Carcinogenesis vol.28 no.1 pp.38–48, 2007
doi:10.1093/carcin/bgl086
Advance Access publication June 13, 2006

Frequent occurrence of uniparental disomy in colorectal cancer

Claus Lindbjerg Andersen, Carsten Wiuf, Mogens Kruhøffer, Marianne Korsgaard, Søren Laurberg and Torben Falck Ørntoft*

[1]Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby DK-8200, Aarhus N, Denmark, [2]Bioinformatics Research Center (BiRC), University of Aarhus, DK-8000 Aarhus C, Denmark and [3]Department of Surgery, THG, Aarhus University Hospital, DK-8000, Aarhus C, Denmark

*To whom correspondence and requests for reprints should be addressed. Tel: +45 89495100; Fax: +45 89496018; Email: orntoft@ki.au.dk

We used SNP arrays to identify and characterize genomic alterations associated with colorectal cancer (CRC). Laser microdissected cancer cells from 15 adenocarinomas were investigated by Affymetrix Mapping 10K SNP arrays. Analysis of the data extracted from the SNP arrays revealed multiple regions with copy number alterations and loss of heterozygosity (LOH). Novel LOH areas were identified at chromosomes 13, 14 and 15. Combined analysis of the LOH and copy number data revealed genomic structures that could not have been identified analyzing either data type alone. Half of the identified LOH regions showed no evidence of a reduced copy number, indicating the presence of uniparental structures. The distribution of these structures was non-random, primarily involving 8q, 13q and 20q. This finding was supported by analysis of an independent set of array-based transcriptional profiles, consisting of 17 normal mucosa and 66 adenocarcinoma samples. The transcriptional analysis revealed an unchanged expression level in areas with intact copy number, including regions with uniparental disomy, and a reduced expression level in the LOH regions representing factual losses (including 5q, 8p and 17p). The analysis also showed that genes in regions with increased copy number (including 7p and 20q) were predominantly upregulated. Further analyses of the SNP data revealed a subset of the identified alterations to be specifically associated with TP53 inactivation (including 8q gain and 17p loss) and lymph node metastasis status (gain of 7q and 13q). Another subset of the identified alterations was shown to represent intratumor heterogeneity. In conclusion, we demonstrate that uniparental disomy is frequent in CRC, and identify genomic alterations associated with TP53 inactivation and lymph node status.

Abbreviations: CGH, comparative genomic hybridization; CRC, colorectal cancer; FDR, false discovery rate; LOH, loss of hetrozygosity; LMD, laser microdissection; *SGCZ,* sarcoglycan zeta; SNP, single nucleotide polymorphism; *TUSC3,* tumor suppressor candidate 3.

Introduction

The development and progression of colorectal cancer (CRC) is a multistep process leading to an accumulation of genomic alterations (1,2). The nature of these alterations varies from minute point mutations to gross chromosomal rearrangements. In particular, recurrent alterations are thought to be important for tumor pathogenesis (3). Several studies have focused on mapping genomic alterations occurring recurrently in CRC, commonly using genome-wide technologies like either comparative genomic hybridization (CGH) or loss of heterozygosity (LOH) analysis (4–6). These methods identify regions experiencing an altered DNA copy number or LOH. Importantly, results obtained by either method are negatively influenced by factors like normal cell contamination and intratumor heterogeneity. Normal cell contamination occurs because tumor tissues are only rarely homogenous, i.e. varying amounts of cells with normal DNA content (e.g. stromal and inflammatory cells) surround and infiltrate the cancer cells. In copy number analysis the signals from the normal cells potentially dilute out the signals from the cancer cells, and in LOH analysis polymorphic makers might appear heterozygous, because of normal cell admixture, despite the cancer cells being homozygous. Intratumor heterogeneity is important because the genomic alterations found in one part of the tumor may be different from alterations found in another part, thus making the alterations identified depended on the site of biopsy. In spite of these problems several genomic regions have been reported to be recurrently altered, including 5q, 8p, 8q, 17p, 18 and 20q (4–9). The order in which these alterations appear during tumorigenesis is a subject of debate (7–9). One reason why this subject receives attention is because of the potential use of genomic alterations as markers for clinically important stages in disease course, e.g. markers of lymph node involvement or TP53 tumor suppressor gene mutation.

Uniparental disomy is a type of genomic alteration that until very recently has not received much attention. Uniparental disomy is caused by loss of one allele and gain of the remaining allele, giving raise to LOH but no copy number alteration. The specific mechanisms underlying the emergence of uniparental structures remain to be identified; however, precedents have been described for the emergence of uniparental disomy through mitotic recombination, non-disjunction or deletion and re-duplication events (10–13).

Both LOH and copy number information are needed for detection of uniparental disomy. In most previous studies of genomic alterations, the applied technologies have only provided one of these data types, practically causing uniparental disomy to be neglected. With the recent development of single nucleotide polymorphism (SNP) arrays, which provide both genotype (LOH) and copy number information, this situation has now changed and uniparental structures have

been observed in acute myeloid leukemias, medulloblastomas and basal cell carcinomas (14–16).

A recent study of genomic alterations in CRC cell lines applied SNP arrays in conjunction with CGH arrays (17). The data revealed several genomic regions experiencing uniparental disomy and importantly indicated that at least in the cell lines these occurred in a non-random pattern. Whether this also holds true in clinical CRC specimens needs to be verified.

In a recent study by Tsafrir and coworkers (8), combined analysis of genomic and gene expression alterations in CRC revealed a correlation between gene expression level and DNA content; in particular in relation to losses involving 1p, 4, 5q, 8p, 14q, 15q and 18 and gains involving 7p, 8q, 13q and 20q. Unfortunately, this study did not include investigations of uniparental disomies and thus it still remains to be clarified how this type of genomic alteration affects gene expression.

In the present article we applied SNP arrays and gene expression arrays to clinical CRC specimens in order to address the following four issues: (i) to identify and characterize genomic alterations (including uniparental disomy) associated with the development and progression of CRC; (ii) to investigate whether specific alterations are specifically associated with clinical important parameters like lymph node involvement and TP53 mutation; (iii) to investigate the extent of genomic intratumor heterogeneity in colorectal adenocarcinomas; and finally (iv) to investigate the relationship between the specific types of genomic alterations and the expression of the affected genes.

Materials and methods

Patient material

Fifteen patients diagnosed with left-sided colorectal adenocarcinomas were included in the study. Tumor and adjacent normal mucosa biopsies were collected directly after surgical resection at Aarhus University Hospital, Denmark. These were immediately embedded in Tissue-Tek O.C.T Compound (Sakura Prohosp, DK) and snap-frozen in liquid nitrogen. A summary of the histopathological characteristics are given in Table I. For 13 of the 15 patients a blood sample was collected as well. All patients gave informed consent, and the study was approved by the local Scientific Ethical Committee. With the purpose of generating a set of references for the copy number analysis we collected germline DNA (from blood) from additional 98 individuals.

Six of the fifteen patients were included in a subinvestigation of intratumor heterogeneity, using whole tumor cross sections. For each patient a minimum of two cancer cell subpopulations from the luminal surface and invasive front were procured by laser microdissection (LMD). A total of 17 tumor areas were dissected (Supplementary Table I).

Microdissection

Crude dissection. From Tissue-Tek embedded tumor tissue fifteen 20 μm thick sections were cut and mounted. Slides were stained with hematoxylin. The fraction of cancer cells was enriched by scraping off non-cancerous tissue parts with a scalpel.

LMD. From Tissue-Tek embedded tissue 5 μm thick sections were cut and mounted on membrane slides (PALM, Bernried, Germany). Slides were stained with hematoxylin. LMD was performed with a LMD and pressure catapulting microscope (PALM).

DNA extraction

Matched cancer and germline DNA was extracted using the PUREGENE DNA extraction system (Gentra SYSTEMS, Minneapolis, MN, USA) according to the manufacturer's instructions. Cancer DNA was extracted from LMD procured cancer cells, while germline DNA was extracted from blood. In two cases blood was not available and germline DNA was extracted from normal colon mucosa biopsies.

Genechip Mapping 10K early access array analysis

The Single Primer Assay Protocol (labeling, hybridization, washing, staining and scanning) was performed according to the manufacturers instructions (Affymetrix, Santa Clara, CA, USA) (18–20). Unless stated otherwise, the used cancer DNA was extracted from nearly 100% pure cancer cells procured by LMD. A total of 139 samples were analyzed (30 samples for LOH and copy number analysis—matched germline and cancer DNA from 15 patients—an additional 11 cancer samples for heterogeneity analysis and an additional 98 germline DNA samples). The 113 germline samples (15 + 98) were used as reference samples in the copy number analysis.

