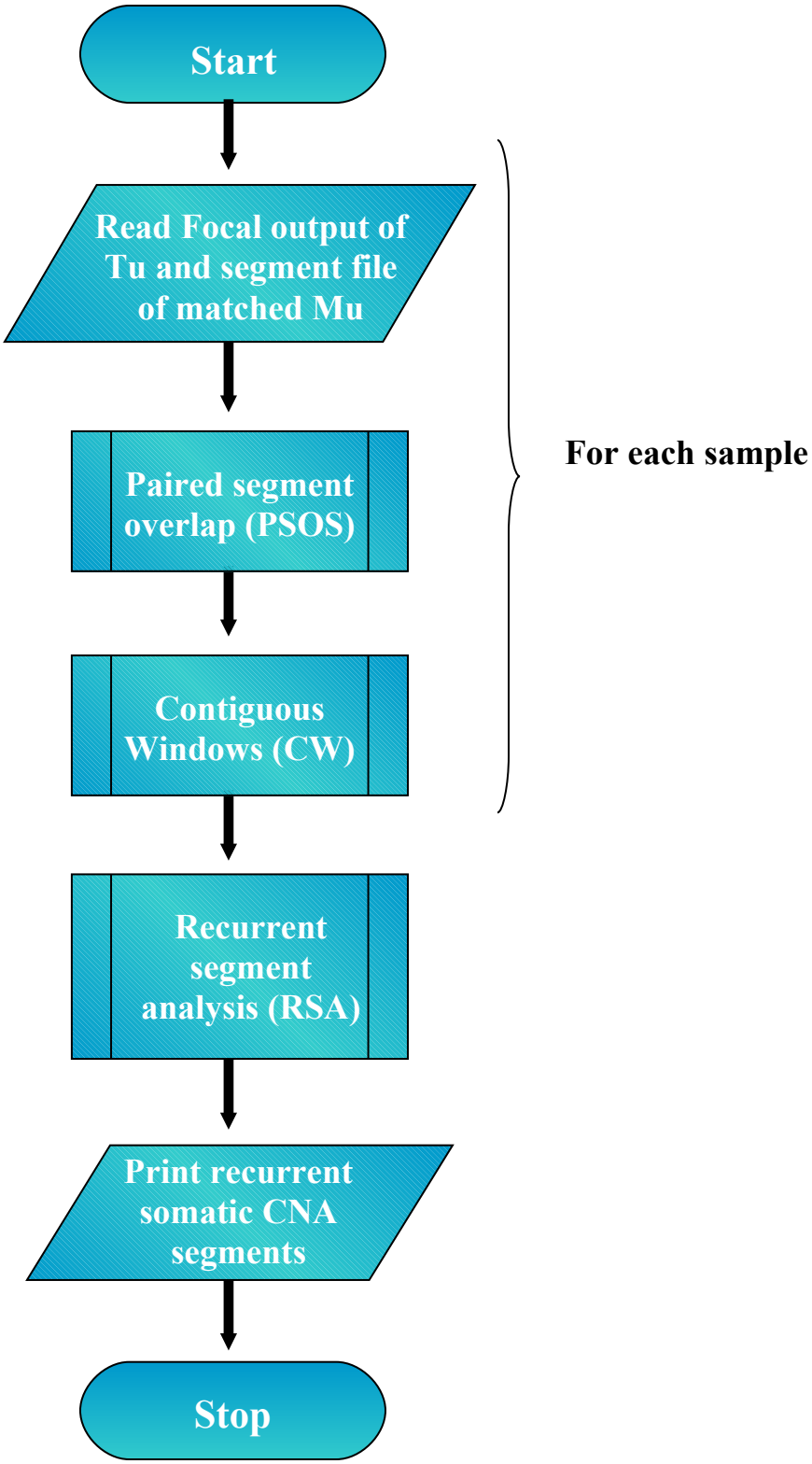

Figure 29
Flowchart to reveal recurring somatic CNA segments in tumor samples.
Tu: tumor sample; Mu: mucosa

5.3.2.1 Somatic Focal CNAs

The *FoCyA* analysis has been performed on 29 matched tumor/normal mucosa samples (5 MSI and 24 MSS). The percentage of patients showing a specific range of focal CNA are shown in Fig. 30. In MSI tumors, all samples, but one, have a low number of focal CNA (0-10 focal aberrations), while in MSS tumors only 30% of the samples show 0-10 focal CNAs (Fig. 30) and 46% of them accumulate more than 20 focal CNAs with 4% reaching the highest level of 70-80 CNAs.

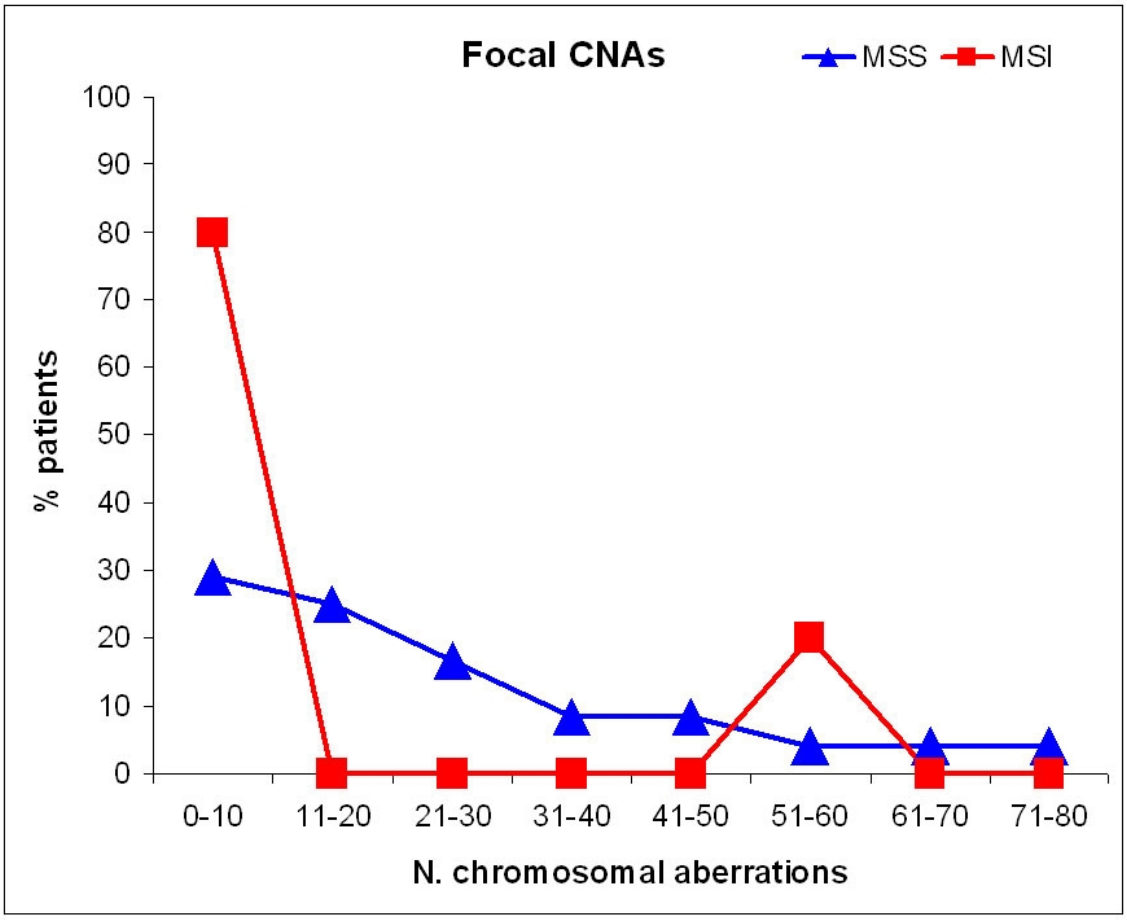

Figure 30
Frequency distribution of focal CNAs in microsatellite instability samples (MSI; red) and microsatellite stability samples (MSS; blue).

5.3.2.2 Recurrent somatic focal CNAs

In order to reveal recurring somatic segments in tumor samples, the analysis in matched tumour/mucosa samples was performed by different scripts (*CW* and *RSA*), as described above (Fig. 29).

For each type of somatic focal CNAs, the minimal region of overlap among the samples has been reported in Tables 9-10.

The most recurrent focal gain, occurring in 20,7% of cases (6 samples), was harbored within chromosome 1, while the most frequent focal losses were localized on chromosomes 16, 20 and 1.

The recurrent somatic focal CNAs found in 29 CRC samples contain interesting candidate cancer genes, which can play a driver role in CRC cancer. For this reason, the genes contained in the focal CNA regions have been searched by using the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and are reported in Table 9 and 10.

Chrom	Start	End	Size	Freq.	Gene
1	3090	3192	102	6	PRDM16
2	151237	151355	118	6	RND3
2	98159	98270	111	5	ANKRD36B , COX5B
11	2142	2224	82	5	IGF2, INS-IGF2, IGF2AS, INS, TH
12	3230	5386	2156	5	TSPAN9, PRMT8, HRMT1L3, EFCAB4B, PARP11, FGF23, RAD51AP1, DYRK4, AKAP3, NDUFA9, GALNT8, KCNA6, KCNA1, KCNA5
12	34482	34558	76	5	No coding regions
12	125748	130993	5245	5	TMEM132B, TMEM132C, SLC15A4, GLT1D1, TMEM132D, FZD10, PIWIL1, RIMBP2
15	22410	22589	179	5	OR4N2
16	2869	3057	188	5	PRSS21, ZG16B, PRSS22, FLYWCH2, FLYWCH1, KREMEN2, PKMYT1
17	39609	39625	16	5	KRT32
17	73466	73642	176	5	CASKIN2, TSEN54, LLGL2, RECQL5
17	74158	77369	3211	5	Gene rich region
2	157692	157807	115	4	No coding regions
2	165844	165965	121	4	SCN3A

