CHAPTER 6

Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression

Gut 2009, 58: 79-89

Beatriz Carvalho
Cindy Postma
Sandra Mongera
Erik Hopmans
Sharon Diskin
Mark A. van de Wiel
Wim van Criekinge
Olivier Thas
Anja Matthäi
Miguel A. Cuesta
Jochim S. Terhaar sive Droste
Mike Craanen
Evelin Schröck
Bauke Ylstra
Gerrit A. Meijer

Abstract

Objective: This study aimed to identify the oncogenes at 20q involved in colorectal adenoma to carcinoma progression by measuring the effect of 20q gain on mRNA expression of genes in this amplicon.

Methods: Segmentation of DNA copy number changes on 20q was performed by array CGH in 34 non-progressed colorectal adenomas, 41 progressed adenomas (i.e. adenomas that present a focus of cancer) and 33 adenocarcinomas. Moreover, a robust analysis of altered expression of genes in these segments was performed by microarray analysis in 37 adenomas and 31 adenocarcinomas. Protein expression was evaluated by immunohistochemistry on tissue microarrays.

Results: The genes *C20orf24, AURKA, RNPC1, TH1L, ADRM1, C20orf20* and *TCFL5,* mapping at 20q were significantly overexpressed in carcinomas compared to adenomas as consequence of copy number gain of 20q.

Conclusion: This approach revealed *C20orf24, AURKA, RNPC1, TH1L, ADRM1, C20orf20* and *TCFL5* genes to be important in chromosomal instability-related adenoma to carcinoma progression. These genes therefore may serve as highly specific biomarkers for colorectal cancer with potential clinical applications.

Introduction

The majority of cancers are epithelial in origin and arise through a stepwise progression from normal cells, through dysplasia, into malignant cells that invade surrounding tissues and have metastatic potential. The colorectal adenoma to carcinoma progression is a classic example of this process [1-2].

Genomic instability is a crucial step in this progression and occurs in two ways in colorectal cancer (CRC) [3]. DNA mismatch repair deficiency leading to microsatellite instability (MIN), explains about 15% of cases [4-6]. In the other 85%, genomic instability occurs at the chromosomal level (CIN) giving rise to aneuploidy. While for a long time chromosomal aberrations have been regarded as random noise, it is now well established that these DNA copy number changes occur in specific patterns and are associated with different clinical behaviour [7-9]. Nevertheless, neither the cause of chromosomal instability in human cancer progression nor its biological consequences have been fully appreciated.

Chromosomal aberrations frequently reported in CRC are 7pq, 8q, 13q, and 20q gains and 4pq, 5q, 8p, 15q, 17p, and 18q losses [10-13]. Of these, especially 8q, 13q and 20q gains and 8p, 15q, 17p and 18q losses are associated with colorectal adenoma to carcinoma progression.

Gain of 20q is observed in more than 65% of CRCs [14]. Gains of 20q are also common in other tumour types and have been associated with poor outcome in gastric and CRC [15-20]. The 20q13 amplicon has been studied in detail in breast and gastric cancers with restricted contig arrayCGH, pinpointing several genes as targets of amplification [21-22]. Analysis of DNA copy number changes at gene level by multiplex ligation-dependent probe amplification (MLPA) showed that in CRC, besides 20q13, also 20q11 is frequently amplified [23].

This study aims to investigate dosage effects of putative 20q oncogenes in colorectal adenoma to carcinoma progression.

Material and Methods

Tumour samples

Forty-one formalin-fixed and paraffin-embedded progressed adenomas (with a focus of carcinoma present, also referred as malignant polyps) collected from the tissue archive of the department of pathology at the VU University medical center (VUmc), Amsterdam, the Netherlands and 73 prospectively collected snap-frozen colorectal tumour samples (37 non-progressed adenomas and

36 carcinomas) were investigated. All samples were used in compliance with the institution's ethical regulations.

The 41 progressed adenomas corresponded to 19 females and 18 males (three patients presented more than one lesion). Mean age was 67 (range 45-86). From these, adenoma and carcinoma components were analysed separately adding to a total of 82 archival samples (41x2).

The 73 frozen specimens corresponded to 31 females and 34 males (six patients had multiple tumours). Mean age was 69 (range 47-89). All histological sections were evaluated by a pathologist (G.A.M.).

Array CGH was performed on both sets of samples while expression microarrays were performed on the frozen samples only.

DNA and RNA isolation

DNA from paraffin was obtained as described previously [24]. RNA and DNA from snap-frozen tissues were isolated using TRIzol (Invitrogen, Breda, The Netherlands) following the supplier's instructions with some modifications, described on http://www.english.vumc.nl/afdelingen/microarrays. Isolated RNA was subjected to purification using RNeasy Mini Kit (Qiagen, Venlo, The Netherlands). RNA and DNA concentrations and purities were measured on a Nanodrop ND-1000 spectrophotometer (Isogen, IJsselstein, The Netherlands) and integrity was evaluated on a 1% agarose ethidium bromide-stained gel.

Array CGH

A BAC/PAC (bacterial artificial chromosome/phage artificial chromosome) array platform was used as described elsewhere [25]. Arrays were scanned (Agilent DNA Microarray scanner G2505B- Agilent technologies, Palo Alto, USA) and Imagene 5.6 software (Biodiscovery, Marina del Rey, California) was used for automatic feature extraction with default settings. Local background was subtracted from the signal median intensities of both test and reference DNA. The median of the triplicate spots was calculated for each BAC clone and log2 ratios (tumour/normal) were normalized by subtraction of the mode value of BAC clones on chromosomes 1-22 (UCSC July2003 freeze of the Human Golden Path – NCBI Build 34). Clones with standard deviation of the intensity of the three spots >0.2 and with >20% missing values were excluded.

Expression microarrays

The Human Release 2.0 oligonucleotide library, containing 60mer oligonucleotides representing 28830 unique genes, designed by Compugen (San Jose, California, USA) was obtained from Sigma-

Genosys (Zwijndrecht, The Netherlands). Printing of slides was done as described elsewhere [26]. Tumour RNA (30 µg) was hybridised against Universal Human reference (Stratagene, Amsterdam, The Netherlands). cDNA labelling and hybridisation procedures are described elsewhere [26]. Scanning of arrays and feature extraction were performed as described above. Overall quality of experiments was judged on microarray plots of intensities of raw data. Normalization was done either with TIGR Midas (http://www.tm4.org/midas.html), using "Lowess" correction [27] or with "Median" normalization and implemented in the maNorm function (Marray R bioconductor package), with identical results. Interarray normalization was also performed. Low intensity values were replaced by the intensity value of 50. Genes with >20% missing values were excluded.
Array CGH and expression microarray data sets are available at Gene Expression Omnibus (GEO) http://www.ncbi.nlm.nih.gov/geo/ [28], accession number GSE8067.

Microarray data analysis

Below, the steps of data analysis are discussed for array CGH data, expression data and integrative analysis. To account for multiple testing, either a false discovery rate (FDR) correction was applied to the P values, or a very stringent P value cut-off was used.

Array CGH data

To segment DNA copy number alterations, a smoothing algorithm "aCGH-Smooth" was applied [29]. Smoothed log2 ratios of -0.15 and 0.15 were used as thresholds to define gains and losses (99% CIs) obtained for 15 normal-to-normal hybridisations. Only gains and losses covering at least three consecutive BAC clones were included. Amplifications were called when log2 ratios exceeded 1.0. DNA copy number data were stored in ArrayCGHbase [30] (http://arraydb.vumc.nl/arrayCGHbase). Median absolute deviation (MAD) was determined for each case as a quality control. Cases with MAD ≥0.2 were excluded. Array CGH profiles were visualized in ArrayCGHbase. Supervised analysis, comparing two groups, was done using CGHMultiArray [31]. For analysis of paired samples (adenoma and carcinoma components within progressed adenomas) an adapted version of CGHMultiArray was used, based on the Wilcoxon sign-rank test corrected for ties. Reported p values are adjusted for multiple testing (FDR), unless stated otherwise.
For defining the most frequent smallest regions of overlap (SRO) for gains on 20q, throughout all cases, STAC (Significance Testing for Aberrant Copy-number) was used [32].

Microarray expression data

As all hybridisations were performed against a common reference, all comparisons were relative between colorectal adenomas and carcinomas.