Mapping and exclusion of SNPs

The physical position of all SNPs ($n = 10043$) on the Mapping 10K Array were mapped according to the April 2003 genome assembly. SNPs that did not map or mapped to more than one position in the genome assembly were excluded from the analysis. To avoid gender-related complications SNPs mapping to chromosome X were also excluded. A total of 606 SNPs were excluded leaving 9437 uniquely mapped SNPs for further analysis.

Table I. Summary of patients and tumors[a]

Patient ID	Age[b]	Tumor location	Histology	Grade of differentiation	pT-stage	pN-stage	pM-stage	Dukes	10K mapping array	Heterogeneity study	Sequencing	
											TP53	CLU
474	52	Rectosigmoid	Ac	1	1	0	0	A	x	n.d.	Wt	n.d.
305	89	Sigmoid	Ac	2	2	0	0	A	x	n.d.	Wt	Wt
336	69	Rectum	Ac	3	2	0	0	A	x	n.d.	Mut	Wt
368	72	Rectum	Ac	2	2	0	0	A	x	n.d.	Mut	Wt
431	74	Rectosigmoid	Ac	2	2	0	0	A	x	n.d.	Mut	Wt
504	75	Sigmoid	Ac	2	2	0	0	A	x	Hetero	Mut	Wt
496	61	Decendens	Ac	2	3	0	0	B	x	Hetero	Mut	Wt
497	85	Sigmoid	Ac	2	3	0	0	B	x	Hetero	Mut	Wt
505	72	Sigmoid	Acm	2	4	0	0	B	x	Homo	Wt	n.d.
347	59	Rectosigmoid	Ac	2	2	1	0	C	x	n.d.	Wt	n.d.
281	79	Rectosigmoid	Ac	2	3	1	0	C	x	n.d.	Mut	Wt
326	89	Sigmoid	Ac	3	3	1	0	C	x	n.d.	Mut	n.d.
469	57	Sigmoid	Ac	2	3	1	0	C	x	Hetero	Mut	n.d.
531	78	Sigmoid	Ac	2	3	1	0	C	x	Homo	Mut	Wt
304	55	Rectosigmoid	Ac	2	2	2	0	C	x	n.d.	Wt	Wt

Ac, adenocarcinoma; Acm, mucinous adenocarcinoma; Hetero, intratumor heterogeneity observed; homo, homogenous tumor with no intratumor heterogeneity observed; Mut, mutated; n.d., not done; Wt, wild-type.
[a]All tumors were sporadic and microsatellite stable.
[b]Age at surgery.

LOH analysis

Genotype information was extracted from the Mapping 10K Arrays using the Genechip DNA Analysis Software (GDAS) from Affymetrix. Genotypes derived from germline and cancer DNA were loaded into the software package dChip (http://www.dchip.org/) which was used for LOH analysis.

LOH calls were inferred using a method described previously (21). Basically, a hidden Markov model (HMM) was applied to statistically infer retention or loss for all SNPs using the genotype calls of matched germline and cancer DNA.

Mapping of minimal genomic regions commonly showing LOH

The frequency of inferred LOH (in 15 patients) was calculated for all SNPs ($n = 9437$). SNPs with LOH frequency >25% were defined as SNPs commonly showing LOH. Within an uninterrupted sequence of SNPs commonly showing LOH, the minimal common genomic region was defined as the uninterrupted segment of SNPs with the highest LOH frequency in the sequence.

Intratumor heterogeneity analysis

We define intratumor heterogeneity to be the case when a genomic region is lost or retained in a cancer cell subpopulation but not in another from the same tumor. Intratumor heterogeneity was assessed using LOH data from 17 LMD procured cancer samples from six patients. The method and the statistics applied to infer genomic regions showing intratumor heterogeneity is described in detail in the supplementary information. Briefly, the regions that show disagreement in observed LOH pattern in two samples from the same tumor were outlined (i.e. SNPs that are observed to be lost/retained in one of the samples but not the other). Then a statistical permutation test was applied to evaluate the likelihood of observing such regions by chance.

Extraction of weighted signal intensity values

Normalization. 128 arrays (15 CRC and 113 germline samples) were normalized and signal values for the individual SNPs were extracted as described previously (22). Owing to different physical properties of the individual probes on the array, signal values from different SNPs are not directly comparable. To make them comparable the dataset were normalized SNP-wise using the mean and standard deviation (SD) of the germline samples, i.e. $z_{ij} = (x_{ij} - \text{mean}_j)/\text{SD}_j$ where x_{ij} is the observed signal of SNP j in sample i (germline or tumor), and mean_j and SD_j are mean and standard deviation of SNP j for all 113 germline samples. As a consequence, the mean and SD of z_{ij} is 0 and 1, respectively, for all the SNPs in the germline samples. To further reduce the noise level in the signal we use the average of $M = 11$ SNPs weighted by genomic distance; i.e. $a_{ij} = \Sigma z_{i(j+l)} \exp(-d_{j(j+l)})/\Sigma \exp(-d_{j(j+l)})$ where the sum (Σ) is over the five neighboring SNP on either side of SNP j ($l = -5, -4, \ldots, 0, \ldots, 4, 5$) and $d_{j(j+l)}$ is the genomic distance between SNP j and $j + l$. For simplicity we refer to a_{ij} as the signal intensity.

Mapping of genomic regions commonly showing copy number alterations

Genomic regions commonly showing copy number alterations were identified as segments of SNPs ($\geq$10 SNPs) for which the average signal intensity (over all tumor samples) was significantly different from the average signal intensity of 15 germline samples ($P \leq 0.00001$). The null distribution of germline samples was obtained by randomly drawing 15 samples among the 113 germline samples 100 000 times.

Transcriptional profiling

The transcriptional profiling data used in the present article is a subset of a larger dataset published previously by our group (23). Briefly, total RNA from 17 normal mucosa samples and 66 adenocarcinomas of the colorectum were transcriptionally profiled using Affymetrix HG-U133A GeneChip arrays. Importantly, the adenocarcinomas (18 Dukes B and 48 Dukes C) were selected according to the same criteria as the tumors investigated by SNP arrays (originating from sporadic, microsatellite stable and localized tumors). However, it was not possible to make a complete stage match, but the Dukes B an C samples were the majority in both sets, and no Dukes D's, that may be necrotic, were included. The two sets were completely independent. The expression data were GC_RMA normalized using ArrayAssist Software (Stratagene, La Jolla, CA).

Correlation of gene expression and genomic alterations

Genes mapping to regions displaying genomic alterations, by LOH and copy number analysis, were identified by querying the RefSeq database (24) (RefSeq mapped to the April 2003 genome assembly) with the physical boundaries of the regions. RefSeq IDs mapping to more than one location in the genome were excluded. We report individual genes by their Entrez Gene ID (25).

The expression patterns of the identified genes were evaluated using array-based transcriptional profiles of 17 normal mucosa and 66 adenocarcinoma samples. Profiles of genes either not expressed or generated by non-functional probesets were excluded from analysis. Upregulated and downregulated genes were identified by comparing the expression values (log scale, base 2) of the tumors and normals. The statistical significance of the expression differences was evaluated using *t*-tests with a significance level of 5%, and genes were labeled upregulated, downregulated or unchanged accordingly. To justify the significance level we estimated the false discovery rate (FDR) (no. of false positives/ no. of positives) and found FDR $\approx$2% for this dataset (26). This value of FDR was considered reasonable. To test whether the gene expression patterns observed for the altered genomic regions were random or associated with alterations, they were compared with gene expression patterns of genomic regions without alterations. These neutral genomic regions were identified based on the following criteria. For LOH, neutral regions were defined as genomic regions containing stretches of SNPs for which no more than one tumor shows LOH; and for copy number, neutral regions were defined as genomic regions containing stretches of SNPs for which the average signal intensities were within the ±5% significance level of the reference samples. This level is fairly conservative excluding all regions (SNPs) likely to have abnormal copy numbers. The identified neutral regions were divided into subregions of the same size as the LOH or copy number regions. Only the subregions with gene density within 70–140% of the density in the reported regions were maintained for further analysis.