2	191629	191863	234	4	GLS, STAT1
2	194531	194633	102	4	No coding regions
2	209089	209217	128	4	IDH1, PIKFYVE
2	236101	236277	176	4	No coding regions
6	169097	169171	74	4	No coding regions
7	1889	2222	333	4	MAD1L1
7	70064	70215	151	4	AUTS2
7	84245	84359	114	4	No coding regions
8	41449	42999	1550	4	AGPAT6, NKX6-3, ANK1, MYST3, AP3M2, PLAT, IKBKB, POLB, DKK4, CHRNB3, THAP1, RNF170, HOOK3, FNTA, SGK196, HGSNAT
9	118882	118997	115	4	PAPPA
10	5019	5120	101	4	AKR1C1, AKR1C2, AKR1C3
10	122980	124358	1378	4	FGFR2, ATE1, NSMCE4A TACC2, BTBD16, PLEKHA1, ARMS2, HTRA1, DMBT1
10	125764	129668	3904	4	CHST15, OAT, NKX1-2, LHPP, ZRANB1, CTBP2, EDRF1, C10orf137, MMP21, BCCIP DHX32, FANK1, ADAM12, DOCK1, FOXI2
11	1968	2065	97	4	MRPL23
11	130914	133804	2890	4	NTM, OPCML, SPATA19, IGSF9B
12	52912	53015	103	4	KRT5,KRT71, KRT74, KRT72, KRT6, KRT73
12	54154	54567	413	4	HOXC13, HOXC12 , HOXC10, HOXC11, HOXC9, HOXC8, HOXC6, HOXC5, HOXC4
17	32193	37752	5559	4	Gene rich region
17	44215	44365	150	4	KIAA1267
19	31748	31907	159	4	TSHZ3
19	33749	33925	176	4	CEBPA, CEBPG, PEPD
21	46848	46925	77	4	COL18A1, SLC19A1
1	248710	248815	105	3	OR2T29, OR2T34, OR2T10, OR2T11, OR2T35, OR2T27 (olfactory receptor, family 2, subfamily T)
2	22789	22889	100	3	No coding regions
2	121643	121746	103	3	GLI2
2	132882	133137	255	3	No coding regions
2	223058	223191	133	3	PAX3, CCDC140
4	96776	96919	143	3	---------
5	480	619	139	3	SLC9A3, CEP72
7	3918	4063	145	3	SDK1 (sidekick 1 precursor)
8	47013	47198	185	3	No coding regions
9	1418	1565	147	3	No coding regions
9	76946	77055	109	3	No coding regions
9	117034	117153	119	3	COL27A1, ORM1, ORM2, AKNA

9	124463	124603	140	3	DAB2IP
10	72689	72855	166	3	No coding regions
10	114751	114928	177	3	TCF7L2
10	132127	132275	148	3	No coding regions
11	284	405	121	3	NLRP6, ATHL1, IFITM5, IFITM2, IFITM1, B4GALNT4, PKP3
11	516	673	157	3	HRAS, LRRC56, RASSF7, PHRF1, SCT, DRD4, DEAF1
11	70241	70398	157	3	CTTN, SHANK2
12	41802	45799	3997	3	PDZRN4, GLT8D3, PRICKLE1, ADAMTS20, PUS7L, IRAK4, TMEM117, NELL2, DBX2, PLEKHA9, ANO6
12	101250	101351	101	3	ANO4
12	114473	114629	156	3	No coding regions
12	115852	118037	2185	3	MED13L , MAP1LC3B2, RNFT2, HRK, FBXW8, TESC, FBXO21, NOS1, KSR2
13	27413	27556	143	3	No coding regions
16	573	870	297	3	SOLH, PIGQ, RAB40C, WFIKKN1, WDR90,WDR24, FBXL16, METRN, HAGHL, NARFL, MSLN, RPUSD1 (RNA pseudouridylate synthase domain containing), CHTF18, GNG13
16	9471	15735	6264	3	Gene rich region
16	79228	79527	299	3	WWOX
17	50025	50148	123	3	CA10
17	50938	52481	1543	3	KIF2B
19	1381	1506	125	3	NDUFS7, GAMT,DAZAP1, RPS15, APC2, PCSK4, ADAMTSL5
19	43425	43562	137	3	PSG7, PSG11, PSG1, PSG2, PSG3, PSG4, PSG6
21	42620	42724	104	3	BACE2

Table 9
Recurrent somatic focal gains found by *FoCyA* tool in 29 matched tumor/normal mucosa couples (5 MSI and 24 MSS)

Chrom	Start	End	Size	Freq.	Gene
20	14905	14961	56	7	MACROD2
1	24091	24238	147	5	LYPLA2, GALE, HMGCL, FUCA1, CNR2
1	24918	24957	39	5	C1orf130
1	25516	25647	131	5	SYF2, RHD
1	26782	27627	845	5	DHDDS, RPS6KA1, ARID1A, PIGV, ZDHHC18, GPN2, GPATCH3, NR0B2, NUDC, TRNP1, WDTC1
1	28725	28841	116	5	PHACTR4, SNHG3-RCC1
1	29008	29122	114	5	GMEB1, YTHDF2
16	6579	6674	95	5	A2BP1
1	161524	161630	106	4	FCGR3B
3	60422	60442	20	4	FHIT
4	188951	189140	189	4	No coding regions
16	5853	5954	101	4	No coding regions
17	57767	57873	106	4	CLTC, PTRH2, TMEM49
1	9802	9938	136	3	CLSTN1, CTNNBIP1
1	51590	51614	24	3	C1orf185
1	53113	53231	118	3	FAM159A, C1orf163, ZYG11B
3	49237	49327	90	3	C3orf62, USP4
4	113581	113676	95	3	No coding regions
4	170735	171029	294	3	MFAP3L, AADAT
4	172432	172535	103	3	No coding regions
4	177771	177896	125	3	No coding regions
5	60208	60450	242	3	ERCC8, NDUFAF2
5	68436	68569	133	3	CCNB1, CENPH, CDK7
6	162440	162602	162	3	PARK2
6	162687	162790	103	3	PARK2
11	48743	48803	60	3	No coding regions
11	107683	107791	108	3	SLC35F2
11	108003	108155	152	3	ACAT1, NPAT, ATM
12	51713	52677	964	3	BIN2, CELA1, GALNT6, SLC4A8, SCN8A, ACVRL1, ANKRD33, ACVR1B, GRASP, NR4A1, KRT80, KRT7, KRT86
15	50896	51020	124	3	TRPM7, SPPL2A
19	50989	51114	125	3	JOSD2, LRRC4B

Table 10
Recurrent somatic focal losses found by *FoCyA* tool in 29 matched tumor/normal mucosa couples (5 MSI and 24 MSS)

5.4 Copy neutral loss of heterozygosity (CN-LOH) in CRC samples

A major advantage of the SNP array is its ability to identify loss of heterozygosity (LOH) regions that occurs without concurrent changes in the gene copy number. Such defects, called CN-LOH, are attributed to mitotic recombination occurring in somatic cells or to abnormalities following repair of DNA double-strand breaks. In addition, a new parameter obtained by GTC, called "*allele difference*" (difference of allele A signal and allele B signal, each standardized with respect to their median values in the reference population) was extremely useful for visualization of CN-LOH region. (Fig. 31). Although extended regions of CN-LOH can be constitutional (germ-line origin), the majority of CN-LOH regions higher than 5 Mb are somatic aberrations. With this in mind, an analysis of CN-LOH regions larger than 5 Mb (called broad CN-LOH) was performed in all 51 tumor samples.

The most common broad CN-LOHs have been observed in chromosome 22 (23.5%), in chromosome 6 (25.5%), in chromosome 9 (15.7%), in chromosome10 (15.7%), and in chromosome17 (15.7%) (Fig. 32). The somatic nature of the chromosomal aberration was confirmed in those samples in which the corresponding normal mucosa was available (29 samples).

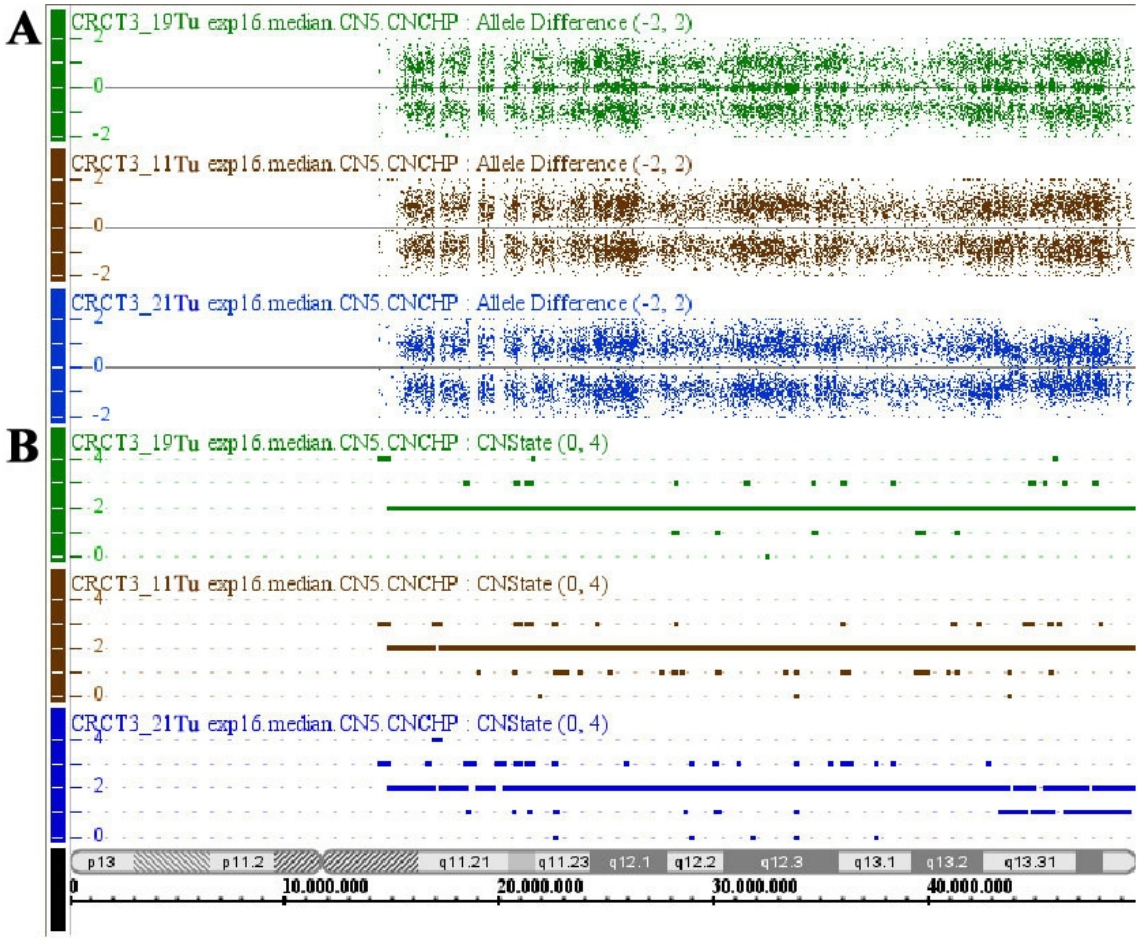

Figure 31
CN-LOHs in CRC tumors.
View of CN-LOH in chromosome 22 in CRC samples: 19Tu, 11Tu, and 21Tu.
A) Allele Difference is the signal intensity of allele A minus the signal intensity of allele B. In the diploid status each SNP marker can have three possible genotypes (AA, AB, BB, top plot, green dots) and allele difference values fluctuate around three values (-1, 0, 1). In case of CN-LOH two genotypes are possible (AA, BB) around two values (-1, +1, middle plot, brown dots). In case of deletion status two genotypes are possible (AA, BB) with two values (-0.5, and +0.5, bottom plot, blue dots in a terminal q arm).
B) The Log_2Ratio between signal for each marker in each patient sample and the corresponding median value in a reference group (270 HapMap individuals) provides an estimate of copy number (CN). Hidden Markov Model converts Log_2Ratio value in CN States (CN 0 = homozygous deletion, CN 1 = heterozygous deletion, CN 2 = diploid state, CN 3 = single copy gain, 4 = amplification $\geq$ 4). In sample 19Tu, CN State is diploid as indicated by a continuous line in correspond of value 2 (green dots). In sample 11Tu CN State 2 confirm the presence of CN-LOH (brown blot). In case of deletion (blue dots, sample 21Tu, q13.31-terminal deletion) appears CN State 1.

