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 TRlZol (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 deparaffinization 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 2 Mb (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, ADMR1, 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.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>GenBank accession no.</th>
<th>Location (bp position)</th>
<th>Wilcoxon P-value</th>
<th>Thas P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C20orf1(TPX2)</td>
<td>NM_012112</td>
<td>31103374</td>
<td>2E-06</td>
<td>8E-05</td>
</tr>
<tr>
<td>MYRL2</td>
<td>NM_006097</td>
<td>35859501</td>
<td>5E-06</td>
<td>4E-05</td>
</tr>
<tr>
<td>C20orf24 (RIP5)</td>
<td>NM_018840</td>
<td>35923014</td>
<td>2E-07</td>
<td>2E-05</td>
</tr>
<tr>
<td>TOMM34</td>
<td>NM_006809</td>
<td>44265329</td>
<td>8E-08</td>
<td>0</td>
</tr>
<tr>
<td>RBPSUHL</td>
<td>NM_014276</td>
<td>44626010</td>
<td>2E-07</td>
<td>6E-06</td>
</tr>
<tr>
<td>BCA54</td>
<td>NM_017843</td>
<td>50138063</td>
<td>2E-06</td>
<td>6E-05</td>
</tr>
<tr>
<td>AURKA (STK6)</td>
<td>NM_003600</td>
<td>55641283</td>
<td>4E-10</td>
<td>0</td>
</tr>
<tr>
<td>TFJ37465 (BMP7)</td>
<td>AK094784</td>
<td>56477906</td>
<td>1E-09</td>
<td>0</td>
</tr>
<tr>
<td>RNPC1</td>
<td>NM_017495</td>
<td>56660843</td>
<td>8E-07</td>
<td>7E-05</td>
</tr>
<tr>
<td>THIL</td>
<td>NM_016397</td>
<td>58253070</td>
<td>1E-06</td>
<td>1E-05</td>
</tr>
<tr>
<td>ADRM1</td>
<td>NM_007002</td>
<td>61566389</td>
<td>9E-07</td>
<td>8E-05</td>
</tr>
<tr>
<td>C20orf20</td>
<td>NM_018270</td>
<td>62156238</td>
<td>9E-09</td>
<td>0</td>
</tr>
<tr>
<td>TCFL5</td>
<td>NM_006602</td>
<td>62211152</td>
<td>2E-09</td>
<td>0</td>
</tr>
<tr>
<td>C20orf11</td>
<td>NM_017896</td>
<td>62299593</td>
<td>4E-07</td>
<td>0</td>
</tr>
</tbody>
</table>

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, THIL, 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.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>GenBank accession no.</th>
<th>Location (bp position)</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM13</td>
<td>NM_030789</td>
<td>30874805</td>
<td>0.03</td>
</tr>
<tr>
<td>C20orf1 (TPX2)</td>
<td>NM_012112</td>
<td>31103374</td>
<td>0.03</td>
</tr>
<tr>
<td>CDC91L1</td>
<td>NM_080476</td>
<td>33922394</td>
<td>0.02</td>
</tr>
<tr>
<td>C20orf44</td>
<td>NM_018244</td>
<td>34608051</td>
<td>0.07</td>
</tr>
<tr>
<td>DLGAP4</td>
<td>NM_014902</td>
<td>35761669</td>
<td>0.05</td>
</tr>
<tr>
<td>TGIF2</td>
<td>NM_021809</td>
<td>35897616</td>
<td>0.03</td>
</tr>
<tr>
<td>C20orf24 (RIP5)</td>
<td>NM_018840</td>
<td>35923014</td>
<td>0.0006</td>
</tr>
<tr>
<td>YWHAB</td>
<td>NM_014052</td>
<td>44210177</td>
<td>0.0002</td>
</tr>
<tr>
<td>AURKA (STK6)</td>
<td>NM_003600</td>
<td>55641283</td>
<td>0.02</td>
</tr>
<tr>
<td>RNPC1</td>
<td>NM_017495</td>
<td>56660843</td>
<td>0.04</td>
</tr>
<tr>
<td>TH1L</td>
<td>NM_016397</td>
<td>58253070</td>
<td>0.007</td>
</tr>
<tr>
<td>ADRM1</td>
<td>NM_007002</td>
<td>61566389</td>
<td>0.05</td>
</tr>
<tr>
<td>SLCO4A1</td>
<td>NM_016354</td>
<td>62015102</td>
<td>0.08</td>
</tr>
<tr>
<td>C20orf20</td>
<td>NM_018270</td>
<td>62156238</td>
<td>0.04</td>
</tr>
<tr>
<td>TCFL5</td>
<td>NM_006602</td>
<td>62211152</td>
<td>0.03</td>
</tr>
<tr>
<td>C20orf11</td>
<td>NM_017896</td>
<td>62299593</td>
<td>0.0009</td>
</tr>
<tr>
<td>C20orf59</td>
<td>NM_022082</td>
<td>62323360</td>
<td>0.007</td>
</tr>
<tr>
<td>PRPF6</td>
<td>NM_012469</td>
<td>63364789</td>
<td>0.03</td>
</tr>
</tbody>
</table>