Supervised analysis for comparing carcinomas and adenomas was done using the Wilcoxon signed rank test, and a modified version of this test- total Thas score (http://www.cvstat.ugent.be/index.php?page=techrep/techrep.htm) that is powerful when the distributions of the expression levels of both groups do not differ over the whole range of expression levels. This occurs when not all cases in the carcinomas and adenomas groups have differentially expressed genes, but differences rather appear in subpopulations. Genes were considered as differentially expressed when a Wilcoxon test p value <1e-5 and a Thas P value <0.05, corresponding to a FDR <0.05.
To disclose genes which expression is influenced by 20q gain, tumours with and without 20q gain were compared. Gene expression was regressed on copy number count using a linear model.
To evaluate the discriminatory power of candidate genes for classifying adenomas vs carcinomas, a stepwise linear discriminant analysis with leave one out cross-validation was performed on mRNA expression data (SPSS 15.0 for Windows, SPSS, Chicago, Illinois, USA).

Integration of copy number and expression data

ACE-it (Array CGH Expression integration tool) was applied to test whether gene dosage affects RNA expression [33]. Only genes on chromosome 20 are presented. We used a cut-off value of 0.15 for gains and losses, a default group value of 9 and a FDR ≤0.10.

Quantitative reverse transcription-PCR (qRT-PCR)

RNA (1 μg) was treated with DNase I and reverse transcribed to cDNA using oligo(dT)20 Primer with Superscript II reverse transcriptase (Invitrogen).
qRT-PCR was performed in duplicate on 15 adenomas and 15 carcinomas for six candidate genes. A master mix was prepared with 12.5 μl of SYBR Green PCR master mix (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands), 0.5 μM of each primer in 22.5 μl. cDNA (25 ng in 2.5 μl) was added to the mix. Reactions were performed in a 7300 Real-time PCR System (Applied Biosystems). Amplification conditions comprised a denaturation step at 95°C for 10 min and 50 cycles at 95°C for 15 sec and annealing temperature for 1 min (Supplementary table 1). Relative expression levels were determined following the 2ΔΔCt method [34], using β-2-microglobulin gene as a reference. This gene was previously demonstrated not to differ in expression between adenomas and carcinomas [35].

Immunohistochemistry on tissue microarrays (TMAs)

A TMA was constructed with 57 tumours (32 adenomas and 25 carcinomas) of which array CGH and/or expression microarray data were available. Of each tumour three cores from different

locations within the tumour were included in the array. A 4 µm section of the array was used for immunohistochemistry. After deparaffination in xylene, and rehydration through graded alcohol to water, endogenous peroxidase was blocked with hydrogen peroxide (0.3% H2O2/methanol) for 25 min. Antigen retrieval was done by autoclaving in citrate buffer (10 mM; pH 6.0). Primary Aurora A monoclonal antibody NCL-L-AK2 from Novocastra Laboratories (Newcastle, UK) was incubated overnight at 4°C in a dilution of 1:50. The secondary antibody, K4006, mouse, from Envision kit (DAKO, Heverlee, Belgium) was incubated for 30 min at room temperature. Counterstaining was done with Mayer's hematoxylin. Incubation without primary antibody was used as negative control. The colorectal cancer cell line Caco-2, which has a 20q gain and is known to express Aurora A, was used as positive control. Caco-2 cells were fixed and paraffin embedded, sections of which were taken along in the same run of immunohistochemistry as the TMA was processed. Caco-2 produced strong nuclear, mostly along with cytoplasmic, staining in >75% of tumour cells and this pattern was taken as reference for intense staining. Next, the spectrum of staining in the respective cores on the TMA was surveyed. in terms of intensity and positive nuclei. Only staining in tumour cells (i.e. either adenoma or carcinoma cells) was considered. Cores of the TMA typically contained 4-17 crypts with in every crypt >100 cells which all were evaluated. Basically, three staining patterns were seen; no staining at all, strong staining comparable to that observed in Caco-2 cells, and an intermediate pattern that showed positive staining, but clearly less intense than in Caco-2 cells. The intensity of staining was taken as most important parameter. In pattern 2, typically 50% to 75% of nuclei showed intense staining, while in pattern 1 typically 25% to 75% of nuclei showed weak staining. For score 0, no more than a scattered weakly positive cell was tolerated. Based on evaluation of up to three cores by two independent observers, a score ranging from 0 to 2 was assigned per tumour, with score 0 corresponding to no signal, score 2 corresponding to the strong signal that was observed in the positive control Caco-2 and score 1 for an intermediate intensity staining. In case of disagreement between observers, a third observer was consulted and the majority score was noted.

Cochran-Armitage test analysis was performed to compare protein expression with lesion type (adenoma, carcinoma). Jonckheere-Terpstra test was performed to compare protein expression with log2 ratios (microarray expression data). Both tests make explicit use of the ordinality of the protein levels of expression. Differences were considered significant when $P<0.05$.

Results

Delimiting gained regions on 20q

We analysed by array CGH 41 progressed adenomas, which were previously studied by classical CGH. We analysed the adenoma and carcinoma components of these samples separately. Gain of 20q occurred in >60% of the cases (Figure 1A, 1B). The pattern of copy number changes did not differ between adenoma and carcinoma components (as determined by CGHMultiArray), although sometimes showed lower amplitudes in the adenoma component (Figure 1A, 1B).

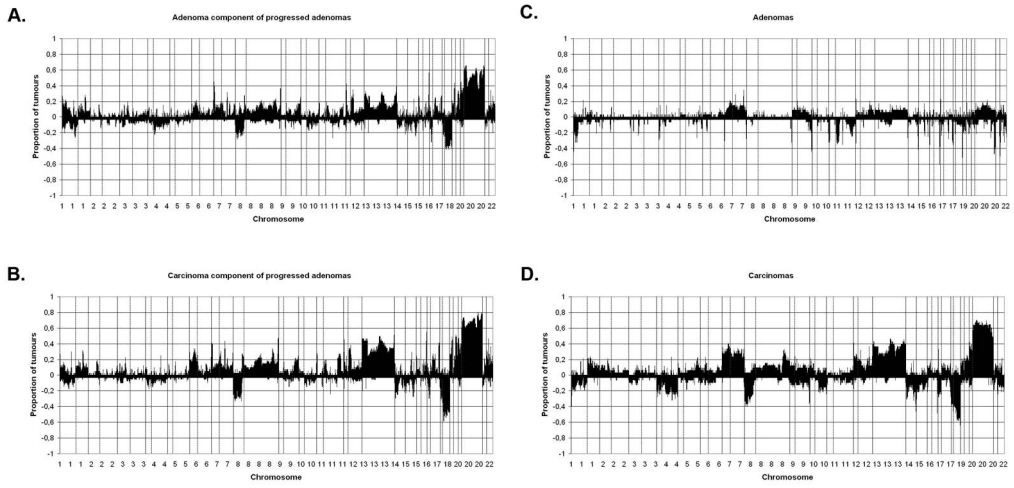


Figure 1. Frequency plot of DNA copy number gains and losses determined by bacterial artificial chromosome (BAC) array comparative genomic hybridisation in **A**) adenoma components of 41 progressed colorectal adenomas, **B**) adenocarcinoma components of 41 progressed colorectal adenomas, **C**) 34 non-progressed colorectal adenomas and **D**) 33 adenocarcinomas. Y axis displays the fraction of tumours with either a gain (positive sign) or loss (negative sign) for all clones that are sorted by chromosome and base pair position.

Next, we analysed DNA copy number status of 37 non-progressed adenomas and 36 carcinomas. From these 73 tumours, 67 (34 adenomas and 33 carcinomas) showed high quality genomic profiles with MAD values <0.2, giving an 8% drop-out. In these 67 tumours, chromosome 20 gain occurred in <15% of the adenomas but in >60% of the carcinomas (P<0.00001, as determined by CGHMultiArray), mostly affecting either all of chromosome 20 or the q-arm only, similar to the progressed adenomas (Figure 1C, 1D).

To determine the most relevant regions within 20q harbouring putative oncogenes with a role in colorectal adenoma to carcinoma progression, STAC [32] was applied to the combined set of

paraffin-embedded malignant polyps (n=41) and frozen carcinomas (n=33). This revealed 3 relevant regions of aberrant copy gains on 20q, one spanning 4 Mb (32-36 Mb), one spanning 3 Mb (56-59 Mb) and the third one spanning 2Mb (61-64 Mb) (Figure 2). These three regions (SROs) contained 80, 35 and 94 known genes, respectively.