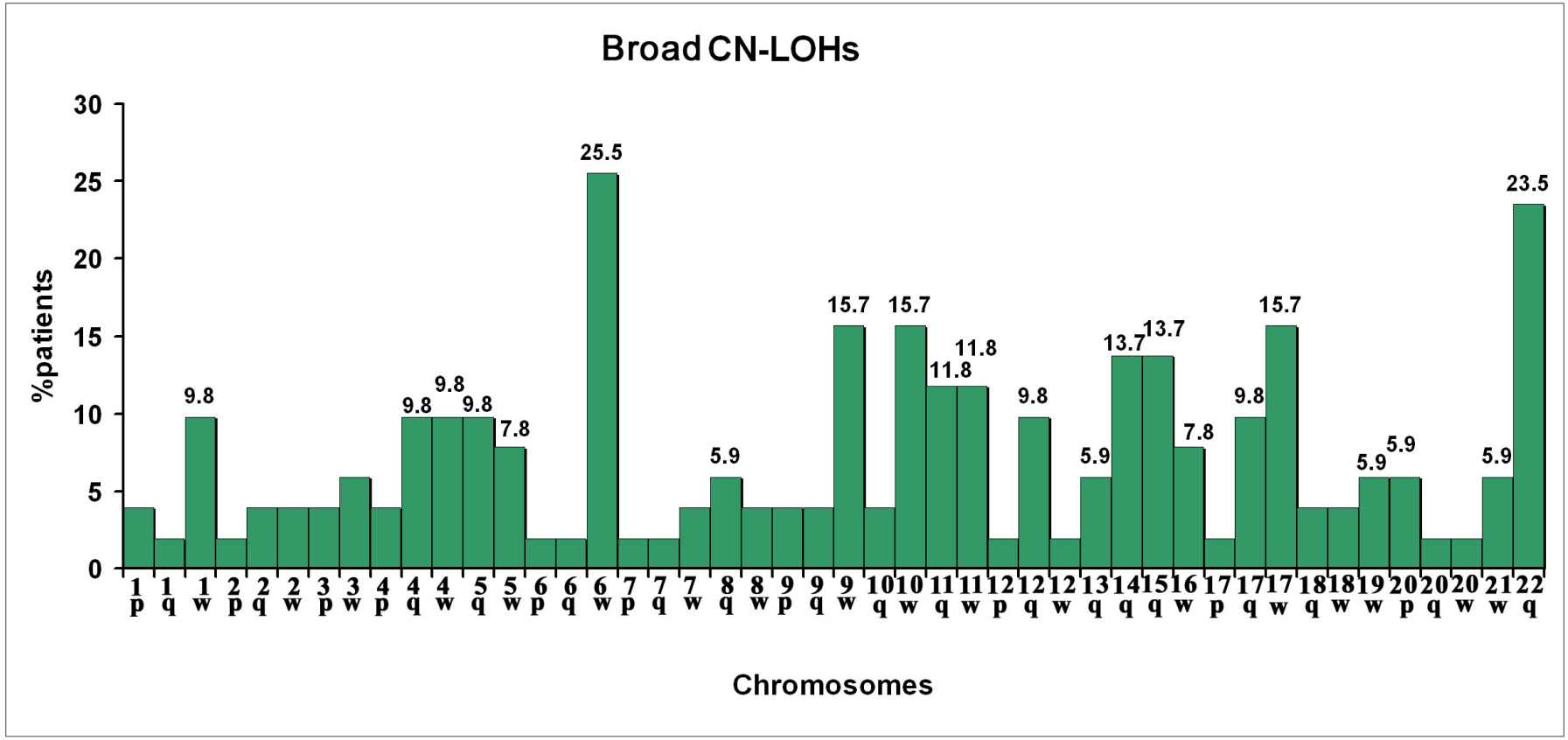

Figure 32
Percentuage of broad CN-LOHs in all CRC samples

Fig. 33 shows the frequency distribution of broad CN-LOH number per sample in the population of CRC samples (7 MSI tumors and 44 MSS tumors). About 60% of the MSI tumors bear >1 broad CN-LOHs, showing that this type of chromosomal abnormality is significantly present in karyotypically quasi-euploid samples. However, MSS tumors show a significantly higher number of broad CN-LOH segments in comparison to MSI tumors.

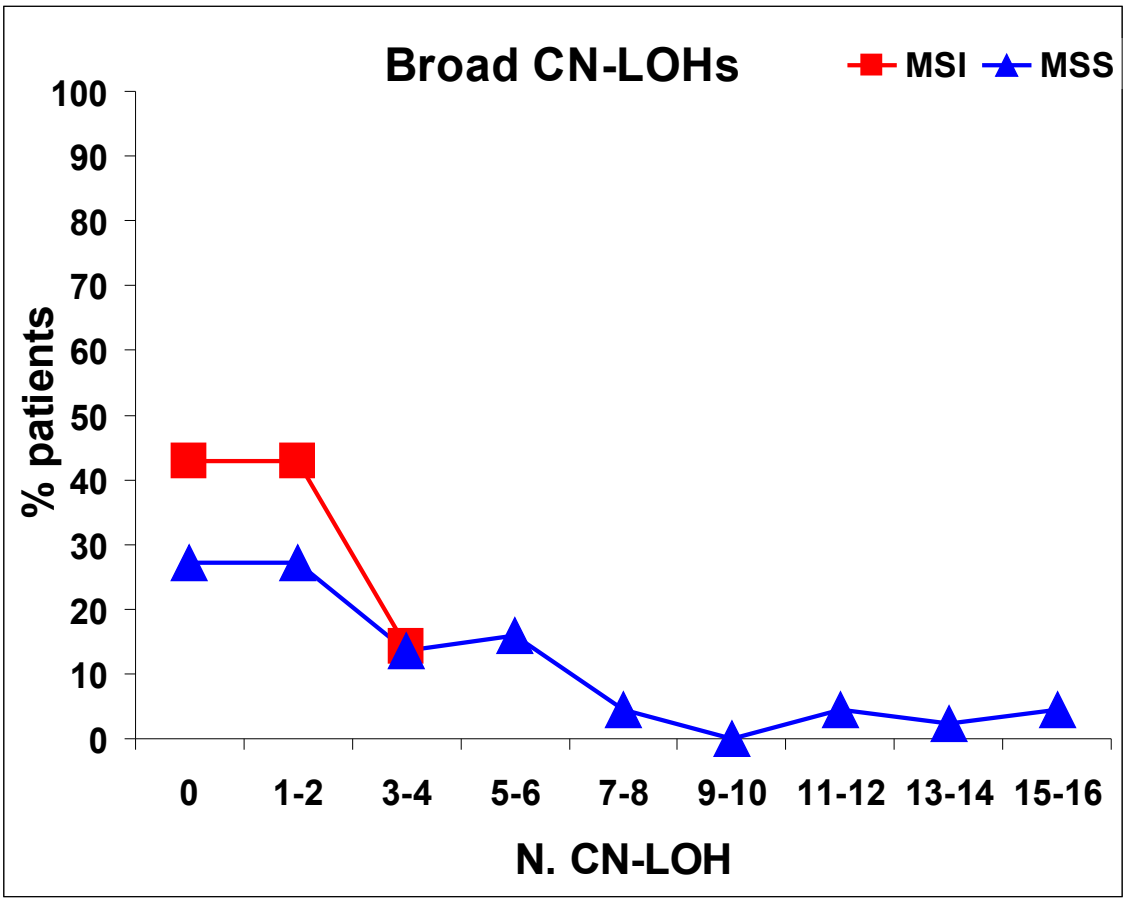

Figure 33
Frequency distribution of CN-LOH regions in microsatellite instability samples (MSI; red) and microsatellite stability samples (MSS; blue).

5.5 Correlation between broad CNAs and broad CN-LOH in MSS and MSI tumors

After obtaining the frequency of each type of CNAs in MSS and MSI tumors, the correlation between them in each status could been determined. As shown in Fig. 34, a low correlation (r=0.35) has been found between gains and losses in MSS tumors, while in MSI this type of correlation is higher (r=0.61) (Fig. 35). The number of broad losses found in MSS tumors was not significantly correlated with the number of broad CN-LOHs in MSS tumors (r=0.2) (Fig. 36). An inverse correlation has been revealed between broad losses and broad CN-LOHs (r=-0.8) and between gains and broad CN-LOHs (r=-0.48) in MSI tumors (Fig. 37-38). Finally a low but significant relationship between broad gains and broad CN-LOHs in MSS tumors (r=0.58) has been observed (Fig. 39).

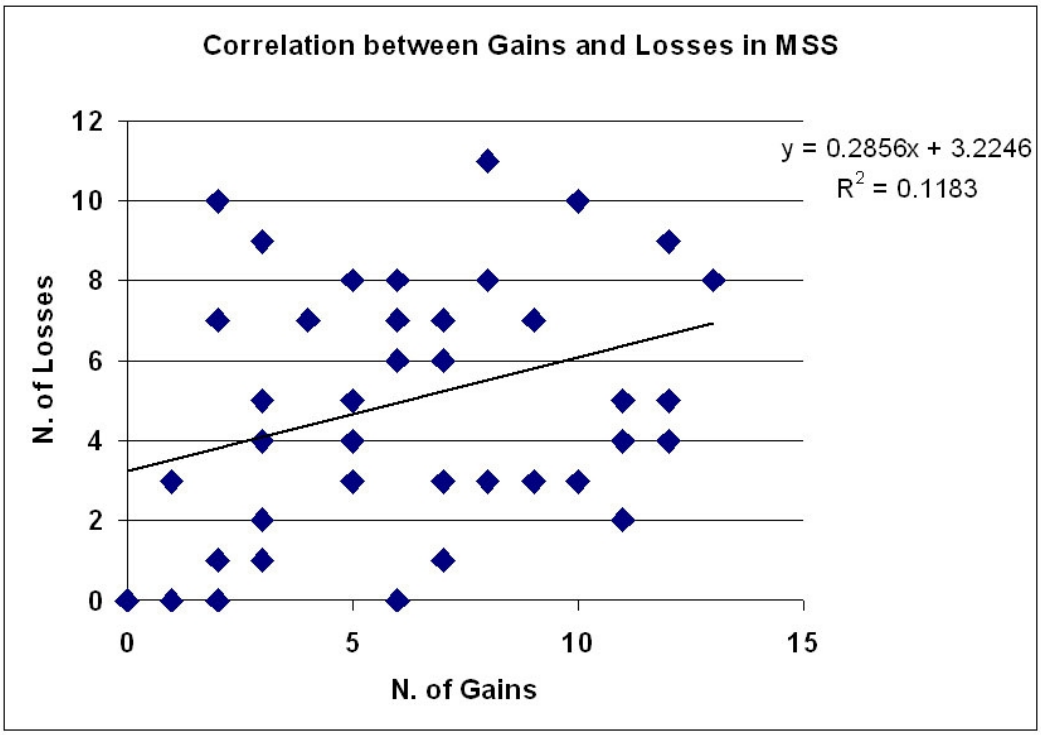

Figure 34
Trend line indicates the correlation between broad Gains vs broad Losses in MSS status