We then calculated the proportion of times the numbers of upregulated and downregulated genes in the neutral regions were higher than in the regions reported.

Genomic copy number determination using real-time PCR

Real-time PCR was performed on an ABI PRISM 7000 Sequence Detection System (Applied Biosystems, Foster City, CA) using the SYBR GREEN PCR Master Mix (Applied Biosystems). Quantification of the gene sarcoglycan zeta (*SGCZ*) located at 8p was based on dilution curves constructed using serially diluted germline DNA (four dilution points, each of 10-fold). The copy number of SGCZ was determined relative to a reference Line-1 repetitive element with similar copy numbers per haploid genome in normal and malignant cells, and normalized by using normal genomic DNA as calibrator, a method described previously by others in detail (22). All reactions were done in triplicate using the following PCR conditions: 50°C for 2 min; 95°C for 10 min; 40 cycles of 95°C for 15 s; and 60°C for 1 min. Primers specific to Line-1 are forward 5′-AAAGCCGCTCAACTACATGG-3′ and reverse 5′-TGCTTTGAATGCGTCCCAGAG-3′ [the primers for Line-1 has been described previously (22)]. Primers spanning a 106 bp non-repetitive region of *SGCZ* for *SGCZ* was designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) forward 5′-GATTACCATTGGGGCTGAAAAG-3′ and reverse 5′-AAGAAGCTCCTTGTGCAGTTGT-3′. The specificity of the SGCZ primer set was validated by electronic PCR as well as BLAT searches against the complete human genome (http://genome.ucsc.edu). Furthermore, melting analysis of real-time PCR end product along with agarose gel electrophoresis confirmed the primer set to generate only a single amplicon.

Real-time RT–PCR

Quantitative real-time RT–PCR for the target genes SGCZ and TUSC3 was performed on an ABI PRISM 7000 Sequence Detection System (Applied Biosystems). For normalization the gene *UBC* was used. We have previously demonstrated the suitability of *UBC* as a normalization gene for analysis of normal mucosa and CRC specimens sample sets (27). Total RNA from 12 normal mucosa and 14 adenocarcinoma samples were investigated; hereof 10 samples represented matched normal mucosa and adenocarcinoma from five patients. RNA integrity was evaluated using RNA integrity numbers (RIN) generated on a Bioanalyzer instrument (Agilent Technologies, Palo Alto, CA). Only samples with high RNA quality were included, mean RIN was 7.9, minimum was 6.6 and maximum was 9.6. The target genes were investigated using pre-designed TaqMan Gene Expression Assays (Applied Biosystems). For SGCZ assay ID Hs00292381_m1 was used and for TUSC3 assay ID Hs00185147_m1 was used. The normalization gene *UBC* was investigated using a SYBR green assay. The primers for the *UBC* assay have been published previously (27). The assays were performed basically as described by the manufacturer, except for smaller reaction volumes of 25 µl. Each measurement was performed in triplicate and no-template controls were included for each assay. Relative expression values were obtained using a dilution curve consisting of four 10-fold serial dilution points. The dilution curve was created using a cDNA pool constructed by pooling 2 µl of each of the 26 test cDNAs.

Sequencing

The genomic integrity of TP53 was analyzed by bidirectional sequencing using the BigDye Terminator Kit (Applied Biosystems) and the ABI 3100 genetic analyzer (Applied Biosystems). Primer details can be found in the Supplementary Table 2. A two-phased sequencing approach was applied. In phase 1 the mutation hotspot covering exons 5–8 were sequenced. If no mutation was identified in phase 1 then phase 2 was initiated entailing sequencing of the remaining coding exons 2–4 and 9–11.

Genomic differences between tumor subgroups

Genomic differences between tumor subgroups, defined by either TP53 mutation status or lymph node metastasis status were identified based on both copy number and LOH data. The statistical approaches applied are described in detail in the supplementary information. In brief, the difference between tumor subgroups were calculated for each SNP based on either signal intensities (copy number) or frequency of LOH. The significance of the observed differences was evaluated by permutation tests.

Results

Comparison of crude dissection and LMD

Contamination of DNA from cancer cells with DNA from normal cells influences our ability to generate correct cancer genotypes, and thus our ability to identify LOH. In order to evaluate the importance of sample purity for accurate identification of LOH, cancer cells from the tumor of patient 336 were dissected using both a crude dissection protocol and a more precise LMD protocol. We found 88 of 3087 SNPs to have LOH in the crude sample and 228 SNPs in the laser microdissected sample. Thus, LMD was quite superior for identification of LOH and, although laborious, was used for all samples in this study.

Mapping of minimal genomic regions commonly showing LOH

Genome-wide LOH information was obtained for the 15 cancer samples by comparing the genotypes of 9437 SNPs extracted from LMD cancer DNA and matched germline DNA. A search for the genomic regions most frequently $\geq$27% ($\geq$4 of 15 cases) displaying LOH revealed 20 regions (Table II). These regions involved 12 different chromosomes of which chromosomes 13, 14 and 15 have not previously been reported to show frequent LOH in CRC. The 9437 SNPs provided good coverage of the genome and, thus, sufficient resolution to enable identification of small LOH regions, in particular the regions with IDs 4–8, 14 and 18–19 involving chromosomes 8, 15 and 20. The identified minimal LOH regions ranged in size from 0.67 to 33.08 Mb. A survey of the physical positions of all known human genes revealed that the gene content of the identified regions varied from one gene up to 359 genes.

Table II. Identification of common genomic alterations in CRC: minimal genomic regions commonly showing LOH[a]

Region ID	Chr. band	Position in HG15[b]		Region size (Mb)	LOH Frequency[c]	Correlates with reduced copy number[d]	Genes[e]	No. of Expressed genes[f]	Dysregulated genes[g]		Notes
		Start	End						Down	Up	
1	3p25–p24	11.38	22.43	11.05	33	No	46	20	5	9	Not incl. MLH1
2	4q34–q35	172.85	191.30	18.45	27	Yes	50	24	10	5	
3	5q15–q22	97.10	114.67	17.57	87	Yes	35	15	5	2	Incl. APC
4	8p22	13.30	15.45	2.15	67	Yes	2	0	0	0	
5	8q21	87.84	91.40	3.56	27	No	8	4	0	2	
6	8q22–q23	104.39	106.99	2.60	27	No	6	1	1	0	
7	8q24	121.81	127.79	5.98	27	No	24	12	0	8	
8	9p21	26.49	29.69	3.21	27	Yes	9	5	3	1	
9	13q12–q14	30.84	43.41	12.57	27	No	48	25	3	17*	Not incl. RB1
10	13q21–q22	60.34	77.53	17.20	27	No	21	12	2	8*	Not incl. RB1
11	14q11–q21	18.48	42.64	24.16	40	Yes	143	85	17	24	
12	14q22–q32	50.75	102.38	51.63	40	Yes	283	160	53	45	
13	15q11–q21	19.70	56.78	37.08	47	Yes	196	98	27	32	
14	15q25	76.40	81.42	5.03	47	No	31	16	3	5	
15	17p13–p12	0.90	15.62	14.72	80	Yes	194	88	30*	16	Incl. TP53
16	18p11	0.24	12.54	12.30	60	Yes	44	27	16*	3	
17	18q21–q23	47.11	76.90	29.79	60	Yes	83	36	21*	2	Incl. SMAD4 & DCC
18	20p13	0.10	1.22	1.12	47	No	17	9	1	6	
19	20p11	20.26	20.93	0.67	53	No	1	0	0	0	
20	22q11–q13	15.69	48.76	33.08	27	No	359	189	63*	56	

[a]Only regions showing LOH in $\geq$25% of cases are listed.
[b]The positions are defined by the SNPs at the boundary of the region, and given as the distance in megabases (Mb) from the p-terminus.
[c]The frequency was calculated from LOH calls inferred by dChip and is listed in %.
[d]The number of SNPs showing loss, no change and increase in copy number was counted for each region. The thresholds for calling loss or increase were the 0.5% significance levels displayed in supplementary figure 2. Correlation of LOH and loss (LOH = loss) is reported when >60% of the SNPs in the region have loss calls.
[e]The number of genes (Entrez Gene ID's) mapping to the region.
[f]The number of genes for which at least a single probeset on the U133A GeneChip shows a dynamic expression pattern across the 83 expression arrays included in the study. Thereby excluding genes either not expressed or whose expression data were generated using non-functional probesets.
[g]The number of genes (expressed) in the given region whose expression was significantly different ($P < 0.05$) in the adenocarcinomas ($n = 66$) than in the normal mucosa ($n = 17$).
*Marks the situations where the observed number of deregulated genes (up or down) were significantly different ($P < 0.05$) from what was observed in regions of genome not displaying LOH.