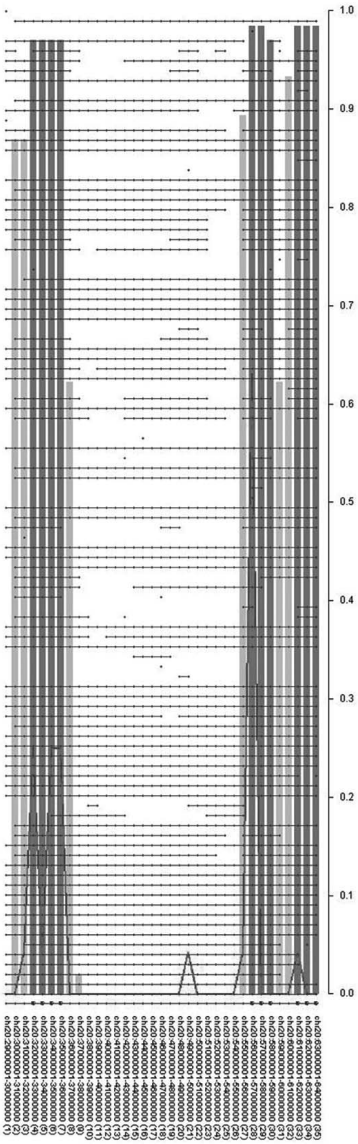


Figure 2. Delimitation of the smallest regions of overlap (SRO's) by STAC analysis for 115 samples (41 non-progressed adenomas, 41 adenocarcinoma components of progressed adenomas, and 33 adenocarcinomas). Results for the long arm of chromosome 20 are displayed. Rows represent samples, and columns represent chromosomal locations. A black dot indicates a gain called in a sample at a location. Consecutive black dots are connected by a line to represent an interval of aberration. Grey bars track the maximum STAC confidence (1 – P value), darker bars are those with confidence >0.95. The line graph indicates the actual frequencies in the sample set.

Identification of differentially expressed genes

We performed microarray expression analysis on the 37 non-progressed adenomas and 36 carcinomas of which snap-frozen material was available. High quality expression array data were obtained from 68 cases (37 adenomas and 31 carcinomas, 7% drop-out).

Supervised data analysis for identifying putative oncogenes on 20q, was done in two different ways; we compared carcinomas to adenomas, and we compared tumours with 20q gain to tumours without 20q gain. The first approach revealed genome-wide 122 up-regulated genes and 219 downregulated genes (a total of 341 differentially expressed genes), in carcinomas when compared to adenomas (Wilcoxon test p value <1e-5 (FDR <0.05) and Thas P-value <0.05). Of these 122 upregulated genes, 14 map at chromosome 20q (Table 1). For the second approach, only tumours (adenomas and carcinomas) that had both array CGH data and expression data available (n=64) were included. As a preselection, we used genes differentially expressed (both up and down) between carcinomas and adenomas, as we focus on genes at 20q that are involved in progression, using a less stringent cut-off (Thas P value <0.05). Thereby, we identified 127 genes genome-wide out of 931 differentially expressed genes (regression analysis; FDR ≤0.1), whose expression levels are influenced by the occurrence of 20q gain. Of these 127 genes, 21 are mapped at 20q (Table 2). Nine genes common to these two approaches emerged, namely *TPX2, C20orf24, AURKA, RNPC1, TH1L, ADRM1, C20orf20, TCFL5* and *C20orf11*.

Integration of array CGH and expression data

BAC array CGH data were related to oligonucleotide expression array data, independently of adenoma or carcinoma status, using a dedicated integration tool called ACEit [33]. We obtained a list of 151 genes located at chromosome 20 for which gene dosage affected expression levels (FDR ≤0.1), 120 of which are on the q-arm (Supplementary table 2). Combining this information with the results of the two supervised approaches for expression data analysis (carcinoma vs adenoma and 20q gain vs no 20q gain), seven genes were shared (Figure 3). For these genes, *C20orf24, AURKA, RNPC1, TH1L, ADMR1, C20orf20* and *TCFL5*, combined box plots with dot plots of mRNA expression in adenomas vs carcinomas (Figure 4A) and scatter plots of mRNA expression vs DNA copy number ratio (Figure 4B) are shown.

Table 1. Genes significantly upregulated in carcinomas, when compared with adenomas, mapping at 20q (Wilcoxon ranking p value <1e-5 (FDR <0.05) and Thas P value < 0.05), ordered by chromosomal position (Location in base pairs (bp) according to Freeze July 2003; NCBI Build 34) with HUGO gene symbols and GenBank accession ID.

Gene symbol	GenBank accession no.	Location (bp position)	Wilcoxon P-value	Thas P-value
C20orf1(TPX2)	NM_012112	31103374	2E-06	8E-05
MYRL2	NM_006097	35859501	5E-06	4E-05
C20orf24 (RIP5)	NM_018840	35923014	2E-07	2E-05
TOMM34	NM_006809	44265329	8E-08	0
RBPSUHL	NM_014276	44626010	2E-07	6E-06
BCAS4	NM_017843	50138063	2E-06	6E-05
AURKA (STK6)	NM_003600	55641283	4E-10	0
FLJ37465 (BMP7)	AK094784	56477906	1E-09	0
RNPC1	NM_017495	56660843	8E-07	7E-05
TH1L	NM_016397	58253070	1E-06	1E-05
ADRM1	NM_007002	61566389	9E-07	8E-05
C20orf20	NM_018270	62156238	9E-09	0
TCFL5	NM_006602	62211152	2E-09	0
C20orf11	NM_017896	62299593	4E-07	0

FDR, false discovery rate

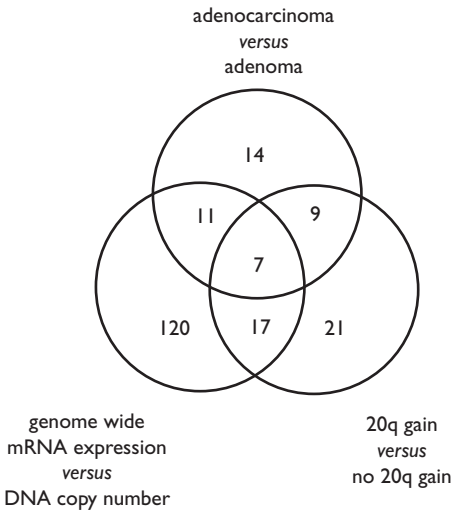


Figure 3. Venn diagram integrating results of three different data analysis approaches (comparing colorectal adenocarcinomas vs adenomas; colorectal tumours with 20q gain vs tumours without 20q gain; and genome wide integration of mRNA expression data with DNA copy number data). Seven genes (*C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*) emerge with all three approaches.

Table 2. Genes whose expression is related to the 20q gain (FDR ≤0.10), ordered by chromosomal position (location in base pairs (bp) according to freeze July 2003; NCBI Build 34) with HUGO gene symbols and GenBank accession ID.

Gene Symbol	GenBank accession no.	Location (bp position)	FDR
HM13	NM_030789	30874805	0.03
C20orf1 (TPX2)	NM_012112	31103374	0.03
CDC91L1	NM_080476	33922394	0.02
C20orf44	NM_018244	34608051	0.07
DLGAP4	NM_014902	35761669	0.05
TGIF2	NM_021809	35897616	0.003
C20orf24 (RIP5)	NM_018840	35923014	0.0006
YWHAB	NM_014052	44210177	0.0002
UBE2C	NM_007019	45128792	0.01
DPM1	NM_003859	50248672	0.000001
NFATC2	AK025758	50769018	0.003
AURKA (STK6)	NM_003600	55641283	0.02
RNPC1	NM_017495	56660843	0.04
TH1L	NM_016397	58253070	0.007
ADRM1	NM_007002	61566389	0.05
SLCO4A1	NM_016354	62015102	0.08
C20orf20	NM_018270	62156238	0.04
TCFL5	NM_006602	62211152	0.03
C20orf11	NM_017896	62299593	0.0009
C20orf59	NM_022082	62323360	0.007
PRPF6	NM_012469	63364789	0.03

FDR, false discovery rate

Of these seven candidate genes, 6 map within the SROs determined by STAC analysis. The seventh gene (*AURKA*) maps approximately 400 kb proximal to SRO2 at 55.6 Mb (20q13.31). *C20orf24* maps within SRO1 at 35.9 Mb (20q11.23), *RNPC1* and *TH1L* map within SRO2 at position 56.7 and 58.3 Mb, respectively (20q13.32), and genes *ADMR1, C20orf20* and *TCFL5* map within SRO3, the first at 61.6 and the other two at 62.2 Mb (20q13.33).