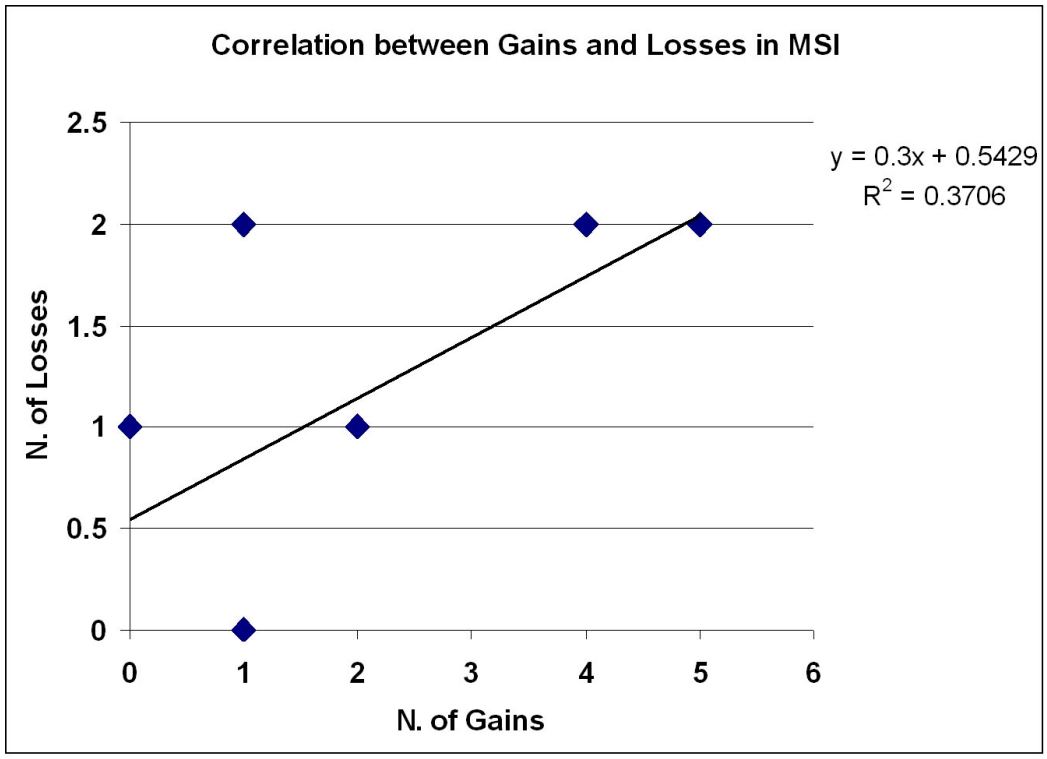

Figure 35
Trend line indicates the correlation between broad Gains vs broad Losses in MSI status

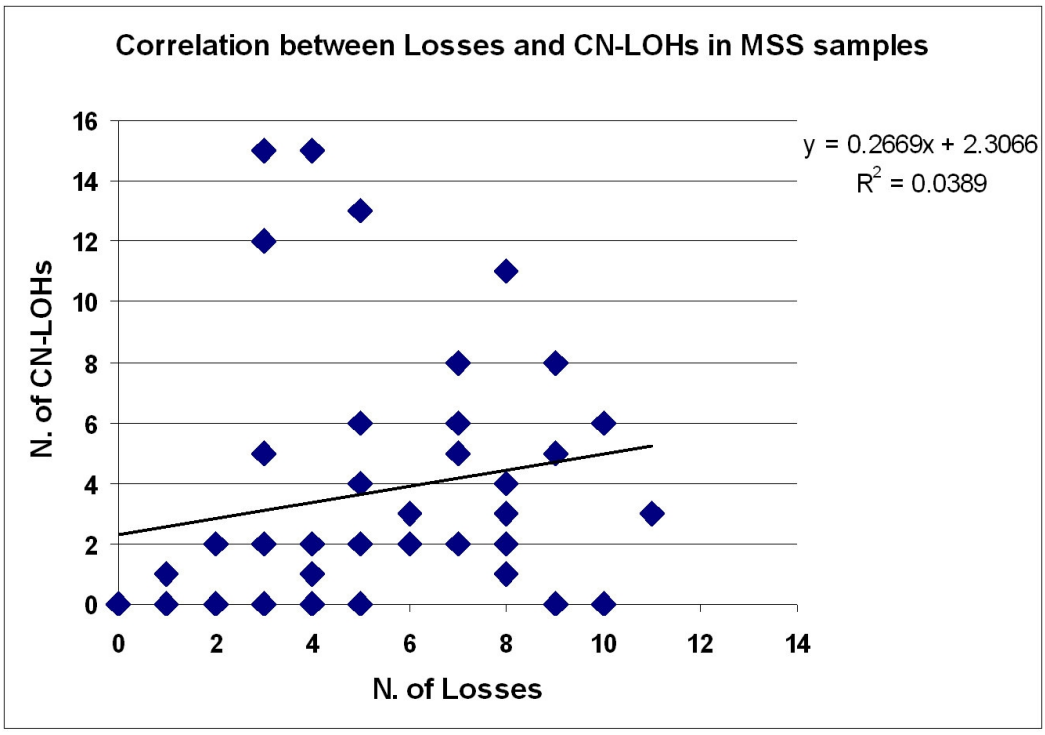

Figure 36
Trend line indicates the correlation between broad Losses vs broad CN-LOHs in MSS status

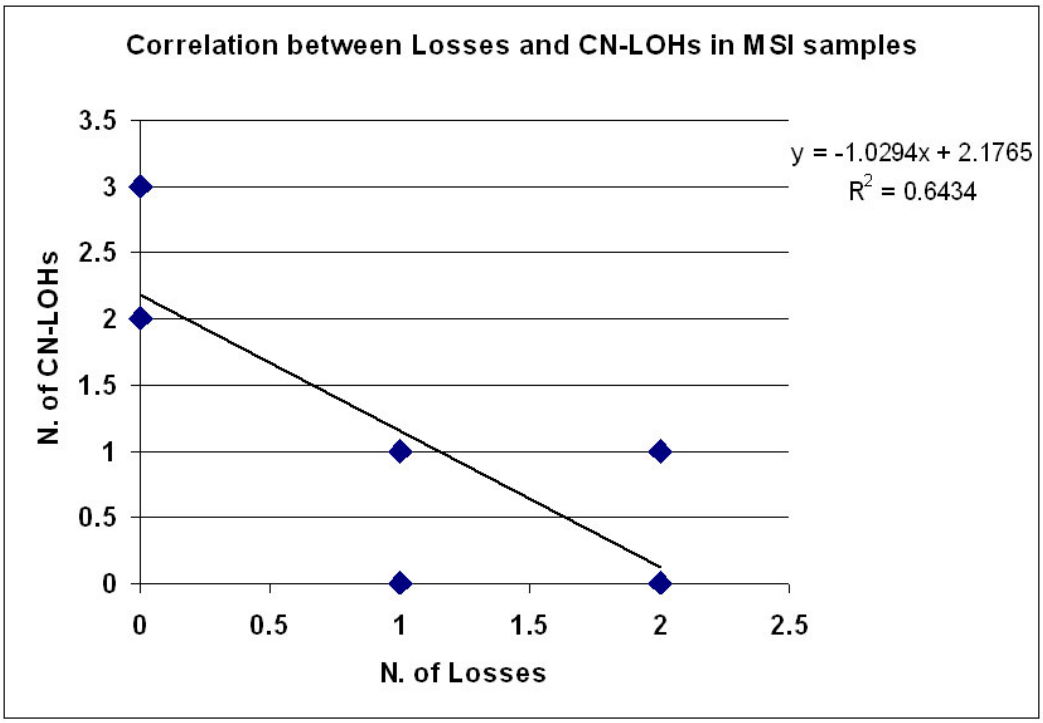

Figure 37
Trend line indicates the correlation between broad Losses vs broad CN-LOHs in MSI status

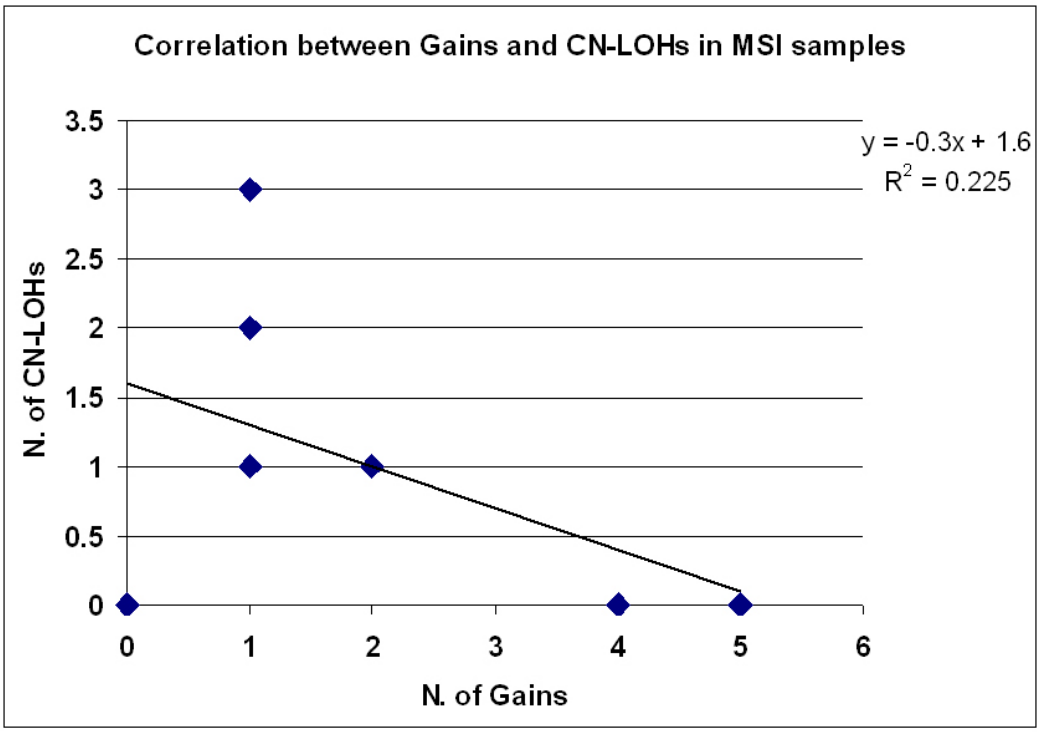

Figure 38
Trend line indicates the correlation between broad Gains vs broad CN-LOHs in MSI status