Mapping of genomic regions commonly showing copy number alterations

In addition to the genotype information used for LOH analysis, the signal intensities of the SNP array also carries information on DNA copy numbers in the investigated sample. The average DNA copy numbers of the 15 investigated CRC tumors differed in many regions significantly from the average copy numbers of 113 germline DNAs used as a reference. This implies that specific regions of the tumor genomes commonly carry aberrant DNA copy numbers (for an illustrative overview of the average copy numbers of the tumors see Supplementary Figure 1). Thirty-seven genomic regions stood out when we applied stringent criteria to both the significance of the observed copy number and the number of consecutive SNPs with this significance (Table III). These regions distribute across 13 chromosomes. The 9437 SNPs provided good coverage of the genome and, thus, sufficient resolution to enable identification of small regions, ranging in size from 1.73 to 19.4 Mb. The gene content of the regions varied from 0 to 108 candidate genes.

In order to validate the above copy number data we determined the copy number of the genomic loci for the gene *SGCZ*, located at chromosome arm 8p, by quantitative real-time PCR. *SGCZ* is located in a region showing both copy number loss (Table III, region ID no. 15) and LOH (Table II, region ID no. 4) in 10 out of 15 samples. For the

Table III. Identification of common genomic alterations in CRC: genomic regions with altered DNA copy numbers[a]

Region ID	Chr. band	Position in HG15[b]		Region size (Mb)	Gain/loss	No. of SNPs in region	Genes[c]	No. of expressed genes[d]	Dysregulated genes[e]		Notes
		Start	End						Down	Up	
1	1p31	78.03	81.65	3.62	Loss	23	5	3	0	0	
2	1p21	101.58	106.15	4.57	Loss	13	4	1	0	1	
3	4p15	11.38	15.16	3.77	Loss	34	4	1	1	0	
4	4p15	19.78	23.75	3.97	Loss	18	6	3	3*	0	
5	4p15	29.02	35.61	6.59	Loss	16	1	1	1	0	
6	4q28	134.92	138.78	3.86	Loss	16	0	0	0	0	
7	4q32	156.28	158.91	2.63	Loss	10	9	5	1	0	
8	4q32	161.36	167.70	6.34	Loss	28	15	3	0	1	
9	4q33–q34	172.24	174.17	1.94	Loss	14	0	0	0	0	
10	4q34	180.14	182.91	2.77	Loss	12	0	0	0	0	
11	5q12	58.88	62.25	3.37	Loss	12	10	3	1	2	Not incl. APC
12	5q14	83.61	89.41	5.80	Loss	22	6	5	3	2	Not incl. APC
13	7p15	24.12	26.14	2.02	Gain	10	10	7	0	4	
14	8p23	2.55	8.85	6.30	Loss	41	17	5	1	2	
15	8p23–p21	10.48	18.87	8.38	Loss	37	36	14	6	3	
16	8p21	23.17	26.20	3.03	Loss	11	16	8	5*	1	
17	8p21–p12	27.14	28.87	1.73	Loss	15	17	11	4	3	Incl. CLU
18	8p12	30.13	34.14	4.01	Loss	22	13	9	3	3	
19	10p14	8.38	10.65	2.27	Loss	14	0	0	0	0	
20	10q21	54.40	56.79	2.39	Loss	20	2	0	0	0	
21	13q12	25.97	28.87	2.90	Gain	13	14	6	0	5*	
22	14q21	38.14	41.33	3.19	Loss	11	1	0	0	0	
23	14q21	44.59	47.37	2.77	Loss	10	1	0	0	0	
24	14q24–q31	76.84	78.63	1.79	Loss	11	1	1	0	0	
25	14q31	81.72	86.36	4.64	Loss	35	1	1	0	0	
26	15q21	51.26	53.07	1.81	Loss	16	1	1	0	0	
27	17p13–p11	9.76	20.11	10.35	Loss	38	77	35	13	5	Not incl TP53
28	18p11	4.12	5.95	1.83	Loss	11	3	1	1	0	
29	18q11–q12	22.01	41.41	19.40	Loss	73	39	20	9*	4	
30	18q21	48.16	52.96	4.79	Loss	18	14	7	5*	0	Incl. SMAD4 and DCC
31	18q21	57.26	59.47	2.21	Loss	12	4	1	0	1	
32	18q21–q22	60.67	71.64	10.97	Loss	48	26	7	5	1	
33	18q22–q23	72.65	75.76	3.11	Loss	11	7	2	1	0	
34	20q11–q12	30.54	40.00	9.45	Gain	18	108	61	6	49*	
35	20q12–q13	40.87	52.50	11.63	Gain	30	99	51	6	32*	
36	20q13	54.96	62.58	7.62	Gain	21	67	39	7	29*	
37	21q21	22.27	25.26	2.99	Loss	20	1	0	0	0	

[a]Listed are regions with $\geq$10 consecutive SNPs displaying a significant difference in the average intensity of the adenocarcinoma samples compared with germline samples ($P < 0.00001$) . The multiple regions listed for chromosomes 4, 8, 14 and 18 are not necessarily independent and distinct regions, but could very well be parts of larger regions (Supplementary Figure 1).

[b]The positions are defined by the SNPs at the boundary of the region, and given as the distance in megabases (Mb) from the p-terminus.

[c]The number of genes (Entrez Gene ID's) mapping to the region.

[d]The number of genes for which at least a single probeset on the U133A GeneChip shows a dynamic expression pattern across the 83 expression arrays included in the study. Thereby excluding genes either not expressed or whose expression data were generated using non-functional probesets.

[e]The number of genes (Expressed) in the given region whose expression was significantly different ($P < 0.05$) in the adenocarcinomas ($n = 66$) than in the normal mucosa ($n = 17$).

*Marks the situations where the observed number of deregulated genes were significantly different from what was observed in regions of genome not displaying copy number alterations.

10 samples showing 8p loss laser microdissected cancer cells and matched germline cells were investigated. Four samples of laser microdissected normal cells were also investigated. The results showed an average copy number of 0.72 ± 0.3 (mean ± SD) ($n = 10$) in the cancer cells as compared with 2.0 ± 0.2 ($n = 10$) in germline cells (blood) and 1.8 ± 0.1 ($n = 4$) in normal cells (the data are shown in supplementary material), thereby confirming the copy number losses found by SNP arrays.

Correlation of LOH and copy number alterations unveils uniparental disomy as a common event in CRC

The LOH and genomic copy number data extracted from the 10K mapping arrays often supported and complemented each other, as illustrated for chromosome 20 in Figure 1. Comparison of the LOH and copy number data revealed genomic regions where LOH correlates positively to copy number loss, as well as regions where it correlates positively to two copies or even an increased copy number (Figure 1 and Supplementary Figure 2). The affected regions covered whole chromosomes, chromosome arms and interstitial chromosome segments.

Of the 20 minimal genomic regions commonly showing LOH, we found half to be positively correlated with genomic loss, representing factual losses, and the other half to be positively correlated with copy numbers of two or higher, probably representing uniparental disomy and polysomy (Table II). Our data indicate that the generation of regions with uniparental disomy occurs in non-random fashion primarily affecting chromosome arms 3p, 8q, 13q and 20q. Uniparental di/polysomy was predominantly observed in regions with low LOH frequencies, whereas factual losses were predominantly observed in regions with high LOH frequencies, indicating that uniparental di/polysomy occurs less frequently than factual losses in CRC (Table IV).

CRCs often show intratumor heterogeneity

In order to reveal genomic intratumor heterogeneity we performed a genome-wide search for genomic alterations in cancer cell subpopulations laser microdissected from the luminal surface and the invasive front of tumors from six patients (Figure 2A and Table I). Our statistical analysis of the genotypes derived from the individual cancer cell subpopulations revealed four of six tumors to show intratumor heterogeneity at the LOH level (Figure 2B and Supplementary table 3). None of the genomic regions commonly associated with CRC (e.g. 5q, 8p, 17p and 18) displayed intratumor heterogeneity. Generally, the regions in which we observed intratumor heterogeneity in specific tumors displayed LOH very infrequently when evaluated across the whole set of tumors (observed in <25% of tumors).