Stepwise linear discriminant analysis with leave one out cross validation showed that mRNA expression levels of two out of the seven candidate genes – that is, *RNPC1* and *TCFL5*- allowed to correctly classify 88.2% of the cases (60/68) as adenomas or carcinomas (Figure 5 and table 3).

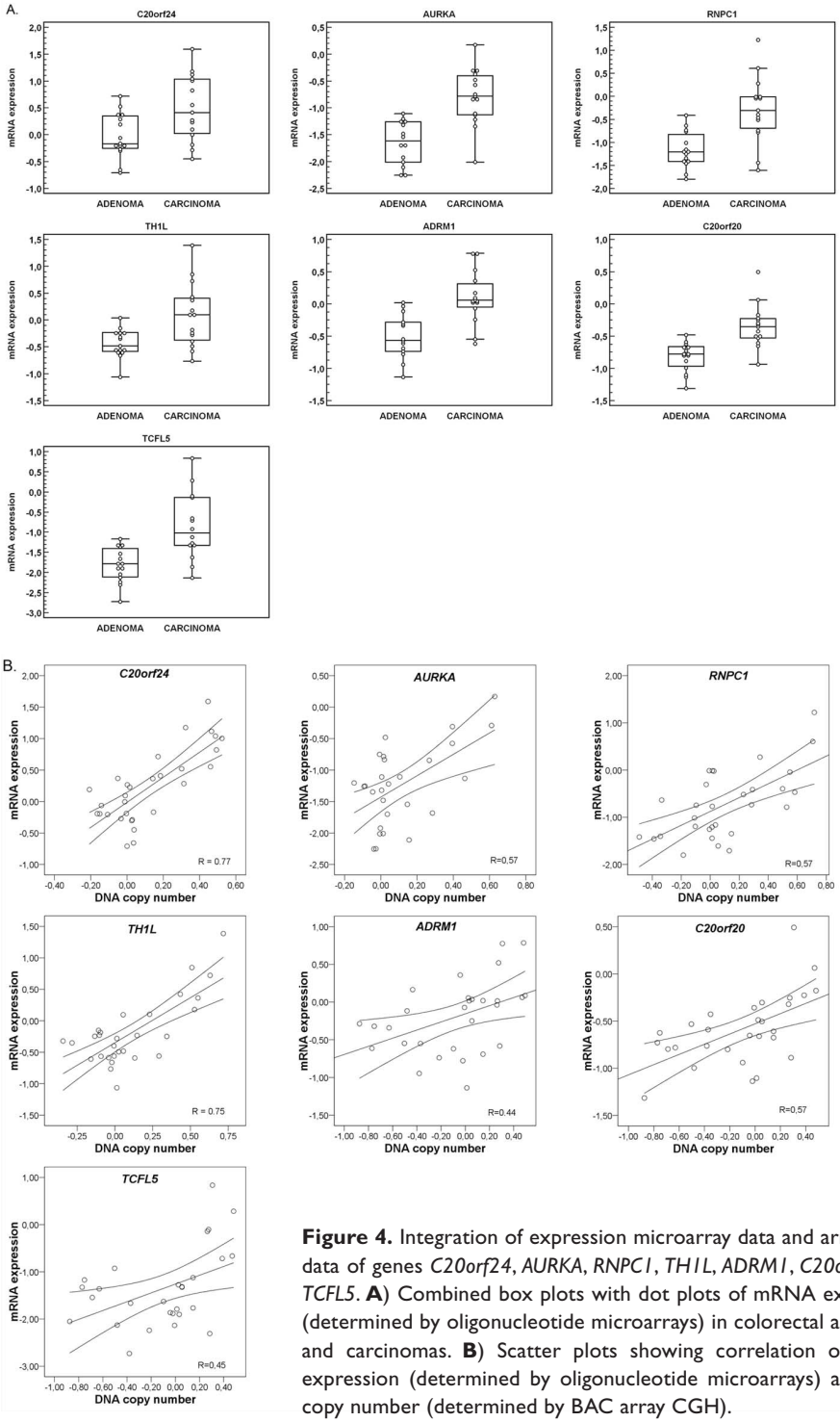


Figure 4. Integration of expression microarray data and array CGH data of genes *C20orf24*, *AURKA*, *RNPC1*, *TH1L*, *ADRM1*, *C20orf20* and *TCFL5*. **A**) Combined box plots with dot plots of mRNA expression (determined by oligonucleotide microarrays) in colorectal adenomas and carcinomas. **B**) Scatter plots showing correlation of mRNA expression (determined by oligonucleotide microarrays) and DNA copy number (determined by BAC array CGH).

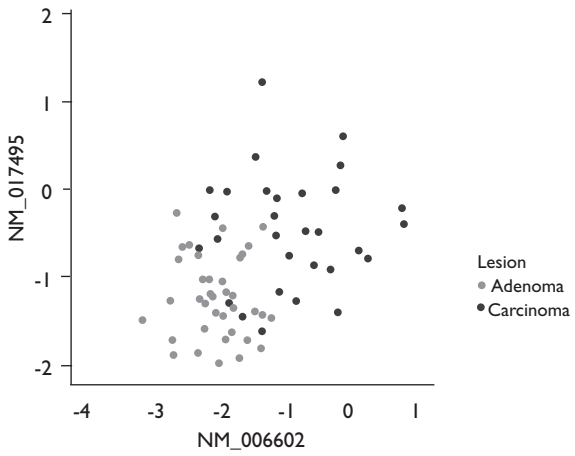


Figure 5. Scatter plot of mRNA expression levels of *RNPC1* (NM_017495) and *TCFL5* (NM_006602), by lesion (grey circles: adenomas; black circles: carcinomas) showing a good separation of colorectal adenomas vs adenocaricnomas.

Table 3. Results of linear stepwise discriminant analysis with leave one out cross-validation of the seven candidate genes.

			Predicted Group Membership		
		Lesion	**Adenoma**	**Carcinoma**	**Total**
Original	Count	Adenoma	35	2	37
		Carcinoma	6	25	31
	%	Adenoma	94.6	5.4	100.0
		Carcinoma	19.4	80.6	100.0

From 68 tumours in total, 60 were correctly classified (88.2%), using expression levels of *RNPC1* and *TCLF5* only.