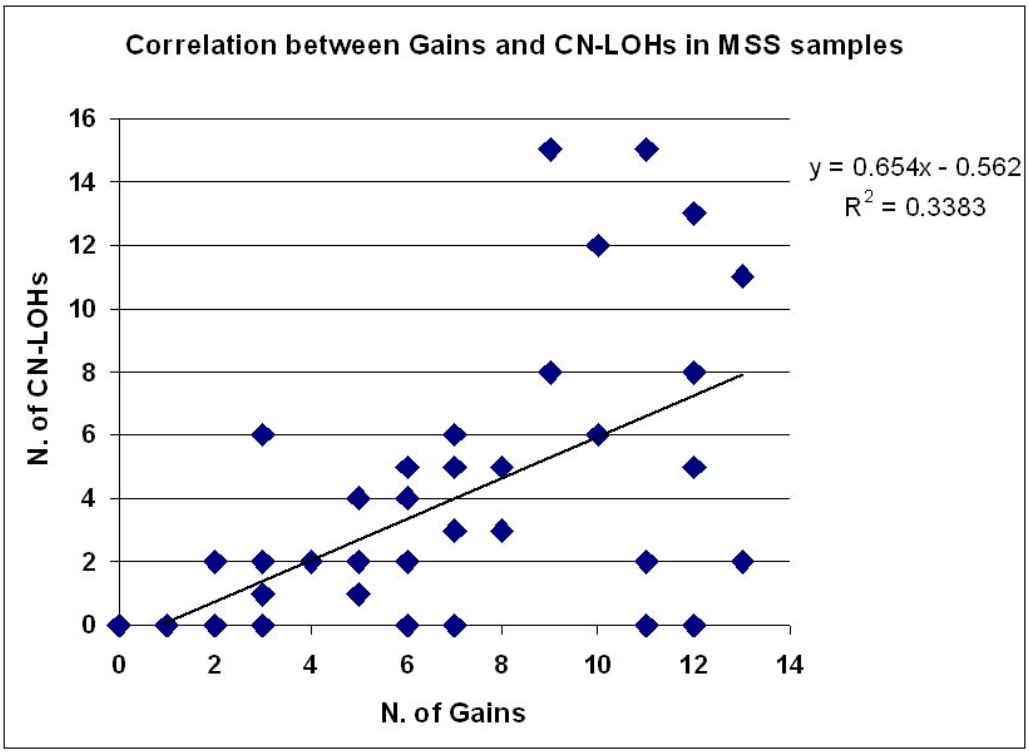

Figure 39
Trend line indicates the correlation between broad Gains vs broad CN-LOHs in MSS status

6 Discussion

The data of this study demonstrates and confirms the heterogeneity of CRC and underlines the importance of the molecular characterization of CRC.

MSI tumors were separated by MSS tumors using Microsatellite Instability test. Microsatellite instability status was detected in 13.73% of samples and this figure is in accordance with the relevant literature where MSI tumors usually make out ~15% of a representative series of colorectal carcinomas [88].

MSI status has a prognostic role in the adjuvant setting, while its predictive role is not clear. In fact Elsalesh et al. (2001) showed that MSI was a factor predictive of response to the 5-FU-based adjuvant therapy in stage III MSI CRCs [36]. In contrast to these results, several studies showed that MSI CRC patients do not benefit from 5-FU based therapy, as compared to patients with MSS CRC [14, 63, 88, 95]. The role of MSI as predictive marker is, probably, influenced by mutations of other genes involved in CRC.

In addition, K-RAS gene is considered a predictive factor for lack of response to epidermal growth factor receptor inhibitors such as cetuximab and panitumumab in patients with CRC [5, 69, 83, 117]. Patients with KRAS mutation in codon 12 or 13 did not benefit from treatment with cetuximab or panitumumab and these mutations are associated with colorectal development through both CIN and MSI pathways. In both groups, the analysis of KRAS mutation was performed and K-RAS mutation occurred in about 41.18% of tumours, which is coherent with the results of the literature [42, 70].

The KRAS mutation was found more frequently in MSS patients, but in present study this is, probably, caused by the low number of MSI tumors. In addition, in MSS tumors a correlation between stage and frequency of KRAS mutation was observed: high frequency of KRAS mutation were revealed in CRC stages III and IV.

Moreover, the most frequent KRAS mutation was Gly12Val. Andreyev et al. (2001) showed that the presence of the mutation Gly12Val predicts a more aggressive behaviour in CRC patients [8].

Using high-resolution 6.0 SNP-arrays it has been possible to describe a comprehensive map of the genetic abnormalities present in colorectal cancer. In the majority of cases, this technique revealed a stable karyotype (euploid or quasi-euploid karyotype) in MSI tumors. On the contrary several chromosomal aberrations were observed in MSS tumors. Moreover the combined use of SNP array and microsatellite assay allowed the detection of a small subpopulation of CRC tumors that are stable both at the chromosomal and microsatellite level (MSS, CIN-), confirming previous observations [120]. Such subgroup could be investigated for prognosis and response to specific therapy.

SNP-array analysis, performed in this thesis, has confirmed the non random distribution of chromosomal aberration as previously reported using chromosome banding techniques [32], CGH [30], aCGH [20, 55, 65, 67, 112], low-resolution 50k SNP-arrays [108] and 500k SNP-arrays [104]. Typical broad chromosomal aberrations are amplifications on chromosome 7, 8q, 13q, 20 or deletions on chromosome 4, 5q, 8p, 17p, 20p, 22q.

Different previous studies have proved that trisomies of chromosomes 7 and 13 occur early in colorectal tumorigenesis because they are frequently seen in the early stages of colorectal carcinogenesis, and trisomy 7 is often found as the sole aberrations in colorectal adenomas

[51]. Such observations suggest that these changes may participate in the initial events of colorectal tumorigenesis.

The amplification of 8q has been associated with the presence of more oncogenes, as MYC, a key regulators of cell proliferation, whose deregulation contributes to the genesis of most human tumours.

Gain of 20q has been associated with colorectal tumorigenesis: this aberration has been seen more often in CRC carcinomas than in low-grade and high-grade adenomas [97], and with cell immortalization [103].

In addition, it has been seen that the loss of chromosome 4 is more frequent in late stages than in early stages of colorectal carcinogenesis [12, 48, 65, 97]. Deletion of 8p, the second common loss in the present study (47.7% in MSS tumors), followed by gain of 8q occur more often in colorectal carcinomas and metastases than in adenomas, suggests that alteration of chromosome 8 is associated with tumor progression. Losses of chromosome arms 17p and 18q are important events in colorectal tumorigenesis [39]. Loss of 18, often ascribed to monosomy 18, and 17p was seen in 65.9% and 47.7% of MSS tumors of this study, respectively. Loss of 17p and 18q has therefore been suggested to be a late and important event in the adenoma–carcinoma sequence [39], and most likely reflects inactivation of the TP53 [39], DCC [24], SMAD2, and SMAD4 [9, 24] genes. Finally, allelic loss on chromosome 22q is present not only in CRC but also in oral (40%) [76], brain (40%) [98], ovarian (55%) [37], breast (40%) [57], pancreatic endocrine tumor (30%) [10], gastrointestinal stromal tumor (77%) [26], and even hepatocellular carcinoma[44].

In addition, 6.0 SNP-arrays and implementation of *BroCy*A tool made it possible to reveal other significant aberrations in MSS tumors (e.g. amplification of chromosome 2, 5, 9, 12, 17q, 19 or deletion on chromosome 15q, 18, 19p, 22q). In MSI tumor group a low number of

broad chromosomal aberrations were also observed but they hit the same chromosomes involved in MSS tumors (gains in chromosomes 7, 8, 9, 11, 13 and 20 and losses on chromosomes 17p, 19p, 19 and 22q).

HLAs were revealed in this study only in MSS tumors. Their mechanisms of formation is not clear, but Sheffer et al. (2009) [107] detected different HLAs containing several known oncogenes as MYC, LYN, MET. In the present thesis, high number of recurrent HLAs on chromosome 20 were observed. Interestingly a frequent HLA localized on 20q13.12 contains the STK4 gene. The protein encoded by this gene is a cytoplasmic kinase that is structurally similar to the yeast Ste20p kinase, which acts upstream of the stress-induced mitogen-activated protein kinase cascade. The encoded protein can phosphorylate myelin basic protein and undergoes autophosphorylation. A caspase-cleaved fragment of the encoded protein has been shown to be capable of phosphorylating histone H2B. The particular phosphorylation catalyzed by this protein was correlated with apoptosis, and it is possible that this protein induces the chromatin condensation observed in this process. Moreover, Babel et al. described this gene as an autoantibody targets for the diagnosis of colorectal cancer [11].

Another interesting gene, localized on the same recurrent HLA regions, at 20q13.12, is the inhibitor of differentiation/DNA binding 1 (Id-1), a negative regulator of basic helix-loop-helix transcription factors. It plays an important role in the regulation of cell proliferation and differentiation. Id-1 was found to be involved in the invasion and metastasis of gastric [61], oral [77], breast [40] and prostate cancers [28]. Zeng-Ren Zhao et al. (2008) [121] in their study noted that the frequency of the Id-1 stronger expression was significantly increased in the advanced Dukes' stage and the cases with lymph node metastasis. In conclusion,

these results suggested that Id-1 protein is involved in the development of colorectal cancers and its overexpression may be a marker in tumor progression.

As previously observed, also the chromosomal regions 20q13.3 is an hotspots for potential tumor suppressor genes and the most interesting ones for their probable role in CRC pathogenesis is PMEPA1. It is a TGF-β-induced transmembrane protein that is overexpressed in several cancers, as breast, ovarian cancer. How PMEPA1 expression relates to malignancy is unknown. Prajjal K. Singha et al (2010) in their study reported high expression of PMEPA1 in ER/PR-negative and HER2-negative breast cancer cell lines and primary breast cancers [111]. They suggested with their study that overexpression and/or increased or altered function of PMEPA1 may be a "*molecular switch*" that converts TGF-β from tumor suppressor to tumor promoter.

Homozygous deletions in general occur over recessive cancer genes, where they can confer selective growth advantage, and over fragile sites, where they are thought to reflect an increased local rate of DNA breakage. However, most homozygous deletions in cancer genomes are unexplained. In study of Bignell et al. (2010) [16] over 2428 somatic homozygous deletions in different cancer cell lines were identified. Several recessive cancer genes , such as CDKN1A, PTEN etc, were detected on regions characterized by homozygous deletions. However, many unexplained homozygous deletions in cancer genomes have signatures indicative of fragility.

In CRC samples of this thesis, somatic HoD regions contained different loci characterized by recessive cancer genes, such as D4S234E , BTG4 and PARK2.