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 RNPCI (NM_017495) and TCFLS (NM_006602), by lesion (grey circles: adenomas; black circles: carcinomas) showing a good separation of colorectal adenomas vs adenocarcinomas.

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

<table>
<thead>
<tr>
<th>Lesion</th>
<th>Predicted Group Membership</th>
<th>Adenoma</th>
<th>Carcinoma</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Count</td>
<td>Adenoma</td>
<td>35</td>
<td>2</td>
<td>37</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>6</td>
<td>25</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>% Adenoma</td>
<td>94.6</td>
<td>5.4</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>19.4</td>
<td>80.6</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

From 68 tumours in total, 60 were correctly classified (88.2%), using expression levels of RNPCI and TCFLS 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)

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

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

<table>
<thead>
<tr>
<th>AURKA staining</th>
<th>Negative</th>
<th>Weak</th>
<th>Strong</th>
<th>Total</th>
<th>p value *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesion Adenoma</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Carcinoma</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>21</td>
<td>7</td>
<td>44</td>
<td>0.01</td>
</tr>
</tbody>
</table>

* 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 (54.5 Mb) 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
Putative oncogenes at chromosome 20q in colorectal carcinogenesis

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 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).
Putative oncogenes at chromosome 20q in colorectal carcinogenesis

References


Putative oncogenes at chromosome 20q in colorectal carcinogenesis | 125

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

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
<th>Anneal.˚C</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADRM1</td>
<td>5’AGGGTCCAAGCGGCTTTT</td>
<td>5’CGGCAATGCTCCTCATCCT</td>
<td>58</td>
</tr>
<tr>
<td>AURKA</td>
<td>5’CCTGAGGAGGAACCTGCCCTA</td>
<td>5’TCTCCAAAGCCCACCTCCTTT</td>
<td>58</td>
</tr>
<tr>
<td>C20orf20</td>
<td>5’GAGCACCATTACGCATCCTCA</td>
<td>5’ACGAAGTTCTCTCTTTGATC</td>
<td>57</td>
</tr>
<tr>
<td>C20orf24</td>
<td>5’CTGGTGTCGCCAGATCATATTGT</td>
<td>5’CAAGAACCCTCGTAATGGCAA</td>
<td>58</td>
</tr>
<tr>
<td>TCFL5</td>
<td>5’GGAGAGGCAATACCGAATTG</td>
<td>5’AAGAGATTCAACCTCACACAGC</td>
<td>57</td>
</tr>
<tr>
<td>TH1L</td>
<td>5’CGCAAAGCACATTTTCTTCTTA</td>
<td>5’GGTATGTGCAATCATCTGTTC</td>
<td>58</td>
</tr>
<tr>
<td>B2M</td>
<td>5’TGACTTTTGCCAGCCCAAGATA</td>
<td>5’AATGCAGGCATCTTAAACCT</td>
<td>57</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>GenBank accession nr.</th>
<th>Location (bp position)</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>C20orf97</td>