Intratumor heterogeneity was not evaluated using copy number alterations.

Genomic alterations seem to impose significant changes on gene expression

Genomic alterations are thought to exert their oncogenic effects by imposing a dysregulated expression pattern on one or more target genes located within the affected region. To search for the target genes of the genomic alterations listed in Tables II and III we therefore decided to correlate with gene expression patterns. As we did not have expression data on the same tumors as SNP data we used an independent set of

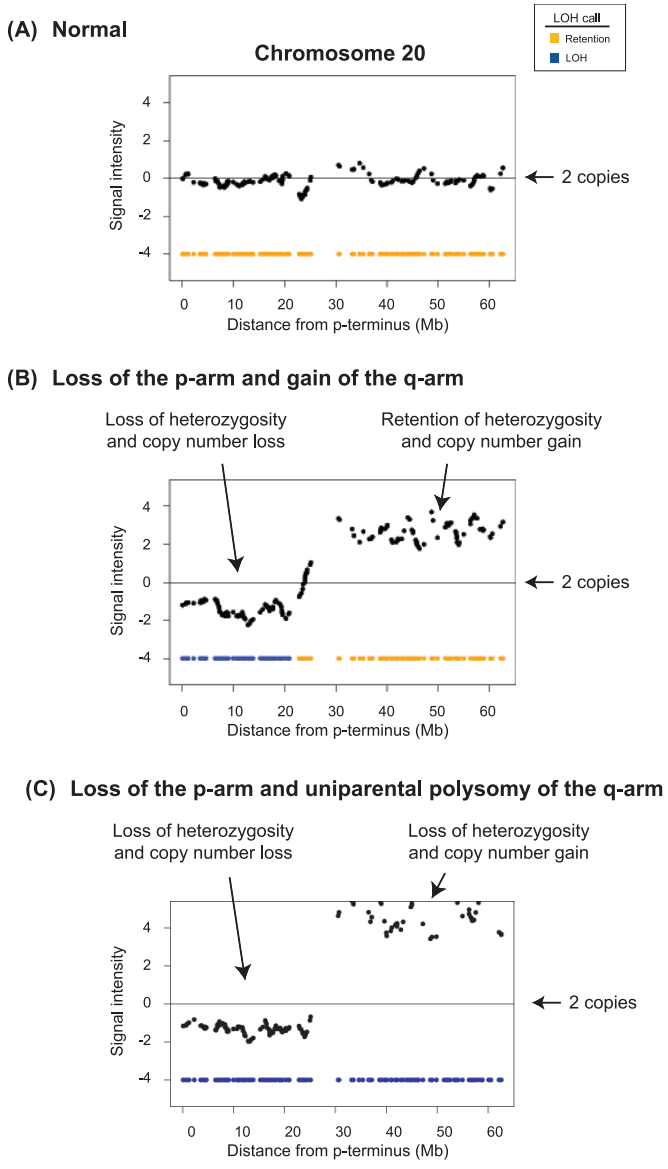

Fig. 1. Correlation between LOH and copy number data. Shown are three examples of how the DNA copy number and LOH data support and complement each other. Plotted are chromosome 20 data for three patients. (**A**) A patient with a normal chromosome 20; (**B**) a patient with factual loss of the p-arm (i.e. LOH concomitant with copy number reduction) and gain of the q-arm (i.e. retention of both alleles and simultaneous copy number increase); (**C**) a patient with factual loss of the p-arm and uniparental polysomy of the q-arm (i.e. LOH and concomitant copy number increase of the remaining allele). The copy number data are visualized SNP-wise by signal intensity. An intensity of 0 corresponds to two copies. The LOH data are visualized SNP-wise by LOH calls. Orange corresponds to retention and dark blue to LOH.

samples. This makes a direct comparison impossible; however, we believe that the general findings detected with one technique should also be detected with the other technique if the sample size is suitable. We therefore searched array-based transcriptional profiles of an independent set of 17 normal mucosa and 66 adenocarcinoma samples for differentially expressed genes mapping to the altered genomic regions. The number of genes differentially expressed in adenocarcinomas compared with normal mucosa is listed in Tables II and III. Complete lists of the upregulated

Table IV. Correlation between LOH frequency and genomic copy numbers

		Number of regions with		Sum
		LOH = loss	LOH ≠ loss	
Regions with LOH frequency	High*	8	3	11
	Low*	2	7	9
	Sum	10	10	20

For each region in Table II the number of SNPs showing loss, no change and increase in copy number was counted. The thresholds for calling loss or increase were the 0.5% significance levels displayed in Supplementary Figure 2. Correlation of LOH and loss (LOH = loss) is reported when >60% of the SNPs in the region have loss calls.
Note that 9 of the 10 regions without correlation had no SNPs showing loss, the last region had 33% loss calls.
* The high/low LOH frequency threshold was set to 40% in order to get a similar number of regions in the two groups.

and downregulated genes can be found in Supplementary tables 4 and 5.

Irrespective of the type of genomic alteration investigated (LOH, increased or reduced copy number) we found most of the aberrant regions to contain multiple dysregulated genes, both upregulated and downregulated (Tables II and III). This showed that not all genes in the affected regions respond in the same manner and at the same time raised the question: 'Is there at all a relationship between the type of genomic alteration and the orientation of the potentially imposed expression change?' We addressed this question by comparing the overall expression patterns observed in the different types of genomic alterations (LOH regions representing uniparental disomies, LOH regions representing factual losses, copy number losses and copy number increases) with that of neutral genomic regions (without alterations). The analysis revealed that all the different alteration types, except uniparental disomies, displayed overall expression patterns significantly different (P-values <0.000001) from that of neutral regions (Table V). First of all, this finding showed that the different types of alterations impose characteristic expression changes on the majority of genes in the affected regions. Second, it indicated that the expression of genes located in regions of uniparental disomy remained largely unaffected, which is expected because the gene dosage is unchanged. However, many reports on LOH have not taken this aspect into consideration.

Quantitative expression analysis of SGCZ and TUSC3 by real-time RT–PCR

Our LOH and copy number data analysis pinpointed a specific 2.15 Mb region located 13.3 Mb from the p-terminus of chromosome 8 displaying both LOH and copy number loss in 67% of cases. Importantly, the region only contains two known genes, *SGCZ* and tumor suppressor candidate 3 (*TUSC3*). The relation between these genes and CRC are poorly described. Probe sets for neither *SGCZ* nor *TUSC3* were on the expression array we used for transcriptional profiling. In order to identify whether these genes were differentially expressed in adenocarcinomas, we therefore performed quantitative real-time RT–PCR. Twelve normal mucosa and 14 adenocarcinoma samples were investigated. The results revealed *TUSC3* to be significantly downregulated in the adenocarcinomas (P = 0.027). On average the