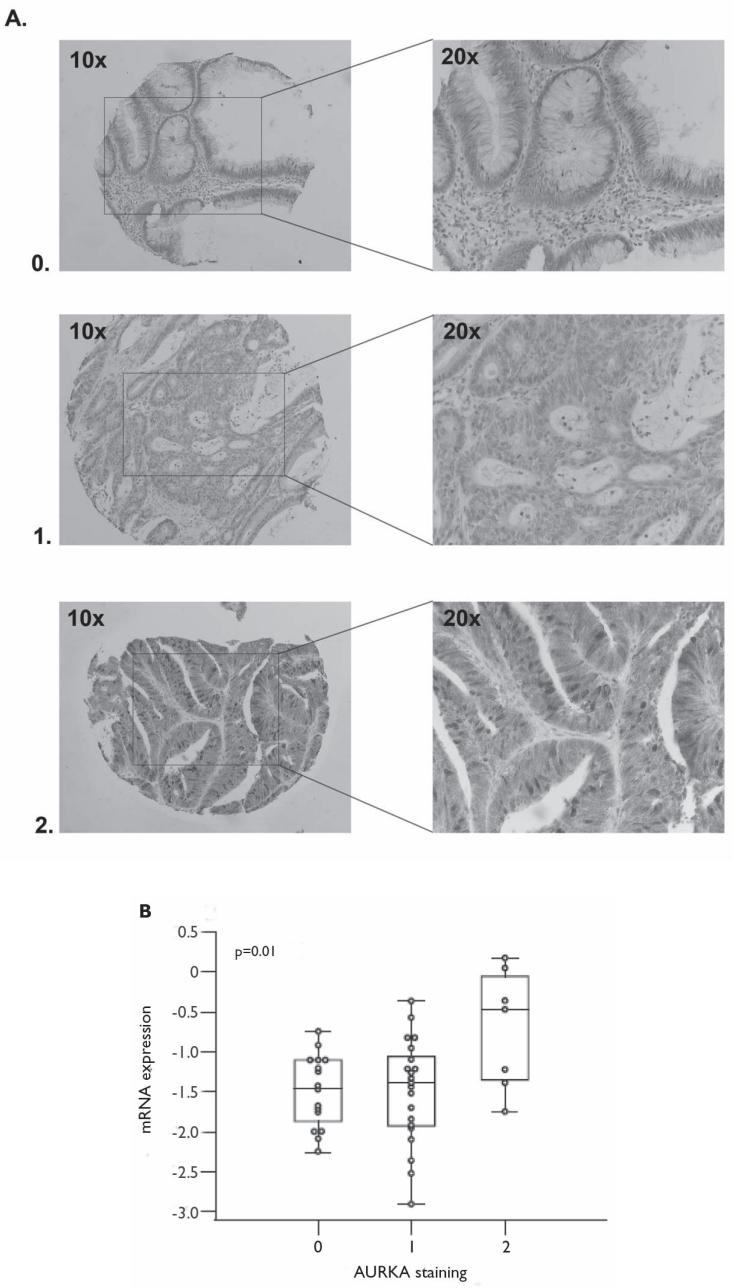


Figure 6. A) Examples of AURKA protein expression in TMA cores of an adenoma showing no expression (0), an adenocarcinoma showing weak expression (1), and an adenocarcinoma showing strong expression (2). **B)** Combined box plot with dot plot of mRNA expression, determined by oligonucleotide microarrays (Y-axis), of colorectal adenomas and carcinomas with a negative (0), weak (1) or strong (2) protein expression of AURKA on immunohistochemistry (X-axis).

Table 4. Expression fold changes and range determined by expression microarray and by qRT-PCR, comparing either carcinomas vs adenomas (Ca/Ad) or tumors with 20q gain vs tumors without 20q gain (20q gain/non-20q gain)

Gene	Comparison	Microarray fold change	qRT-PCR fold change	Microarray expression range*	qRT-PCR expression range*
C20orf24	Ca/Ad	1.54	1.78	(-0.45 to 1.60)/(-0.71 to 0.71)	(1.84 to 6.08)/(-0.26 to 4.81)
	20q gain/non-20q gain	1.68	3.99	(-0.17 to 1.60)/(-0.71 to 0.37)	(-0.26 to 6.08)/(1.85 to 4.95)
AURKA	Ca/Ad	1.91	3.39	(-2.01 to 0.17)/(-2.26 to -1.11)	(-1.78 to 6.06)/(-0.64 to 3.72)
	20q gain/non-20q gain	1.55	4.53	(-2.11 to 0.17)/(-2.26 to -0.48)	(1.03 to 6.06)/(-1.78 to 3.99)
RNPC1	Ca/Ad	1.74	nd	(-1.61 to 1.22)/(-1.80 to -0.41)	nd
	20q gain/non-20q gain	1.58	nd	(-1.71 to –1.22/-1.80 to -0.01)	nd
TH1L	Ca/Ad	1.52	4.98	(-0.77 to 1.39)/(-1.06 to -0.15)	(-1.97 to 6.27)/(-3.57 to 3.72)
	20q gain/non-20q gain	1.59	6.4	(-0.59 to 1.39)/(-1. 06 to 0.10)	(-3.57 to 6.27)/(-3.57 to 3.72)
ADRM1	Ca/Ad	1.45	1.46	(-0.62 to 0.79)/(-1.14 to 0.02)	(-0.30 to 5.58)/(-1.29 to 5.34)
	20q gain/non-20q gain	1.38	2.58	(-0.69 to 0.78)/(-1.14 to 0.36)	(-1.29 to 5.58)/(-0.30 to 5.34)
C20orf20	Ca/Ad	1.36	3.08	(-0.94 to 0.49)/(-1.31 to -0.59)	(-1.32 to 2.07)/(-2.79 to 0.14)
	20q gain/non-20q gain	1.34	3.57	(-0.89 to 0.49)/(-1.31 to -0.36)	(-1.16 to 2.06)/(-2.79 to 0.35)
TCFL5	Ca/Ad	2.2	3.54	(-2.14 to 0.83)/(-2.73 to -1.17)	(2.07 to 6.94)/(-1.28 to 4.21)
	20q gain/non-20q gain	2.02	3.54	(-2.31 to 0.83)/(-2.73 to -0.93)	(-1.28 to 6.94)/(1.99 to 4.41)

not determined; qRT-PCR, quantitative reverse transcription-PCR; * Log2 ratio

Confirmation of differential expression by qRT-PCR

qRT-PCR was performed on a sub-sample (n=30) of frozen tumours (15 adenomas and 15 carcinomas) to confirm the expression levels of six of the seven genes identified.

Carcinomas showed higher expression of all 6 genes compared to adenomas and tumours with 20q gain (4 adenomas and 8 carcinomas) showed higher expression compared to tumours without 20q gain (11 adenomas and 7 carcinomas). Table 4 shows the fold changes observed between either, carcinomas and adenomas or, tumours with 20q gain vs tumours without 20q gain, by microarrays and by qRT-PCR.

In situ confirmation of AURKA expression by immunohistochemistry on TMAs yielded higher expression of AURKA in carcinomas compared to adenomas (P=0.01) (Table 5) as well as a significant positive correlation with the mRNA expression levels (P=0.01) (Figure 6). Validation of other genes was hampered by the absence of adequate antibodies.

Table 5. AURKA protein expression in colorectal adenomas vs carcinomas by immunohistochemistry on tissue array

		AURKA staining				
		Negative	**Weak**	**Strong**	**Total**	**p value***
Lesion	Adenoma	12	12	1	25	
	Carcinoma	4	9	6	19	
Total		16	21	7	44	0.01

* Cochran-Armitage test

Discussion

One of the most frequent chromosomal aberrations observed in CRC is gain of the long arm of chromosome 20. Nonetheless, which of the many genes mapping at 20q show altered expression due to DNA copy number alterations and play a role in the progression of colorectal adenoma to carcinoma, is not yet fully understood. In order to try to identify these putative oncogenes, we analysed a series of colorectal tumours, both adenomas and carcinomas at the DNA and RNA level.

In this study we confirmed that chromosome 20 was the most frequently altered in the progressed adenomas and carcinomas (in >60% of cases). In non-progressed adenomas, gains of 20q were detected in <20%, supporting a role of 20q gain in colorectal adenoma to carcinoma progression consistent with earlier observations [7]. Narrowing down the gained region by array CGH across

all tumours analysed yielded three smallest regions of overlap: SRO1 at 20q11.22-q11.23 (32-36 Mb), SRO2 at 20q13.32-q13.33 (56-59 Mb) and SRO3 at 20q13.33 (61-64 Mb). Douglas and collaborators [12], in a series of 48 cell lines and 37 primary CRCs, defined 20q13.3 as the most common region of gain by array CGH (at 62.3 Mb), which is included in the SRO3. Moreover, Nakao and collaborators [13], also by array CGH in a panel of 125 primary CRCs, observed like us more than one prominent region of gain on 20q, centring at 32.3, 37.8, 45.4, 54.7, 59.4 and 65 Mb (according to the August 2001 freeze of the human genome). Based on these array CGH data the previous studies listed several candidate oncogenes located at these genomic regions, such as *AIB4* (35 Mb), *ZNF217* (53.9 Mb), *CYP24* (54.5Mb) and *AURKA* (56.6 Mb), and at 62.3 Mb the genes *LIVIN, PTK6, HD54* and *EEF1A2*. In the present study, the defined regions SRO1, 2 and 3 harbour 80, 35 and 94 genes, respectively, including *AURKA, PTK6* and *EEF1A2*. Altogether, based on copy number alterations alone the number of genes is still too high (209 genes, in total) to really pinpoint the genes that play a role in the progression of colorectal adenoma to carcinoma.