D4S234E, also designed NEEP21 or NSG1, is located on chromosome 4p16.3. It is an endosomal protein expressed in neuronal cells under normal conditions and contributes to the regulation of synaptic transmission and plasticity in slice cultures by affecting the recycling and targeting of AMPA receptors to the synapse. From the study of Ohnishi et al. (2010) [80] emerged that D4S234E is a direct transcriptional target gene of tumor suppressor p53 and it may plays a critical role in apoptosis as a mediator of p53.

The gene BTG4, located on 11q23, is a novel member of the PC3/BTG/TOB family of growth inhibitor genes [18]. The BTG family genes exert antiproliferative effects and have the ability to induce cell cycle arrest and are thus thought to act as tumor suppressors [114]. Although much remains unclear about the function of BTG4, Auer et al (2005) in their study demonstrated that deletion of 11q is a common abnormality in chronic lymphocytic leukemia, and that inactivation of BTG4 may contribute to the disease's pathogenesis [10].

In study of Dong et al. (2003) [33], it was described that BTG4 undergoes promoter CpG island hypermethylation-associated inactivation in gastric cancer and 5'-aza-2'-deoxycytidine (DAC) treatment restores BTG4 expression. Dong et al (2003) [33] also found that BTG4 levels were significantly reduced in primary gastric cancer but not in normal gastric tissues. Taken together, their data supports BTG4 as a candidate tumor suppressor gene that is epigenetically silenced in the majority of gastric cancers [33].

In 6q25-27 region a deletion was found, which was identified both as HoD and recurrent somatic focal loss. This deletion involves PARK2, the gene encoding PARKIN, the E3 ubiquitin ligase whose deficiency is responsible for a form of autosomal recessive juvenile parkinsonism.

George Poulogiannis et al. (2010) in their study showed that deficiency in expression of PARK2 is significantly associated with adenomatous polyposis coly (APC) deficiency in colorectal patients [89]. They concluded that PARK2 is a tumor suppressor gene whose haploinsufficiency cooperates with mutant APC in colorectal carcinogenesis.

Some of the recurrent somatic focal CNAs, individuated in this study, contain candidate cancer pathway genes not previously known to play a role in CRC. The most recurrent focal gain, occurring in 20,7% of cases (6 samples), was localized in chromosome 1, while the most frequent focal losses were present on chromosomes 16 and 20. The amplification on chromosome 1q36.32 contains PRMD16, a PR domain-containing 16 involved in human leukemic translocations and highly expressed in some karyotypically normal acute myeloblastic leukemias. Shing et al. (2007) [108] demonstrated that overexpression of sPRDM16 and disruption of the p53 tumor suppressor pathway cooperate in leukemogenesis, both in human AML patients and a murine model of leukemia. p53 is deleted in 5 out of 6 samples bearing the PRDM16 gain (in 4 samples due to a broad 17p deletion and 1 sample due to a focal loss).

The focal recurrent somatic deletion on 16p13.3 contains A2BP1, ataxin 2-binding protein 1. Andersen et al. (2010) [6] showed that structural rearrangements involving chr16p13.3 are very frequent in colorectal neoplasia, often leading to homozygous deletions, and are associated with poor clinical outcome.

Finally, recurrent somatic focal losses located on chromosome 20 (20p12.1) include the MACROD2 gene. Davison et al. (2005) [29] found MACROD2 deletions in 23% of their primary colorectal cancer and 55% of colorectal cell lines. They provided evidence that RNA molecules encoded

in the region 20p12.1 may have tumour suppressor activity, but it is also probable that the high frequency of deletions may in part be attributable to the instability of the region (chromosomal fragile site).

With the use of SNP arrays, it has been possible to study another important biological phenomenon involved in colorectal tumorigenesis: copy neutral LOH. CN-LOH is sometimes seen in tumors as a way to inactivate tumor suppressor genes, and has been reported in many different types of cancer [115]. Moreover, in contrast to hemizygous deletion, CN-LOH would not be expected to cause any haploinsufficiency effects, since the total copy number remains normal.

Thus by combining CN-LOH data with data about deletions and gains, new regions of interest can be identified and previously characterized regions can be further defined. However, the possible clinical implications of CN-LOH are at the moment difficult to evaluate, ranging from potentially harmless to resulting in tumor suppressor gene silencing, depending on mutational and imprinting status of the remaining allele.

Results obtained in this thesis demonstrate that this phenomenon is significantly present in karyotipically quasi-euploid samples, in fact about 60% of MSI tumors bear >1 broad CN-LOHs.

7 Conclusion

CRC is a heterogeneous disease with many molecular phenotypes and does not just acquire certain chromosomal copy number aberrations randomly. Within regions of gains, losses or CN-LOHs there are genes, which, on dysregulation, are transformed into cancer-promoting states. These aberrations, which often involve a whole chromosome arm (eg, 8p, 8q, 13q, 20, 22q), can lead to dysregulation of numerous genes in either their wild-type or mutated states. In the present study both microsatellite stable and unstable phenotypes by their copy-number changes were studied.

The utilization of SNP-array technology in present thesis has permitted to detect on a genome-wide scale, not only somatic CNAs (losses and gains), but also chromosomal areas of CN-LOH. The CN-LOH was defined an important mutational event in relation to the carcinogenesis of MSS and MSI tumors, causing the inactivation of a tumor suppressor gene without copy number alteration of the respective region. Indeed, this chromosomal abnormality was relatively frequent in MSI tumors and could participate in pathogenesis of this tumoral subclass.

The analysis of HLAs, HoDs and focal recurrent somatic alterations in samples of the present study allowed to some potential genes that could play an important role in the development of CRC. In particular, genes, located in regions characterized by high level of amplifications, such as ID-1 at 20q11.21, STK4 at 20q13.12, PMEPA1 at 20q13.31, and in homozygous deletion region such as BTG4 at 11q23.1, and D4S234E at 4p16.3, were highlighted. Finally, the frequent alterations of some possible candidate genes, located in deletion region, such as MACROD2 at 20p12.1,

and A2BP1 at 16p13.3; and in gain regions such as PRDM16 at 1q36.32, were demonstrated in this thesis.

Further investigations based on the sequencing of the coding region, or on the analysis of the epigenetic modifications of the identified genes are required in order to confirm their role in colorectal cancer pathogenesis.
Finally, the bioinformatics tools set up in the present thesis (*BroCyA* and *FoCyA*) can be also exploited in the routine clinical setting for the preparation of an accurate report of relevant data obtained by SNP array analysis of colorectal cancer.

8 References

1. Adam, R., et al., *Five-year survival following hepatic resection after neoadjuvant therapy for nonresectable colorectal*. Ann Surg Oncol, 2001. 8(4): p. 347-53.
2. Aguirre AJ, Brennan C, Bailey G, Sinha R, Feng B, et al. *High-resolution characterization of the pancreatic adenocarcinoma genome*. Proc Natl Acad Sci U S A 101:9067–9072 (2004).
3. Al-Mulla F, AlFadhli S, Al-Hakim AH, Going JJ, Bitar MS. *Metastatic recurrence of early-stage colorectal cancer is linked to loss of heterozygosity on chromosomes 4 and 14q*. J Clin Pathol. 2006a, 59(6):624-30.
4. Al-Mulla F, Behbehani AI, Bitar MS, Varadharaj G, Going JJ. *Genetic profiling of stage I and II colorectal cancer may predict metastatic. relapse*. Mod Pathol. 2006b, 19(5):648-58.
5. Amado RG, Wolf M, Peeters M et al. *Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer*, J Clin. Oncol. 2008;26(10):1626-34
6. Andersen CL, Lamy P, Thorsen K, Kjeldsen E, Wikman F, Villesen P, Oster B, Laurberg S, Orntoft TF. *Frequent genomic loss at chr16p13.2 is associated with poor prognosis in colorectal cancer.* Int J Cancer. 2010 Dec 10.
7. Andre, T., et al., *Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer*. N Engl J Med, 2004. 350(23): p. 2343-51.
8. Andreyev, H.J.; Norman, A.R.; Cunningham, D.; Oates, J.; Dix, B.R.; Iacopetta, B.J.; Young, J.; Walsh, T.; Ward, R.; Hawkins, N.; et al. *Kirsten ras mutations in patients with colorectal cancer: the _RASCAL II' study*. Br. J. Cancer 2001, 85, 692-696.

9. Arends JW. *Molecular interactions in the Vogelstein model of colorectal carcinoma.* J Pathol 190:412–416, 2000.
10. Auer RL, Starczynski J, McElwaine S, et al. *Identifi- cation of a potential role for POU2AF1 and BTG4 in the deletion of 11q23 in chronic lymphocytic leukemia*. Genes Chromosomes Cancer 2005;43:1–10.
11. Babel I, Barderas R, Diaz-Uriarte R, Moreno V, Suarez A, Fernandez-Aceñero MJ, Salazar R, Capellá G, Casal JI. *Identification of MST1/STK4 and SULF1 proteins as autoantibody targets for the diagnosis of colorectal cancer by using phage microarrays* Mol Cell Proteomics. 2011 Mar;10(3):M110.001784..
12. Bardi G, Sukhikh T, Pandis N, Fenger C, Kronborg O, Heim S. *Karyotypic characterization of colorectal adenocarcinomas*. Genes Chromosomes Cancer 12:97–109. 1995.
13. Bardi G, Fenger C, Johansson B, Mitelman F, Heim S. *Tumor karyotype predicts clinical outcome in colorectal cancer patients*. J Clin Oncol. 2004 Jul 1;22(13):2623-34.
14. Benatti, P.; Gafa, R.; Barana, D.; Marino, M.; Scarselli, A.; Pedroni, M.; Maestri, I.; Guerzoni, L.; Roncucci, L.; Menigatti, M.; et al. *Microsatellite instability and colorectal cancer prognosis*. Clin. Cancer Res. 2005, 11, 8332-8340.
15. Benson AV, 3rd. *New approaches to the adjuvant therapy of colon cancer*. Oncologist. 2006 Oct; 11 (9):973-80
16. Bignell Graham R, Chris D. Greenman, Helen Davies, Adam P. Butler et al., *Signatures ofmutation and selection in the cancer genome*, Nature, Vol 463| 18 February 2010
17. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling*. Nature, 2001. 411(6835): p. 355-65.