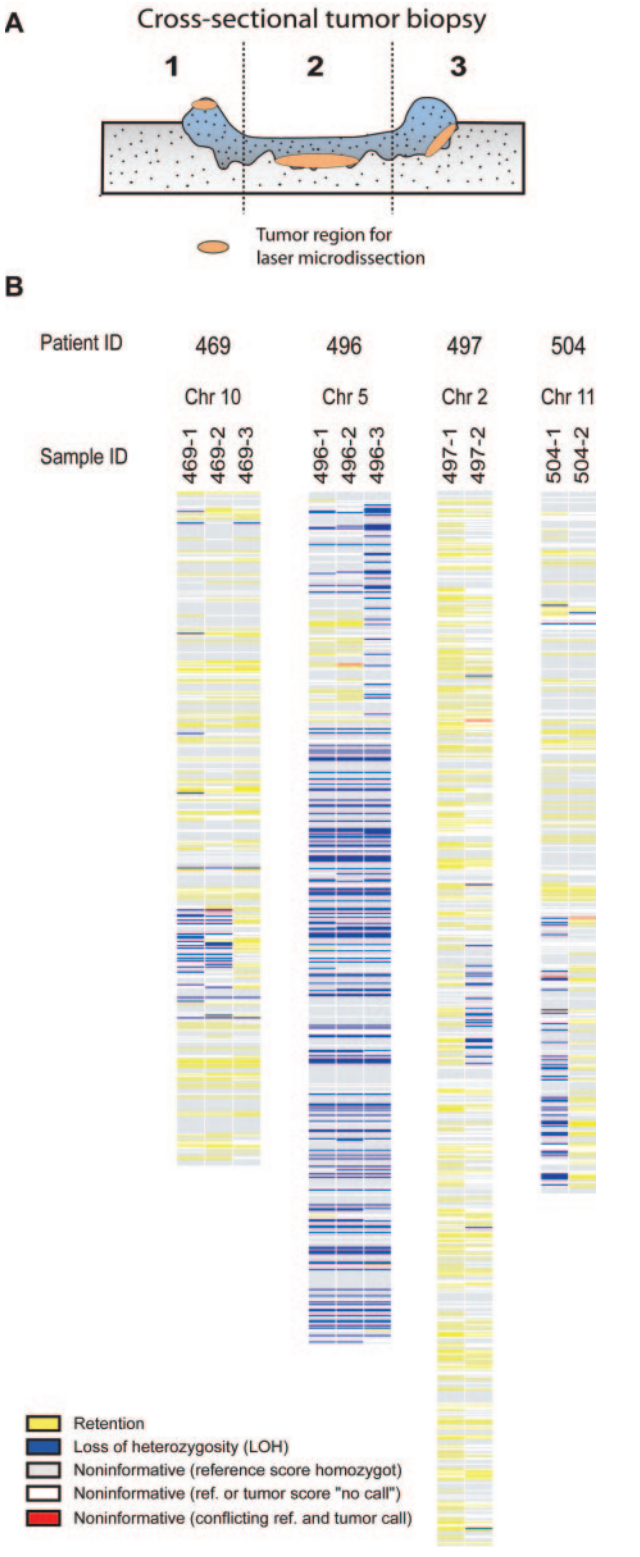

Fig. 2. Intratumor heterogeneity revealed by genome-wide LOH mapping. (**A**) The use of LMD to procure unique cancer cell subpopulations from whole tumor cross-sectional biopsies. At surgery the section was subdivided into three pieces. From each piece a cancer cell subpopulation from either the luminal or invasive front was procured. (**B**) Representative examples of genomic intratumor heterogeneity. Two or three cancer cell subpopulations from each tumor were analyzed. Genomic heterogeneity was identified on different chromosomes in different patients. Observed LOH calls for all SNPs mapping to these chromosomes are shown. The SNPs are ordered according to their genomic position from top (p-terminus) to bottom (q-terminus).

Table V. Correlation of gene expression and genomic alterations

	Genes[a] (*N*)	Observed no. of genes[b]		Expected no. of genes[c]		*P*-value[d]
		Downregulated	Upregulated	Downregulated	Upregulated	
LOH regions in Table II (uniparental disomies)	288	78	111	69	112	0.45
LOH regions in Table II (regular losses)	538	182	130	130	209	<0.000001
Loss regions in Table III	148	63	29	35	58	<0.000001
Gain regions in Table III	164	19	119	39	64	<0.000001

Transcriptional dysregulation is more common than expected in regions with genomic alterations when compared with the overall number of dysregulated genes in neutral regions.
[a]The total number genes observed in the regions listed in Table II.
[b]The total observed number of upregulated and downregulated genes in the regions listed in Table II.
[c]The expected number of upregulated and downregulated genes (based on the distribution of upregulated and downregulated genes in regions of the genome without copy number alterations).
[d]*P*-value for the χ^2-test of the difference between the observed and the expected.

expression of TUSC3 was halved in the adenocarcinomas. Ten of the samples represented paired normal and tumor from five patients. A paired *t*-test on these 10 samples also showed that *TUSC3* is significantly downregulated in adenocarcinomas ($P = 0.017$) (the data are shown in supplementary material). Our real-time RT–PCR data of *SGCZ* indicated that this gene is not expressed by neither normal nor neoplastic epithelia of the colorectum.

Mutational analysis and relation to clinical parameters

We sequenced TP53 and found non-sense or missense mutations in 67% (10/15) of the tumors investigated (Supplementary table 6). No silent mutations were identified. All identified mutations were homozygote and thus in perfect concordance with the observation that mutations were only found in tumors displaying factual losses at the TP53 locus.

We investigated whether clinically important tumor subgroups, defined by TP53 mutational status or lymph node metastasis status, were associated with specific genomic alterations. When based upon copy number data our analysis revealed specific genomic alterations to be significantly associated with both lymph node involvement and TP53 mutational status. Examples of the results are illustrated in Supplementary figure 3. Seventeen regions were specifically associated with TP53 mutational status (including loss of 17p covering the TP53 locus) and three regions with lymph node status (Table VI).

In contrast, when based upon LOH data, the analysis identified only TP53 mutational status to be associated with specific genomic alterations. Three genomic regions were identified with significantly different LOH frequencies in tumors of wild-type and mutated TP53, as expected one of these covered the TP53 locus at 17p (Table VII).

Discussion

The present article is one of the first in which high-density SNP arrays have been used for genome-wide screening for LOH and copy number alterations in CRC specimens. The SNP arrays have two major advantages; first, they enable parallel analysis of thousands of SNP markers yielding high resolution; and, second, they enable concurrent acquisition of genotype and DNA copy number data on a genome-wide scale. Importantly, the tumor analyses of the present study were performed on cell populations procured by LMD, i.e. >95% cancer cells, eliminating the problem of contaminating normal cells often observed in this type studies. The acquired data revealed a set of non-random genomic alterations, some of which were novel with respect to CRC, e.g. the LOH regions on chromosomes 13, 14 and 15, while others where in agreement with data reported by ourselves and other groups, using conventional LOH analysis or CGH (4–7). However, the ~10 000 markers applied in this study provided high resolution and precision in the delineation of the minimally involved regions. Accordingly, we often pinpoint small, previously uncharacterized, regions where other studies typically identify much larger regions or even entire chromosome arms. An example is the short arm of chromosome 8, where still the precise loci that are targeted for loss remains unclear, despite many reports on copy number reductions and LOH (4,6,7,28,29). Our data analysis pinpointed a specific 2.15 Mb region located 13.3 Mb from the p-terminus of chromosome 8 displaying both LOH and copy number loss in 67% of cases. Importantly, the region only contains two known genes, *SGCZ* and *TUSC3*. The relation between these genes and CRC is described poorly. Probe sets for neither *SGCZ* nor *TUSC3* were on the expression array we used for transcriptional profiling, we therefore turned to quantitative real-time RT–PCR analysis to investigate their expression. While we could not find any evidence of *SGCZ* expression in colorectal tissues, we found *TUSC3* to be significantly downregulated in the adenocarcinomas ($P = 0.017$). This finding makes *TUSC3* a good tumor suppressor gene candidate, potentially the one driving the formation of the commonly observed 8p loss. Interestingly, *TUSC3* has previously been suggested to be a tumor suppressor gene with relation to prostate and ovarian cancer (30,31).

Combined analysis of genotype and copy number data revealed the underlying genomic alterations to be much more complex than could have been predicted using either data type alone. Interestingly, the analysis unveiled that 50% of the identified LOH regions showed no evidence of a reduction in DNA copy number, indicating that uniparental disomy plays a significant role in the pathogenesis of CRC. A few regions even displayed uniparental polysomy (note chromosome arms 8q and 20q in Supplementary Figure 2).

Importantly, the distribution of the regions of uniparental disomy was non-random, they were observed primarily on 8q, 13q and 20q. The size of the regions covered everything from whole chromosomes to small interstitial chromosomal regions. Recent studies have shown LOH regions

Table VI. Genomic alterations associated with TP53 mutational status or lymph node metastasis status: copy number alterations