Comparison of the expression of carcinomas relatively to adenomas using strict statistical parameters showed that 14 genes mapping at 20q were overexpressed in carcinomas as compared to adenomas. Most of these genes were not described in previous expression microarray studies comparing adenomas to carcinomas [36-37]. These studies used either Affymetrix arrays representing 6600 genes or cDNA microarrays representing 23000 genes as oppose to the 28830 represented on our oligonucleotides arrays. In addition, the number of samples analysed in the present study is considerably higher, with 4 to 10 times more adenomas and 2 to 3 times more carcinomas. In combination with our multi-angle, integrative data analysis the power of the present study thus is higher. Two genes that have been described as being over-expressed in colon tumours, compared to normal mucosa, are *ADRM1* and *TOMM34* [38]. Moreover, cDNA expression analysis in a series of clinical samples of CRC patients showed frequent up-regulation of *TOMM34* in carcinomas and inhibition of this gene by siRNA in HCT116 colon cancer cell line drastically reduced the cell growth [39]. Together, these findings corroborate our results since *ADRM1* and *TOMM34* are on the list of genes that showed to be up-regulated in carcinomas (compared to adenomas) (Table 1). This lends support to the approach of the present study, and suggests that also the other candidates identified might be relevant.

Looking at the same expression data from a different angle - that is, comparing the expression of tumours with and without 20q gain- we aimed to find genes with a dosage effect on expression. Genome-wide, expression of 127 out of 931 genes was related to 20q gain, 21 of which are located at chromosome 20q itself. Although chromosome 20 has a high gene density, and copy number gains of the long arm are very frequent, certainly not all genes mapping at the gained regions are

recurrently over-expressed. Two hundred and nine genes are mapped to the SRO's defined here, but only 21 genes are recurrently up-regulated in association with 20q gain, consistent with earlier observations [40-41].

Nine genes overlapped between the 14 adenoma vs carcinoma genes and the 21 genes associated with either or not 20q gain, namely *TPX2, C20orf24, AURKA, RNPC1, TH1L, ADRM1, C20orf20, TCFL5* and *C20orf11*.

TPX2 and *AURKA*, differentially expressed between carcinomas and adenomas and associated with 20q gain, are known to be related to aneuploidy and interact with each other [42]. Moreover, in a microarray based chromosomal instability (CIN) gene signature, *TPX2* ranked first [43]. This finding supports the hypothesis that chromosome 20q gain has an important role in colorectal adenoma to carcinoma progression cancer, and that the onset of manifest CIN is instrumental.

Our third approach, integration of DNA copy number changes and gene expression demonstrated that throughout the genome 507 genes showed a statistically significant association between DNA copy number and mRNA expression status, both for amplified/up-regulated and deleted/down-regulated genes, 120 of these being located on chromosome 20q. Some well known genes like, *ZNF217* and *CSE1L*, previously reported to be amplified in colon cancer [44-46], showed relative higher expression when there was allelic gain. From these 120 genes, 17 overlapped with the 20q gain associated list, and 11 overlapped with the adenoma and carcinoma vs carcinoma list. Overlapping these three approaches (expression in adenomas vs carcinomas, expression vs 20q gain, and genome wide expression vs whole genome copy-number changes) showed that seven genes are consistently significant (Figure 4), namely *C20orf24, AURKA, RNPC1, TH1L, ADRM1, C20orf20* and *TCFL5*. In addition to the already stringent data analysis, a permutation analysis was performed, comparing the differential expression of the seven 20q genes with the expression of >50.000 random subsets out of genes 7946 in silent DNA regions (2q, 3, 5, 10p, 11, 16, 21, 22). For each random subset, the Wilcoxon scores of the seven most differentially expressed (adenoma vs carcinoma) genes were selected. The seven genes on 20q showed a significantly higher expression in adenocarcinomas vs adenomas compared to the best performing combination from the permutation test (P=0.001), underlining that the copy number based discovery of putative oncogenes did not yield random differentially expressed genes. The fact these overexpressed putative oncogenes on 20q actually resulted in biologically active components (i.e, proteins) in the tumour cells was demonstrated by immunohistochemistry on TMA for AURKA. For the other candidates, antibodies did not perform adequately in the tissue samples or were not available at all. Little is known about the function of most of these genes. Some are transcription factors, like *TCFL5* [47], or otherwise involved in transcriptional regulation, like *C20orf20* [48]. *TH1L* product is

involved in regulation of A-Raf kinase [49]. *ADRM1* encodes for a putative cell adhesion molecule that recently was shown to be component of the 26S proteosome [50]. *RNPC1* product is predicted to bind to RNA, based on sequence motifs and C20orf24 interacts with Rab-5, although its precise function is still unknown. *AURKA* has been well characterized and is involved in cell cycle regulation. It has been shown to be amplified in CRC [51] and its overexpression induces centrosome amplification, aneuploidy and transformation *in vitro* [52]. Moreover, inhibiting *AURKA* by RNA interference lead to growth suppression of human pancreatic cancer cells [53]. Knocking down *TCFL5* resulted in suppression of the number of multicellular HT29 tumour spheroids, supporting its role in cancer development [54].

In summary, we demonstrated the involvement of three SROs in the 20q amplicon in CRC and showed strong DNA copy number-mRNA expression associations for seven genes in these areas. In addition we demonstrated significant differences between colorectal adenomas and carcinomas at the DNA, mRNA and, for a one of the genes, at the protein level, supporting an important role as oncogenes in colorectal adenoma to carcinoma progression. Furthermore we showed that expression levels of two out of the seven genes allowed to discriminate adenomas from carcinomas with high accuracy. These genes therefore may both serve as highly specific biomarkers for CRC and as possible targets for pharmaceutical intervention in the development of CRC.

Acknowledgements

We thank the mapping core and map finishing groups of the Wellcome Trust Sanger Institute for initial BAC clone supply and verification. We would like to thank Anders Svensson for handling the array CGH database, Wessel van Wieringen for supervising the use of the software tool ACEit, Marianne Tijssen and Sjoerd J Vosse for helping with formatting all the raw data into the MIAME format, and Meike de Wit and Pien van Diemen for helping in the immunohistochemistry analysis.

Funding

This study was financially supported by the Dutch Cancer Society (KWF2002-2618).

References

1. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61(5):759-67.
2. Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. Cancer 1975;36(6):2251-70.
3. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature 1997;386(6625):623-7.
4. Umar A, Boland CR, Terdiman JP et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96(4):261-8.
5. Edelmann L, Edelmann W. Loss of DNA mismatch repair function and cancer predisposition in the mouse: animal models for human hereditary nonpolyposis colorectal cancer. Am J Med Genet C Semin Med Genet 2004;129(1):91-9.
6. di Pietro M, Bellver JS, Menigatti M et al. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129(3):1047-59.
7. Hermsen M, Postma C, Baak J et al. Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. Gastroenterology 2002;123(4):1109-19.
8. Weiss MM, Kuipers EJ, Postma C et al. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 2003;22(12):1872-9.
9. Rajagopalan H, Nowak MA, Vogelstein B et al. The significance of unstable chromosomes in colorectal cancer. Nat Rev Cancer 2003;3(9):695-701.
10. Ried T, Knutzen R, Steinbeck R et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer 1996;15(4):234-45.
11. Meijer GA, Hermsen MA, Baak JP et al. Progression from colorectal adenoma to carcinoma is associated with non-random chromosomal gains as detected by comparative genomic hybridisation. J Clin Pathol 1998;51(12):901-9.
12. Douglas EJ, Fiegler H, Rowan A et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. Cancer Res 2004;64(14):4817-25.
13. Nakao K, Mehta KR, Fridlyand J et al. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. Carcinogenesis 2004;25(8):1345-57.
14. De Angelis PM, Clausen OP, Schjolberg A et al. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. Br J Cancer 1999;80(3-4):526-35.
15. Nessling M, Solinas-Toldo S, Wilgenbus KK et al. Mapping of chromosomal imbalances in gastric adenocarcinoma revealed amplified protooncogenes MYCN, MET, WNT2, and ERBB2. Genes Chromosomes Cancer 1998;23(4):307-16.
16. Korn WM, Yasutake T, Kuo WL 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 1999;25(2):82-90.
17. Kallioniemi A, Kallioniemi OP, Piper J et al. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. Proc Natl Acad Sci U S A 1994;91(6):2156-60.
18. Tanner MM, Grenman S, Koul A et al. Frequent amplification of chromosomal region 20q12-q13 in ovarian cancer. Clin Cancer Res 2000;6(5):1833-9.