18. Buanne P, Corrente G, Micheli L, et al. *Cloning of PC3B, a novel member of the PC3/BTG/TOB family of growth inhibitory genes, highly expressed in the olfactory epithelium*. Genomics 2000;68:253–63.
19. Butler, J., F.F. Attiyeh, and J.M. Daly, *Hepatic resection for metastases of the colon and rectum*. Surg Gynecol Obstet, 1986. 162(2): p. 109-13.
20. Camps J, Grade M, Nguyen QT, Hormann P, Becker S, et al. *Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome*. Cancer Res 68: 1284–95, 2008.
21. Capon F, Allen MH, Ameen M, Burden AD, Tillman D, Barker JN, Trembath RC. *A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups*. Hum Mol Genet. 2004;13:2361-8.
22. Cassidy, J., *Potential of Xeloda in colorectal cancer and other solid tumors*. Oncology, 1999. 57 Suppl 1: p. 27-32.
23. Center MM, Jemal A, Smith RA, Ward E. *Worldwide variations in colorectal cancer.* CA Cancer J Clin. 2009 Nov-Dec;59(6):366-78. Review.
24. Cho KR, Fearon ER. *DCC: linking tumor suppressor genes and altered cell surface interactions in cancer?* Curr Opin Genet Dev 5:72–78. 1995
25. Chollet, P., et al., *Single agent activity of oxaliplatin in heavily pretreated advanced epithelial ovarian cancer*. Ann Oncol, 1996. 7(10): p. 1065-70.
26. Chung DC, Brown SB, Graeme-Cook F, Tillotson LG, Warshaw AL, Jensen RT, Arnold A. *Localization of putative tumor suppressor loci*

by genome-wide allelotyping in human pancreatic endocrine tumors. Cancer Res 1998; 58: 3706-3711

27. Cohen, A.M., et al., *Cancer of the colon, in Cancer:Principles and Practice of Oncology*, V.T. DeVita, S. Hellman, and A.S. Rosenberg, Editors. 1997, Lippincott-Raven: Philadelphia.
28. Coppe JP, Itahana Y, Moore DH, Bennington JL and Desprez PY: *Id-1 and Id-2 proteins as molecular markers for human prostate cancer progression*. Clin Cancer Res 10: 2044-2051, 2004
29. Davison EJ, Tarpey PS, Fiegler H, Tomlinson IP, Carter NP (2005) *Deletion at chromosome band 20p12.1 in colorectal cancer revealed by high resolution array comparative genomic hybridization*. Genes Chromosomes Cancer 44:384–391
30. De Angelis PM, Clausen OP, Schjolberg A, Stokke T, *Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes*. Br J Cancer 80: 526–35 ,1999.
31. Des Guetz G, Schischmanoff O, Nicolas P, Perret GY, Morere JF, Uzzan B. *Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis*. Eur J Cancer. 2009 Jul;45(10):1890-6.
32. Diep CB, Parada LA, Teixeria MR et al., *Genetic profiling of colorectal cancer liver metastases by combined comparative genomic hybridization and G-banding analysis*. Genes Chromosome Cancer 36: 189-97, 2003
33. Dong W, Tu S, Xie J, Sun P, Wu Y, Wang L *Frequent promoter hypermethylation and transcriptional downregulation of BTG4 gene in gastric cancer.* Biochem Biophys Res Commun. 2009 Sep 11;387(1):132-8. Epub 2009 Jul 1.

34. Douillard, J.Y., et al., *Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial*. Lancet, 2000. 355(9209): p. 1041-7.
35. Downward J. Targeting *RAS signalling pathways in cancer therapy*. Nat Rev. *2003;3:11–22.*
36. Elsaleh, H.; Iacopetta, B. *Microsatellite instability is a predictive marker for survival benefit from adjuvant chemotherapy in a population-based series of stage III colorectal carcinoma*. Clin. Colorectal. Cancer 2001, 1, 104-109
37. Englefield P, Foulkes WD, Campbell IG. *Loss of heterozygosity on chromosome 22 in ovarian carcinoma is distal to and not accompanied by mutations in NF2 at 22q12*. Br J Cancer 1994; 70: 905-907
38. Falcone, A., et al., *Phase III trial of infusional fluorouracil, leucovorin, oxaliplatin, and irinotecan (FOLFOXIRI) compared with infusional fluorouracil, leucovorin, and irinotecan (FOLFIRI) as first-line treatment for metastatic colorectal cancer: the Gruppo Oncologico Nord Ovest.* J Clin Oncol, 2007. 25(13): p. 1670-6.
39. Fearon ER, Vogelstein B. *A genetic model for colorectal tumorigenesis.* Cell 61:759–767, 1990.
40. Fong S, Itahana Y, Sumida T, et al: *Id-1 as a molecular target in therapy for breast cancer cell invasion and metastasis*. Proc Natl Acad Sci USA 100: 13543-13548, 2003
41. Forozan F,Mahlamki EH,Monni O, Chen Y, Veldman R, et al. *Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary dna microarray data*. Cancer Res 60:4519–4525 (2000).

42. Fransén K, Klintenäs M, Osterström A, Dimberg J, Monstein HJ, Söderkvist P: *Mutation analysis of the BRAF, ARAF and RAF-1 genes in human colorectal adenocarcinomas*. Carcinogenesis 2004, 25:527-33.
43. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. *Copy number variation: new insights in genome diversity*. Genome Res. 2006;16:949-61
44. Fukasawa T, Chong JM, Sakurai S, Koshiishi N, Ikeno R, Tanaka A, Matsumoto Y, Hayashi Y, Koike M, Fukayama M. *Allelic loss of 14q and 22q, NF2 mutation, and genetic instability occur independently of c-kit mutation in gastrointestinal stromal tumor*. Jpn J Cancer Res 2000; 91: 1241-1249
45. Giacchetti, S., et al., *Long-term survival of patients with unresectable colorectal cancer liver metastases following infusional chemotherapy with 5-fluorouracil, leucovorin, oxaliplatin and surgery.* Ann Oncol, 1999. 10(6): p. 663-9.
46. Grady WM: *Genomic instability and colon cancer*. Cancer Metastasis Rev 2004, 23: 11-27
47. Grem, J.L., *Fluoropyrimidines, in Cancer Chemotherapy and Biotherapy: Principles and Practice*, B.A. Chabner and D.L. Longo, Editors. 1996, Lippincott-Raven: Philadelphia.
48. Griffin CA, Lazar S, Hamilton SR, Giardiello FM, Long P, Krush AJ, Booker SV. *Cytogenetic analysis of intestinal polyps in polyposis syndromes: comparison with sporadic colorectal adenomas*. Cancer Genet Cytogenet 67:14–20, 1993.
49. Gryfe R, Kim H, Hsieh ET, Aronson MD, Holowaty EJ, Bull SB, Redston M, Gallinger S. *Tumor microsatellite instability and clinical*

outcome in young patients with colorectal cancer. N Engl J Med. 2000:13;342(2):69-77.

50. Halling KC, French AJ, McDonnell SK, Burgart LJ, Schaid DJ, Peterson BJ, Moon-Tasson L, Mahoney MR, Sargent DJ, O'Connell MJ, Witzig TE, Farr GH Jr, Goldberg RM, Thibodeau SN. *Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers*. J Natl Cancer Inst. 1999; 91(15):1295-303.
51. Heim S, Mitelman F. *Tumors of the digestive tract. In: Cancer cytogenetics: chromosomal and molecular genetic aberrations of tumor cells*, Second Ed. New York: John Wiley & Sons, p. 332–350, 1995.
52. Heiskanen MA, Bittner ML, Chen Y, Khan J, Adler KE, et al. *Detection of gene amplification by genomic hybridization to cdna microarrays*. Cancer Res 60:799–802 (2000).
53. Holzmann K, Kohlhammer H, Schwaenen C, Wessendorf S, Kestler HA, et al. *Genomic dna-chip hybridization reveals a higher incidence of genomic amplifications in pancreatic cancer than conventional comparative genomic hybridization and leads to the identification of novel candidate genes*. Cancer Res 64:4428–4433 (2004).
54. Hsiang, Y.H. and L.F. Liu, *Identification of mammalian DNA topoisomerase I as an intracellular target of the anticancer drug camptothecin*. Cancer Res, 1988. 48(7): p. 1722-6., Takimoto, C.H., L.V. Kieffer, and S.G. Arbuck, DNA topoisomerase I inhibitors. Cancer Chemother Biol Response Modif, 1997. 17: p. 80-113.
55. Hu XT, Chen W, Wang D, Shi QL, Zhang FB, et al. *The proteasome subunit PSMA7 located on the 20q13 amplicon is overexpressed and*

associated with liver metastasis in colorectal cancer. Oncol Rep 19: 441–6, 2008.

56. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. *Detection of large-scale variation in the human genome*. Nat Genet. 2004;36:949-51
57. Iida A, Kurose K, Isobe R, Akiyama F, Sakamoto G, Yoshimoto M, Kasumi F, Nakamura Y, Emi M. *Mapping of a new target region of allelic loss to a 2-cM interval at 22q13.1 in primary breast cancer*. Genes Chromosomes Cancer 1998; 21: 108-112
58. InternationalHapMapConsortium. *A haplotype map of the human genome.* Science, 437:1299{1320, 2005.
59. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, et al., *A second generation human haplotype map of over 3.1 million SNPs.*, Nature. 2007 Oct 18;449(7164):851-61
60. Iyer L, Ratain MJ.; *Clinical pharmacology of camptothecins*. Committee on Clinical Pharmacology, University of Chicago, IL 60637, USA.,Cancer Chemother Pharmacol. 1998;42 Suppl:S31-43
61. Jang TJ, Jung KH and Choi EA: *Id-1 gene downregulation by sulindac sulfide and its upregulation during tumor development, in gastric cancer*. Int J Cancer 118: 1356-1363, 2006
62. Jen J, Kim H, Piantadosi S, Liu ZF, Levitt RC, Sistonen P, Kinzler KW, Vogelstein B, Hamilton SR. *Allelic loss of chromosome 18q and prognosis in colorectal cancer.* N Engl J Med. 1994 Jul 28;331(4):213-21.
63. Jover, R.; Zapater, P.; Castells, A.; Llor, X.; Andreu, M.; Cubiella, J.; Balaguer, F.; Sempere, L.; Xicola, R.M.; Bujanda, L.; et al. *The efficacy of adjuvant chemotherapy with 5-fluorouracil in colorectal*

cancer depends on the mismatch repair status. Eur. J. Cancer 2009, 45, 365-373.

64. Kinzler KW, Vogelstein B., *Lessons from hereditary colorectal cancer.*; Cell. 1996 Oct 18;87(2):159-70
65. Korn WM, Yasutake T, Kuo WL, Warren RS, Collins C, et al., *Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization*. Genes Chromosomes Cancer 25: 82–90, 1999.
66. Kruglyak L, Nickerson DA. *Variation is the spice of life*. Nat Genet. 2001;27:234-6.
67. Lassmann S, Weis R, Makowiec F, Roth J, Danciu M, et al., *Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal- and microsatellite-unstable sporadic colorectal carcinomas*. J Mol Med 85: 293–304, 2007.
68. Leslie A, et al. *Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas*. Cancer Res 63:4656–4661 (2003).
69. Lievre A, Bachet JB, Boige V, et al. *KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab*. J Clin Oncol 2008;26:374−9
70. Lin JK, Chang SC, Wang HS, Jiang JK, Chen WC, Lin TC, Li AF: *Distinctive clinicopathological features of Ki-ras mutated colorectal cancers*. JSurg Oncol 2006, 94:234-41
71. Lockwood WW, Chari R, Chi B, Lam WLa (2006) *Recent advances in array comparative genomic hybridization technologies and their*

applications in human genetics. European Journal of Human Genetics 14:139–148.