Region ID	Group index[a]	Signal value[b]	Chr. band	Position in HG15[c]		Region size (Mb)	# of SNPs in region	*P*-value[d]	Genes[e]	No. of Dynamic genes[f]	Dysregulated genes[g]		Notes
				Start	End						Down	Up	
1	TP53	<	5q13–q14	67.81	79.99	12.18	31	0.0215	71				
2	TP53	>	8q11–q12	46.97	61.73	14.75	55	0.0199	37				
3	TP53	>	8q13–q21	66.30	76.66	10.35	33	0.0413	42				
4	TP53	>	8q22	93.94	104.39	10.44	28	0.0413	54				
5	TP53	<	14q13–q21	32.16	40.96	8.80	52	0.0300	30				
6	TP53	<	14q21–q23	41.85	59.81	17.96	56	0.0049	78				
7	TP53	<	14q23–q24	60.82	69.12	8.30	36	0.0300	48				
8	TP53	<	14q24–q31	70.16	81.25	11.09	46	0.0170	70				
9	TP53	<	14q31–q32	81.72	93.35	11.64	63	0.0170	58				
10	TP53	<	17p13	0.90	7.53	6.62	15	0.0396	136				Incl. TP53
11	TP53	<	17p13–p12	9.39	15.62	6.23	38	0.0396	29				
12	TP53	<	18p11	1.76	11.52	9.76	52	0.0364	28				
13	TP53	<	18q12–q21	32.45	44.76	12.31	40	0.0154	32				
14	TP53	<	18q21	47.11	55.62	8.51	31	0.0432	29				Incl. DCC
15	TP53	<	18q21–q23	60.81	73.84	13.04	56	0.0088	32				
16	TP53	<	19p13–q12	20.27	33.38	13.11	11	0.0045	13				
17	TP53	<	21q21	20.64	26.77	6.13	31	0.0208	9				
18	Lymphnode	>	7q21–q22	84.62	105.55	20.93	68	0.0134	156	95	8	2	
19	Lymphnode	>	7q22–q31	106.38	121.69	15.31	41	0.0322	43	25	7	3	
20	Lymphnode	>	13q12–q34	28.30	112.85	84.54	371	0.0011	200	109	17	4	Incl. RB1

[a]'TP53' denotes that the groups were constructed according to TP53 mutational status and 'Lymph node' denotes that the groups were constructed according to lymph node metastasis status.
[b]'<' denotes that the average signal value in the tumors with positive group status (e.g. TP53 mutated) is smaller than that of the tumors with negative group status (e.g. TP53 wild-type) and '>' denotes the opposite.
[c]The positions are defined by the SNPs at the boundary of the region, and given as the distance in megabases (Mb) from the p-terminus.
[d]The probability of obtaining by chance a difference in group means that are more extreme than the observed.
[e]The number of genes (Entrez Gene ID's) mapping to the region.
[f]The number of genes for which at least a single probeset on the U133A GeneChip shows a dynamic expression pattern across the 83 expression arrays included in the study.
[g]The number of genes (dynamic) in the given region whose expression was significantly different ($P < 0.05$) in adenocarcinomas ($n = 66$) compared with normal mucosa ($n = 17$).

Table VII. Genomic alterations associated with TP53 mutational status or lymph node metastasis status: LOH regions

Region ID	Group index[a]	LOH most frequent in group	Chr. band	Position in HG15[b]		Region size (Mb)	# of SNPs in region	*P*-value[c]	Genes[d]	Note
				Start	End					
1	TP53	TP53 mutated	17p13–p12	0.90	15.62	14.72	56	0.0366	194	Incl. TP53
2	TP53	TP53 mutated	18p11–q21	0.24	45.88	45.63	163	0.0140	133	Not incl. SMAD4 and DCC
3	TP53	TP53 mutated	20p11	19.37	19.60	0.23	8	0.0326	0	

[a]TP53 denotes that the groups were constructed according to TP53 mutational status, no LOH regions associated with lymph node status.
[b]The positions are defined by the SNPs at the boundary of the region, and given as the distance in megabases (Mb) from the p-terminus.
[c]The probability of obtaining by chance a difference in LOH frequency that are more extreme than the observed.
[d]The number of genes (Entrez Gene ID's) mapping to the region.

representing uniparental disomy also to be common in acute myeloid leukemias, medulloblastomas and basal cell carcinomas (14–16), indicating that uniparental disomy occurs more frequently in human cancers, than has previously been appreciated. Why is there a selection for uniparental disomies in cancer genomes? Recent evidence from studies of myeloid leukemias indicates that uniparental disomy probably represents a mechanism for making an oncogenic event homozygote (activated oncogene or inactivated tumor suppressor gene), without suffering lethal effects from haplo-insufficient genes located within the lost region (32–34).

Genome-wide studies of chromosomal aberrations, based on the detection of copy number differences, have provided important insights into the molecular pathogenesis of CRC. However, our data indicate that such studies underestimate the frequency of genetic losses in this disease, since they do not consider copy number-neutral LOHs. This may therefore represent a significant limitation of CGH-based approaches. Overall, results from our study demonstrate that copy number and LOH analysis detect distinct subsets of genetic abnormalities, which would be overlooked if either of these methods were used alone. Thus, the high-density SNP arrays utilized in the present study, which provide both data types, probably represents at present, the most advantageous methodology for characterization of the genomic alterations involved in CRC development.

We next investigated how genomic alterations influence the expression of genes located in the affected regions. For this purpose we used expression data from an independent set of tumors. The analysis revealed that losses and amplifications imposed characteristic expression patterns on the majority of affected genes, downregulation and upregulation, respectively. This result was in line with previous studies of copy number imbalances and gene expression in breast and prostate cancer (35–37), demonstrating the feasibility of using independent tumor sets for identification of general genomic alterations and general expression changes.

Our analysis further divulged that the expression of genes located in regions of uniparental disomy remained largely unaffected. This finding is interesting in itself, but equally important it confirmed the disomic nature of these regions.

We searched our data for genomic alterations associated with lymph node metastasis and mutational inactivation of TP53, two clinically important transitions during tumor progression. Inactivation of TP53 has been shown to be associated with a poor clinical outcome, genomic instability and a high number of genomic alterations (38–40). Here, we show that TP53 inactivation is not just associated with general genomic instability, but with specific copy number increases and losses. As expected loss of 17p (containing the TP53 locus) was associated with TP53 mutation.

Lymph node metastasis status plays a major role in the clinical management of CRC being the primary determinant for the use of adjuvant chemotherapy. Here, we demonstrate that gain of regions of the long arm of chromosome 7 and most of chromosome 13 is associated with lymph node metastases. The biomarkers of TP53 mutational status and lymph node metastasis status identified in the present article could prove to be valuable for the clinical management of CRC if confirmed in larger studies.

Several reports have demonstrated intratumor genetic heterogeneity in CRC (41–43). However, these reports were all based on investigations of minute parts of the genome represented by either a few specific point mutations, a limited number of microsatellite markers or centromere markers of a few chromosomes. As of yet, the level of intratumor heterogeneity has never been investigated genome-wide. We addressed this and the present study hence represents the first time intratumor heterogeneity has been investigated on a genome-wide scale using more than 10 000 SNP markers. Examination of multiple laser microdissected cancer cell subpopulations per tumor demonstrated intratumor heterogeneity in most of the investigated cancers. In contrast to previous studies we never observed the intratumor heterogeneity at the chromosomal regions frequently showing LOH, including 5q, 8p, 17p, 18p, 18q and 20p. We believe that this is a strong indication that these alterations occur relatively early in CRC development, and are hallmarks if not obligatory for CRC development. The genomic regions we found to display intratumor heterogeneity were only altered in the tumor where they were identified, consistent with them representing the newest abnormalities produced by the ongoing genomic instability, or at least demonstrating that they are not obligatory for the cancer to develop. These regions could be involved in disease progression, invasion and metastasis, but could also be unimportant bystanders produced by the ongoing genomic instability.

In conclusion, our analysis revealed a series of complex genomic alterations, some of which were novel in relation to CRC. We identified a panel of genomic regions showing uniparental disomy. While our combined analysis of copy number and LOH data readily revealed this type of abnormality, regions of uniparental disomy have only very rarely been noticed in previous studies of CRC. The genomic alterations were reflected in the general changes in gene transcript levels, which for uniparental disomy was seen as no general change in transcript level in these regions. Finally, we found that some of the genomic alterations represented intratumor heterogeneity, whereas others were tightly associated with TP53 inactivation or lymph node metastases.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Acknowledgements

The authors thank Karina Dalsgaard Sørensen for designing the primers used for determination of SGCZ genomic copy numbers using quantitative real-time PCR. This study was supported by Grants from The University and County of Aarhus, The Danish Research Council, The Danish Cancer Society and The John and Birthe Meyer Foundation.

Conflict of Interest Statement: None declared.