19. Aust DE, Muders M, Kohler A et al. Prognostic relevance of 20q13 gains in sporadic colorectal cancers: a FISH analysis. Scand J Gastroenterol 2004;39(8):766-72.

20. Wilting SM, Snijders PJ, Meijer GA et al. Increased gene copy numbers at chromosome 20q are frequent in both squamous cell carcinomas and adenocarcinomas of the cervix. J Pathol 2006;209(2):220-30.

21. Weiss MM, Snijders AM, Kuipers EJ et al. Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray comparative genomic hybridization. J Pathol 2003;200(3):320-6.

22. Albertson DG, Ylstra B, Segraves R et al. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet 2000;25(2):144-6.

23. Postma C, Hermsen MA, Coffa J et al. Chromosomal instability in flat adenomas and carcinomas of the colon. J Pathol 2005;205(4):514-21.

24. Weiss MM, Hermsen MA, Meijer GA et al. Comparative genomic hybridisation. Mol Pathol 1999;52(5):243-51.

25. Carvalho B, Buffart TE, Reis RM et al. Mixed gastric carcinomas show similar chromosomal aberrations in both their diffuse and glandular components. Cell Oncol 2006;28(5-6):283-94.

26. Muris JJ, Ylstra B, Cillessen SA et al. Profiling of apoptosis genes allows for clinical stratification of primary nodal diffuse large B-cell lymphomas. Br J Haematol 2007;136(1):38-47.

27. Quackenbush J. Microarray data normalization and transformation. Nat Genet 2002;32 Suppl:496-501.

28. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30(1):207-10.

29. Jong K, Marchiori E, Meijer G et al. Breakpoint identification and smoothing of array comparative genomic hybridization data. Bioinformatics 2004;20(18):3636-7.

30. Menten B, Pattyn F, De Preter K et al. arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays. BMC Bioinformatics 2005;6:124.

31. van de Wiel MA, Smeets SJ, Brakenhoff RH et al. CGHMultiArray: exact P-values for multi-array comparative genomic hybridization data. Bioinformatics 2005;21(14):3193-4.

32. Diskin SJ, Eck T, Greshock J et al. STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. Genome Res 2006;16(9):1149-58.

33. van Wieringen WN, Belien JA, Vosse SJ et al. ACE-it: a tool for genome-wide integration of gene dosage and RNA expression data. Bioinformatics 2006;22(15):1919-20.

34. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25(4):402-8.

35. Dydensborg AB, Herring E, Auclair J et al. Normalizing genes for quantitative RT-PCR in differentiating human intestinal epithelial cells and adenocarcinomas of the colon. Am J Physiol Gastrointest Liver Physiol 2006;290(5):G1067-G1074.

36. Notterman DA, Alon U, Sierk AJ et al. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61(7):3124-30.

37. Lin YM, Furukawa Y, Tsunoda T et al. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21(26):4120-8.

38. Pilarsky C, Wenzig M, Specht T et al. Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. Neoplasia 2004;6(6):744-50.

39. Shimokawa T, Matsushima S, Tsunoda T et al. Identification of TOMM34, which shows elevated expression in the majority of human colon cancers, as a novel drug target. Int J Oncol 2006;29(2):381-6.

40. Platzer P, Upender MB, Wilson K et al. Silence of chromosomal amplifications in colon cancer. Cancer Res 2002;62(4):1134-8.

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

42. Marumoto T, Zhang D, Saya H. Aurora-A - a guardian of poles. Nat Rev Cancer 2005;5(1):42-50.

43. Carter SL, Eklund AC, Kohane IS et al. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet 2006;38(9):1043-8.

44. Hidaka S, Yasutake T, Takeshita H et al. Differences in 20q13.2 copy number between colorectal cancers with and without liver metastasis. Clin Cancer Res 2000;6(7):2712-7.

45. Rooney PH, Boonsong A, McFadyen MC et al. The candidate oncogene ZNF217 is frequently amplified in colon cancer. J Pathol 2004;204(3):282-8.

46. Brinkmann U, Gallo M, Polymeropoulos MH et al. The human CAS (cellular apoptosis susceptibility) gene mapping on chromosome 20q13 is amplified in BT474 breast cancer cells and part of aberrant chromosomes in breast and colon cancer cell lines. Genome Res 1996;6(3):187-94.

47. Siep M, Sleddens-Linkels E, Mulders S et al. Basic helix-loop-helix transcription factor Tcfl5 interacts with the Calmegin gene promoter in mouse spermatogenesis. Nucleic Acids Res 2004;32(21):6425-36.

48. Cai Y, Jin J, Tomomori-Sato C et al. Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. J Biol Chem 2003;278(44):42733-6.

49. Liu W, Shen X, Yang Y et al. Trihydrophobin 1 is a new negative regulator of A-Raf kinase. J Biol Chem 2004;279(11):10167-75.

50. Jorgensen JP, Lauridsen AM, Kristensen P et al. Adrm1, a putative cell adhesion regulating protein, is a novel proteasome-associated factor. J Mol Biol 2006;360(5):1043-52.

51. Bischoff JR, Anderson L, Zhu Y et al. A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. EMBO J 1998;17(11):3052-65.

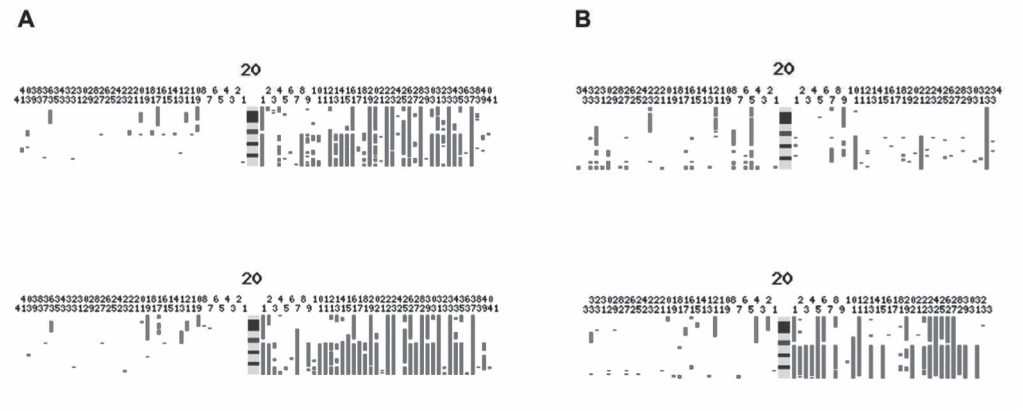
52. Zhou H, Kuang J, Zhong L et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. Nat Genet 1998;20(2):189-93.

53. Hata T, Furukawa T, Sunamura M et al. RNA interference targeting aurora kinase a suppresses tumor growth and enhances the taxane chemosensitivity in human pancreatic cancer cells. Cancer Res 2005;65(7):2899-905.

54. Dardousis K, Voolstra C, Roengvoraphoj M et al. Identification of Differentially Expressed Genes Involved in the Formation of Multicellular Tumor Spheroids by HT-29 Colon Carcinoma Cells. Mol Ther 2007;15(1):94-102.

Supplementary Table 1. Primers sequences and annealing temperature (Anneal.˚C) for the selected genes and housekeeping gene.

Gene symbol	Forward primer sequence	Reverse primer sequence	Anneal. ˚C
ADRM1	5'AGGGTCCAAGCGGCTTTT	5'CGGCAATGCTCCTCATCCT	58
AURKA	5'CCTGAGGAGGAACTGGCATCAA	5'TTCCAAAGCCCACTGCCTCTT	58
C20orf20	5'GAGCACCATGTACGACATGCA	5'ACGAAGTTCCTCTCTGGATTCG	57
C20orf24	5'CTGGTTCCGACAGATCATTGCT	5'CAAGAACCCTCGTAATGGCAAA	58
TCFL5	5'GGAGAGGCATAACCGAATGG	5'AAGAGATTCAACTCATCACAGCAAA	57
TH1L	5'CGCAAAGCAGATTCTATTTTTACTGA	5'GGTATGTGCAATCATCTGTTCCA	58
ß2M	5'TGACTTTGTCACAGCCCAAGATA	5'AATGCGGCATCTTCAAACCT	57



Supplementary figure 1. Karyogram of the copy number aberrations on chromosome 20 in the A) 41 progressed adenomas, adenoma component (top) and carcinoma component (bottom) and B) 34 non-progressed adenomas (top) and 33 carcinomas (bottom). Bars on the left and right represent, losses and gains, respectively.