72. Lu X, Zhang K, Van Sant C, Coon J, Semizarov D. *An algorithm for classifying tumors based on genomic aberrations and selecting representative tumor models.* BMC Med Genomics. 2010 Jun 22;3:23**.**
73. Machover, D., et al., *Two consecutive phase II studies of oxaliplatin (L-OHP) for treatment of patients with advanced colorectal carcinoma who were resistant to previous treatment with fluoropyrimidines*. Ann Oncol, 1996. 7(1): p. 95-8.
74. Malumbres M, Barbacid M. RAS *oncogenes: the first 30 years*. Nat Rev. 2003;3:7–13.
75. Misra A, Pellarin M, Nigro J, Smirnov I, Moore D, et al. *Array comparative genomic hybridization identifies genetic subgroups in grade 4 human astrocytoma*. Clin Cancer Res 11:2907–2918 (2005).
76. Miyakawa A, Wang XL, Nakanishi H, Imai FL, Shiiba M, Miya T, Imai Y, Tanzawa H. *Allelic loss on chromosome 22 in oral cancer: Possibility of the existence of a tumor suppressor gene on 22q13*. Int J Oncol 1998; 13: 705-709
77. Nishimine M, Nakamura M, Mishima K, Kishi M, Kirita T, Sugimura M and Konishi N: *Id proteins are overexpressed in human oral squamous cell carcinomas*. J Oral Pathol Med 32: 350-357, 2003
78. No authors listed, *Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer,* J Clin Oncol, 1992. 10(6): p. 896-903.

79. Norman Wolmark, Howard Rockette, Eleftherios Mamounas, Judy Jones, Sam Wieand, D.L. Wickerham, Harry D. Bear, James N. Atkins, Nikolay V. Dimitrov, Andrew G. Glass, Edwin R. Fisher, Bernard Fisher; *Clinical Trial to Assess the Relative Efficacy of Fluorouracil and Leucovorin, Fluorouracil and Levamisole, and Fluorouracil, Leucovorin, and Levamisole in Patients With Dukes' B and C Carcinoma of the Colon: Results From National Surgical Adjuvant Breast and Bowel Project C-04*; Journal of Clinical Oncology, Vol 17, Issue 11 (November), 1999: 3553-3559
80. Ohnishi S, Futamura M, Kamino H, Nakamura Y, Kitamura N, Miyamoto Y, Miyamoto T, Shinogi D, Goda O, Arakawa H., *Identification of NEEP21, encoding neuron-enriched endosomal protein of 21 kDa, as a transcriptional target of tumor suppressor p53*. Int J Oncol. 2010 Nov;37(5):1133-41.
81. Pagani F, Baralle FE. *Genomic variants in exons and introns: identifying the splicing spoilers*. Nat Rev Genet. 2004;5:389-96.
82. Pal M. *Proportionate increase in incidence of colorectal cancer at an age below 40 years: An observation*. J Can Res Ther 2006;2:97-9.
83. Peeters M., Siena S., Van Cutsem E. et al. *Association of progression-free survival, overall survival, and patients-reported outcomes by skin toxiticity and KRAS status in patients receiving panitumumab monotherapy.* Cancer. 2009;115(7):1544-54
84. Peters, G.J. and C.J. van Groeningen, *Clinical relevance of biochemical modulation of 5-fluorouracil*. Ann Oncol, 1991. 2(7): p. 469-80.
85. Pierre Laurent-Puig et al. *KRAS Mutation Signature in Colorectal Tumors Significantly Overlaps With the Cetuximab Response Signature*. Journal of clinical oncology, 2008

86. Pinkel D, Albertson DG *Array comparative genomic hybridization and its applications in cancer*. Nat Genet 37 Suppl:S11–S17 (2005)
87. Pollack JR, Srlie T, Perou CM, Rees CA, Jeffrey SS, et al. *Microarray analysis reveals a major direct role of dna copy number alteration in the transcriptional program of human breast tumors.* Proc Natl Acad Sci U S A 99:12963–12968 (2002)
88. Popat S, Hubner R, Houlston RS. *Systematic review of microsatellite instability and colorectal cancer prognosis*. J Clin Oncol. 2005; 23(3):609-18.
89. Poulogiannis G, McIntyre RE, Dimitriadi M, Apps JR, Wilson CH, Ichimura K, Luo F, Cantley LC, Wyllie AH, Adams DJ, Arends MJ. *PARK2 deletions occur frequently in sporadic colorectal cancer and accelerate adenoma development in Apc mutant mice*. Proc Natl Acad Sci U S A. 2010 Aug 24;107(34):15145-50. Epub 2010 Aug 9.
90. Powers, HJ. *"Interaction among folate, riboflavin, genotype, and cancer, with reference to colorectal and cervical cancer."* J Nutr (2005);135(12 Suppl.):2960S-2966S
91. Pozzo, C., et al., *Neoadjuvant treatment of unresectable liver disease with irinotecan and 5-fluorouracil plus folinic acid in colorectal cancer patients*. Ann Oncol, 2004. 15(6): p. 933-9.
92. Puppa G, Sonzogni A, Colombari R, Pelosi G. *TNM staging system of colorectal carcinoma: a critical appraisal of challenging issues*. Arch Pathol Lab Med. 2010 Jun;134(6):837-52.
93. Rajagopalan, H., Nowak, M.A., Vogelstein, B., and Lengauer, C. *The significance of unstable chromosomes in colorectal cancer.* Nat Rev Cancer 3, 695-701 (2003).
94. Raymond, E., et al., *Oxaliplatin: a review of preclinical and clinical studies*. Ann Oncol, 1998. 9(10): p. 1053-71.

95. Ribic CM, Sargent DJ, Moore DJ, et al. *Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer.* N Engl J Med 2003; 349: 247-57
96. Ridge, J.A. and J.M. Daly, *Treatment of colorectal hepatic metastase*s. Surg Gynecol Obstet, 1985. 161(6): p. 597-607.
97. Ried T, Knutzen R, Steinbeck R, Blegen H, Schrock E, Heselmeyer K, du MS, Auer G. *Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors.* Genes Chromosomes Cancer 15:234–245, 1996.
98. Rubio MP, Correa KM, Ramesh V, MacCollin MM, Jacoby LB, von Deimling A, Gusella JF, Louis DN. *Analysis of the neurofibromatosis 2 gene in human ependymomas and astrocytomas*. Cancer Res 1994; 54: 45-47
99. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al.; *International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms*. Nature. 2001;409:928-33.
100. Saltz, L.B., et al., *Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer*. Irinotecan Study Group. N Engl J Med, 2000. 343(13): p. 905-14.
101. Samowitz WS, Curtin K, Ma KN, Schaffer D, Coleman LW, Leppert M, Slattery ML. *Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level*. Cancer Epidemiol Biomarkers Prev. 2001:10(9):917-23

102. Santi, D.V., C.S. McHenry, and H. Sommer, *Mechanism of interaction of thymidylate synthetase with 5-fluorodeoxyuridylate*. Biochemistry, 1974. 13(3): p. 471-81.
103. Savelieva E, Belair CD, NewtonMA, DeVries S, Gray JW,Waldman F, Reznikoff CA. *20q gain associates with immortalization: 20q13.2 amplification correlates with genome instability in human papillomavirus 16 E7 transformed human uroepithelial cells*. Oncogene 14:551–560. 1997.
104. Sayagués JM, Fontanillo C, Abad Mdel M, González-González M, Sarasquete ME, Chillon Mdel C, Garcia E, Bengoechea O, Fonseca E, Gonzalez-Diaz M, De las Rivas J, Muñoz-Bellvis L, Orfao A., *Mapping of genetic abnormalities of primary tumours from metastatic CRC by high-resolution SNP arrays*. PLoS One. 2010 Oct 29;5(10):e13752.
105. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi, M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. *Large-scale copy number polymorphism in the human genome.* Science. 2004;305:525-8.
106. Seitz HK, Maurer B, Stickel F. *"Alcohol consumption and cancer of the gastrointestinal tract."* Digestive Diseases (2005);23(3-4):297-303
107. Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, Paty PB, Gerald WL, Notterman DA, Domany E. *Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer*. Proc Natl Acad Sci U S A. 2009 Apr 28;106(17):7131-6.

108. Shing DC, Trubia M, Marchesi F, et al. *Overexpression of sPRDM16 coupled with loss of p53 induces myeloid leukemias in mice*. J Clin Invest. 2007;117(12):3696-3707.
109. Singha PK, Yeh IT, Venkatachalam MA, Saikumar P. *Transforming growth factor-beta (TGF-beta)-inducible gene TMEPAI converts TGF-beta from a tumor suppressor to a tumor promoter in breast cancer*. Cancer Res. 2010 Aug 1;70(15):6377-83. Epub 2010 Jul 7.
110. Strachan T and Read AP. *Human Molecular Genetics 3*. 3rd edition. Garland Publishing, London and New York. 2004:317
111. Syvänen AC. *Accessing genetic variation: genotyping single nucleotide polymorphisms*. Nat Rev Genet. 2001;2:930-42.
112. Tanaka T, Watanabe T, Kazama Y, Tanaka J, Kanazawa T, et al., *Chromosome 18q deletion and Smad4 protein inactivation correlate with liver metastasis: A study matched for T- and N- classification*. Br J Cancer 95: 1562–7, 2006).
113. Tejpar S, Bosman F, Delorenzi M, et al. *Microsatellite instability (MSI) in stage II and III colon cancer treated with 5-FU-LV or 5FULV and Irinotecan (PETACC 3-EORTC 40993-SAKK 60/00 trial)*. Proc ASCO Meeting Abstract 2009; 27: 4001
114. Tirone F. *The gene PC3(TIS21/BTG2), prototype member of the PC3/BTG/TOB family: regulator in control of cell growth, differentiation, and DNA repair?* JCell Physiol 2001;187:155–65
115. Tuna M, Knuutila S, Mills GB: *Uniparental disomy in cancer*. Trends Mol Med 2009, 15:120-128.
116. Urban AE, Korbel JO, Selzer R, Richmond T, Hacker A, et al. *High-resolution mapping of dna copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays*. Proc Natl Acad Sci U S A 103:4534–4539 (2006)

117. Van Cutsem E., Kohne CH, Hiltre E. et al. *Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer*. N Engl J Med. 2009;360(14):1408-17.
118. Vanhoefer, U., et al., Irinotecan in the treatment of colorectal cancer: clinical overview. J Clin Oncol, 2001. 19(5): p. 1501-18.
119. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut. 2008 Jul;57(7):941-50.
120. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer. 2009; 9(7):489-99.
121. Zeng-Ren Zhao1, Zhi-Yong Zhang, Hong Zhang Li Jiang, Ming-Wei Wang And Xiao-Feng Sun; Overexpression of Id-1 protein is a marker in colorectal cancer progression; Oncology Reports 19: 419-424, 2008