References

1. Fearon,E.R. and Vogelstein,B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, **61**, 759–767.
2. Vogelstein,B., Fearon,E.R., Hamilton,S.R., Kern,S.E., Preisinger,A.C., Leppert,M., Nakamura,Y., White,R., Smits,A.M. and Bos,J.L. (1988) Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.*, **319**, 525–532.
3. Albertson,D.G., Collins,C., McCormick,F. and Gray,J.W. (2003) Chromosome aberrations in solid tumors. *Nat. Genet.*, **34**, 369–376.
4. Jones,A.M., Douglas,E.J., Halford,S.E., Fiegler,H., Gorman,P.A., Roylance,R.R., Carter,N.P. and Tomlinson,I.P. (2005) Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene*, **24**, 118–129.
5. Laiho,P., Hienonen,T., Karhu,A. *et al.* (2003) Genome-wide allelotyping of 104 Finnish colorectal cancers reveals an excess of allelic imbalance in chromosome 20q in familial cases. *Oncogene*, **22**, 2206–2214.
6. Nakao,K., Mehta,K.R., Fridlyand,J., Moore,D.H., Jain,A.N., Lafuente,A., Wiencke,J.W., Terdiman,J.P. and Waldman,F.M. (2004) High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis*, **25**, 1345–57.
7. Diep,C.B., Kleivi,K., Ribeiro,F.R., Teixeira,M.R., Lindgjaerde,O.C. and Lothe,R.A. (2006) The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes Chromosomes Cancer*, **45**, 31–41.
8. Tsafrir,D., Bacolod,M., Selvanayagam,Z. *et al.* (2006) Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res.*, **66**, 2129–2137.
9. Vogelstein,B. and Kinzler,K.W. (1993) The multistep nature of cancer. *Trends Genet.*, **9**, 138–141.
10. Cavenee,W.K., Dryja,T.P., Phillips,R.A., Benedict,W.F., Godbout,R., Gallie,B.L., Murphree,A.L., Strong,L.C. and White,R.L. (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature*, **305**, 779–784.
11. Haigis,K.M., Caya,J.G., Reichelderfer,M. and Dove,W.F. (2002) Intestinal adenomas can develop with a stable karyotype and stable microsatellites. *Proc. Natl Acad. Sci. USA*, **99**, 8927–8931.
12. Thiagalingam,S., Laken,S., Willson,J.K., Markowitz,S.D., Kinzler,K.W., Vogelstein,B. and Lengauer,C. (2001) Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc. Natl Acad. Sci. USA*, **98**, 2698–2702.
13. Serra,E., Rosenbaum,T., Nadal,M., Winner,U., Ars,E., Estivill,X. and Lazaro,C. (2001) Mitotic recombination effects homozygosity for NF1 germline mutations in neurofibromas. *Nat. Genet.*, **28**, 294–296.

14. Langdon,J.A., Lamont,J.M., Scott,D.K., Dyer,S., Prebble,E., Bown,N., Grundy,R.G., Ellison,D.W. and Clifford,S.C. (2006) Combined genome-wide allelotyping and copy number analysis identify frequent genetic losses without copy number reduction in medulloblastoma. *Genes Chromosomes Cancer*, **45**, 47–60.
15. Raghavan,M., Lillington,D.M., Skoulakis,S., Debernardi,S., Chaplin,T., Foot,N.J., Lister,T.A. and Young,B.D. (2005) Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res.*, **65**, 375–378.
16. Teh,M.T., Blaydon,D., Chaplin,T. *et al.* (2005) Genomewide single nucleotide polymorphism microarray mapping in Basal cell carcinomas unveils uniparental disomy as a key somatic event. *Cancer Res.*, **65**, 8597–8603.
17. Gaasenbeek,M., Howarth,K., Rowan,A.J. *et al.* (2006) Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res.*, **66**, 3471–3479.
18. Bignell,G.R., Huang,J., Greshock,J. *et al.* (2004) High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.*, **14**, 287–295.
19. Kennedy,G.C., Matsuzaki,H., Dong,S. *et al.* (2003) Large-scale genotyping of complex DNA. *Nat. Biotechnol.*, **21**, 1233–1237.
20. Huang,J., Wei,W., Zhang,J. *et al.* (2004) Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics*, **1**, 287–299.
21. Lin,M., Wei,L.J., Sellers,W.R., Lieberfarb,M., Wong,W.H. and Li,C. (2004) dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics*, **20**, 1233–1240.
22. Zhao,X., Li,C., Paez,J.G. *et al.* (2004) An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res.*, **64**, 3060–3071.
23. Kruhoffer,M., Jensen,J.L., Laiho,P. *et al.* (2005) Gene expression signatures for colorectal cancer microsatellite status and HNPCC. *Br. J. Cancer*, **92**, 2240–2248.
24. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
25. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
26. Storey,J.D. and Tibshirani,R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
27. Andersen,C.L., Jensen,J.L. and Orntoft,T.F. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.*, **64**, 5245–5250.
28. Hermsen,M., Postma,C., Baak,J. *et al.* (2002) Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology*, **123**, 1109–1119.
29. Knosel,T., Petersen,S., Schwabe,H., Schluns,K., Stein,U., Schlag,P.M., Dietel,M. and Petersen,I. (2002) Incidence of chromosomal imbalances in advanced colorectal carcinomas and their metastases. *Virchows Arch.*, **440**, 187–194.
30. Bova,G.S., MacGrogan,D., Levy,A., Pin,S.S., Bookstein,R. and Isaacs,W.B. (1996) Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. *Genomics*, **35**, 46–54.
31. Pils,D., Horak,P., Gleiss,A., Sax,C., Fabjani,G., Moebus,V.J., Zielinski,C., Reinthaller,A., Zeillinger,R. and Krainer,M. (2005) Five genes from chromosomal band 8p22 are significantly down-regulated in ovarian carcinoma: N33 and EFA6R have a potential impact on overall survival. *Cancer*, **104**, 2417–2429.
32. Fitzgibbon,J., Smith,L.L., Raghavan,M., Smith,M.L., Debernardi,S., Skoulakis,S., Lillington,D., Lister,T.A. and Young,B.D. (2005) Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res*, **65**, 9152–9154.
33. Griffiths,M., Mason,J., Rindl,M., Akiki,S., McMullan,D., Stinton,V., Powell,H., Curtis,A., Bown,N. and Craddock,C. (2005) Acquired isodisomy for chromosome 13 is common in AML, and associated with FLT3-itd mutations. *Leukemia*, **19**, 2355–2358.
34. Jones,A.V., Kreil,S., Zoi,K. *et al.* (2005) Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood*, **106**, 2162–2168.
35. Hyman,E., Kauraniemi,P., Hautaniemi,S. *et al.* (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.*, **62**, 6240–6245.
36. Pollack,J.R., Sorlie,T., Perou,C.M., Rees,C.A., Jeffrey,S.S., Lonning,P.E., Tibshirani,R., Botstein,D., Borresen-Dale,A.L. and Brown,P.O. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl Acad. Sci. USA*, **99**, 12963–12968.
37. Wolf,M., Mousses,S., Hautaniemi,S. *et al.* (2004) High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression. *Neoplasia*, **6**, 240–247.
38. Bazan,V., Agnese,V., Corsale,S. *et al.* (2005) Specific TP53 and/or Ki-ras mutations as independent predictors of clinical outcome in sporadic colorectal adenocarcinomas: results of a 5-year Gruppo Oncologico dell'Italia Meridionale (GOIM) prospective study. *Ann. Oncol.*, **16** (Suppl. 4), iv50–iv55.
39. Koed,K., Wiuf,C., Christensen,L.L., Wikman,F.P., Zieger,K., Moller,K., von der Maase,H. and Orntoft,T.F. (2005) High-density single nucleotide polymorphism array defines novel stage and location-dependent allelic imbalances in human bladder tumors. *Cancer Res.*, **65**, 34–45.
40. Leslie,A., Pratt,N.R., Gillespie,K., Sales,M., Kernohan,N.M., Smith,G., Wolf,C.R., Carey,F.A. and Steele,R.J. (2003) Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas. *Cancer Res.*, **63**, 4656–4661.
41. Baises,B., Bouzourene,H., Saraga,E.P., Bosman,F.T. and Benhattar,J. (2001) Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int. J. Cancer*, **93**, 346–352.
42. Di Vinci,A., Infusini,E., Peveri,C., Sciutto,A., Orecchia,R., Geido,E., Monaco,R. and Giaretti,W. (1999) Intratumor heterogeneity of chromosome 1, 7, 17, and 18 aneusomies obtained by FISH and association with flow cytometric DNA index in human colorectal adenocarcinomas. *Cytometry*, **35**, 369–375.
43. Losi,L., Baisse,B., Bouzourene,H. and Benhattar,J. (2005) Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. *Carcinogenesis*, **26**, 916–922.

Received January 25, 2006; revised and accepted May 18, 2006