Supplementary Table 2. Genes significantly overexpressed (FDR ≤0.10) due to copy number dosage (gain) on 20q ordered by chromosomal position (Location in bp according to Freeze July 2003; NCBI Build 34) with HUGO gene symbols and GenBank accession ID.

Gene symbol	GenBank accession nr.	Location (bp position)	FDR
C20orf97	NM_021158	325705	0.02
C20orf140	NM_144628	366092	0.03
C20orf139	NM_080725	577109	0.01
C20orf55	NM_031424	773997	0.09
PI31 (PSMF1)	NM_006814	1093972	0.01
MGC10715	NM_024325	2410624	0.08
MRPS26	NM_030811	2976425	0.006
C20orf27	NM_017874	3687230	0.03
PCNA	NM_002592	5046243	0.01
CDS2	NM_003818	5115333	0.06
LOC149832	AK022998	5126156	0.06
KIAA1434	NM_019593	5476213	0.09
CGI-09	NM_015939	5875080	0.004
CRLS1	AF086526	5924276	0.01
FLJ12676	AK022738	5927485	0.02
LOC54675	NM_019095	5965735	0.01
PLCB1	NM_015192	8670139	0.006
PLCB4	NM_000933	9407945	0.04
C20orf13	NM_017714	13411894	0.0006
KIAA1590	AB046810	16307501	0.07
SNRPB2	NM_003092	16660321	0.04
DSTN	NM_006870	17536305	0.06
RRBP1	NM_004587	17545421	0.03
CSRP2BP	NM_020536	18071256	0.02
RBBP9	NM_006606	18416123	0.08
SEC23B	NM_032986	18489368	0.10
FLJ20941	AK024594	18496078	0.07
NAT5	NM_016100	19946034	0.02
CRNKL1	NM_016652	19966993	0.04
CDABP0105	AY007156	20318412	0.08
C20orf74	BC013749	20423855	0.09
KIAA0255	NM_014742	30212749	0.08
POFUT1	NM_015352	30288385	0.0001
KIF3B	NM_004798	30384299	0.02
KIAA0978	NM_015338	30488639	0.0001
C20orf112	AL122043	30494834	0.0005
DNMT3B	NM_006892	30859803	0.0004
MAPRE1	NM_012325	30900447	0.03
-	BM476468	30910136	0.0002

Gene symbol	GenBank accession nr.	Location (bp position)	FDR
E2F1	NM_005225	31727874	0.09
LOC744651	BC014056	32285332	0.03
AHCY	NM_000687	32342287	0.05
DNCL2A	NM_014183	32592165	0.008
CDC91L1	NM_080476	32612361	0.005
TP53INP2	BC035639	32764711	0.06
NCOA6	NM_014071	32766396	0.0006
FLJ37008	AK094327	32848568	0.007
ACAS2	NM_018677	32973263	0.05
ITGB4BP	NM_002212	33331206	0.003
C20orf44	NM_018244	33354960	0.02
SDBCAG84	NM_015966	33599691	0.03
FER1L4	NM_025206	33609998	0.006
RBM12	NM_006047	33704926	0.04
NFS1	NM_021100	33725686	0.008
C20orf104	NM_016436	33992323	0.08
FLJ23314	AK026967	34304803	0.02
C20orf4	NM_015511	34308105	0.07
KIAA0964	NM_014902	34587733	0.02
MYRL2	NM_006097	34606826	0.001
TGIF2	NM_021809	34653775	2E-05
C20orf24 (RIP5)	NM_018840	34673829	2E-05
NDRG3	NM_032013	34714406	0.06
RBL1	NM_002895	35084572	0.001
RPN2	NM_002951	35260905	0.06
BLCAP	NM_006698	35579832	0.03
CTNNBL1	NM_030877	35827021	0.001
C20orf102	NM_080607	36007097	0.09
KIAA0406	NM_014657	36045374	0.03
C20orf77	NM_021215	36127986	0.0006
TGM2	NM_004613	36191453	0.03
FLJ12683	AK022745	36482793	0.01
U71b	Y11166	36487314	0.0006
LOC128439	NM_139016	36510727	0.03
KIAA1219	AB033045	36640313	0.01
FLJ12785	NM_024855	36812237	0.01
C20orf129	NM_030919	37014104	0.01
DDX35	NM_021931	37101039	0.04
FLJ11409	AK021471	39209554	0.05
PLCG1	NM_002660	39229924	0.03
KIAA0395	AB007855	39246999	0.07

Gene symbol	GenBank accession nr.	Location (bp position)	FDR
C20orf9	NM_016004	41675968	0.002
MYBL2	NM_002466	41754244	0.01
C20orf111	NM_016470	42259525	0.002
-	BC019858	42287375	0.02
C20orf142	BC029662	42364982	0.05
-	AW386993	42494154	0.0002
FLJ31616	AK056178	42556305	0.05
TDE1	NM_006811	42568967	0.04
YWHAB	NM_003404	42949715	0.01
GW128	NM_014052	42969000	0.0001
FLJ30809	AK055371	43001197	0.09
TOMM34	NM_006809	43004464	0.01
RBPSUHL	NM_014276	43374284	0.02
SDC4	BQ183780	43407546	0.07
C20orf35	NM_018478	43472553	0.01
C20orf167	NM_052951	43857408	0.0008
UBE2C	NM_007019	43876451	0.01
PTE1	NM_005469	43903875	0.02
C20orf164	NM_080752	43941017	0.01
C20orf162	NM_080603	43945063	0.04
C20orf165	NM_080608	43948747	0.10
OVCOV1	NM_015945	44417131	0.05
C20orf64	NM_033550	44748895	0.0003
NCOA3	NM_006534	45717814	0.003
ARFGEF2	NM_006420	47083407	0.03
CSE1L	NM_001316	47134042	0.01
DDX27	NM_017895	47293760	0.0006
B4GALT5	NM_004776	47684169	0.06
KIAA0939	AB023156	47940156	0.01
KIAA0757	NM_006038	47956096	0.03
ZNF313	NM_018683	47995349	0.07
CEBPB	NM_005194	48241161	0.04
PAR-6 beta	AB044555	48787877	0.05
DPM1	NM_003859	48985006	2E-05
MOCS3	NM_014484	49010133	0.0007
-	AF095854	49148492	0.08
FLJ22105	AK025758	49438313	0.02
KIAA0611	AB014511	49648670	0.04
ZFP64	NM_022088	50201382	0.01
ZNF217	NM_006526	51617455	0.03
AURKA (STK6)	NM_003600	54394854	0.006

Gene symbol	GenBank accession nr.	Location (bp position)	FDR
CSTF1	NM_001324	54411973	0.01
C20orf43	NM_016407	54526586	0.002
FLJ37465	AK094784	55177288	0.04
RAE1	NM_003610	55386523	0.01
RNPC1	NM_017495	55417321	0.02
TMEPAI	NM_020182	55657093	0.02
VAPB	NM_004738	56452680	0.03
-	CA312715	56459415	0.03
FLJ90166	NM_153360	56468553	0.09
GNAS	NM_080425	56863685	0.007
-	BU682808	56902873	0.09
-	BM512279	56904930	0.06
TH1L	NM_016397	57001935	0.0001
ATP5E	NM_006886	57037199	0.001
C20orf45	NM_016045	57043406	0.0004
EDN3	NM_000114	57334347	0.01
PPP1R3D	NM_006242	57947113	0.003
C20orf177	AL137442	57956376	0.02
TAF4	NM_003185	59983947	0.002
FLJ25473	NM_144703	60139052	0.09
C20orf40	NM_014054	60143137	0.02
PSMA7	NM_152255	60148148	0.004
ADRM1	NM_007002	60317216	0.05
FLJ25011	AK057740	60394790	0.03
CABLES2	BC003122	60398884	0.09
SLC21A12	NM_016354	60773742	0.001
C20orf20	NM_018270	60901606	0.0004
COL9A3	NM_001853	60938880	0.06
TCFL5	NM_006602	60943254	0.001
DATF1	NM_022105	61011623	0.0007