<td>NM_021158</td>
<td>325705</td>
<td>0.02</td>
</tr>
<tr>
<td>C20orf140</td>
<td>NM_144628</td>
<td>366092</td>
<td>0.03</td>
</tr>
<tr>
<td>C20orf139</td>
<td>NM_080725</td>
<td>577109</td>
<td>0.01</td>
</tr>
<tr>
<td>C20orf55</td>
<td>NM_031424</td>
<td>773997</td>
<td>0.09</td>
</tr>
<tr>
<td>PI31 (PSMF1)</td>
<td>NM_006814</td>
<td>1093972</td>
<td>0.01</td>
</tr>
<tr>
<td>MGC10715</td>
<td>NM_024325</td>
<td>2410624</td>
<td>0.08</td>
</tr>
<tr>
<td>MRPS26</td>
<td>NM_030811</td>
<td>2976425</td>
<td>0.006</td>
</tr>
<tr>
<td>C20orf27</td>
<td>NM_017874</td>
<td>3687230</td>
<td>0.03</td>
</tr>
<tr>
<td>PCNA</td>
<td>NM_002592</td>
<td>5046243</td>
<td>0.01</td>
</tr>
<tr>
<td>CDS2</td>
<td>NM_003818</td>
<td>5115333</td>
<td>0.06</td>
</tr>
<tr>
<td>LOC149832</td>
<td>AK022998</td>
<td>5126156</td>
<td>0.06</td>
</tr>
<tr>
<td>KIAA1434</td>
<td>NM_019593</td>
<td>5476213</td>
<td>0.09</td>
</tr>
<tr>
<td>CGI-09</td>
<td>NM_015939</td>
<td>5875080</td>
<td>0.004</td>
</tr>
<tr>
<td>CRLS1</td>
<td>AF086526</td>
<td>5924276</td>
<td>0.01</td>
</tr>
<tr>
<td>FLJ12676</td>
<td>AK022738</td>
<td>5927485</td>
<td>0.02</td>
</tr>
<tr>
<td>LOC54675</td>
<td>NM_019095</td>
<td>5965735</td>
<td>0.01</td>
</tr>
<tr>
<td>PLCB1</td>
<td>NM_015192</td>
<td>8670139</td>
<td>0.006</td>
</tr>
<tr>
<td>PLCB4</td>
<td>NM_000933</td>
<td>9407945</td>
<td>0.04</td>
</tr>
<tr>
<td>C20orf13</td>
<td>NM_017714</td>
<td>13411894</td>
<td>0.0006</td>
</tr>
<tr>
<td>KIAA1590</td>
<td>AB046810</td>
<td>16307501</td>
<td>0.07</td>
</tr>
<tr>
<td>SNRPN2</td>
<td>NM_003092</td>
<td>16660321</td>
<td>0.04</td>
</tr>
<tr>
<td>DSTN</td>
<td>NM_006870</td>
<td>17536305</td>
<td>0.06</td>
</tr>
<tr>
<td>RRB1</td>
<td>NM_004587</td>
<td>17545421</td>
<td>0.03</td>
</tr>
<tr>
<td>CSRP2BP</td>
<td>NM_020536</td>
<td>18071256</td>
<td>0.02</td>
</tr>
<tr>
<td>RBBP9</td>
<td>NM_006606</td>
<td>18416123</td>
<td>0.08</td>
</tr>
<tr>
<td>SEC23B</td>
<td>NM_032986</td>
<td>18489368</td>
<td>0.10</td>
</tr>
<tr>
<td>FLJ20941</td>
<td>AK024594</td>
<td>18496078</td>
<td>0.07</td>
</tr>
<tr>
<td>NAT5</td>
<td>NM_016100</td>
<td>19946034</td>
<td>0.02</td>
</tr>
<tr>
<td>CRNKLI</td>
<td>NM_016652</td>
<td>19966993</td>
<td>0.04</td>
</tr>
<tr>
<td>CDABP0105</td>
<td>AY007156</td>
<td>20318412</td>
<td>0.08</td>
</tr>
<tr>
<td>C20orf74</td>
<td>BC013749</td>
<td>20423855</td>
<td>0.09</td>
</tr>
<tr>
<td>KIAA0255</td>
<td>NM_014742</td>
<td>30212749</td>
<td>0.08</td>
</tr>
<tr>
<td>POFUT1</td>
<td>NM_015352</td>
<td>30288385</td>
<td>0.0001</td>
</tr>
<tr>
<td>KIF3B</td>
<td>NM_004798</td>
<td>30384299</td>
<td>0.02</td>
</tr>
<tr>
<td>KIAA097B</td>
<td>NM_015338</td>
<td>30488639</td>
<td>0.0001</td>
</tr>
<tr>
<td>C20orf112</td>
<td>AL122043</td>
<td>30494834</td>
<td>0.0005</td>
</tr>
<tr>
<td>DNMT3B</td>
<td>NM_006892</td>
<td>30859803</td>
<td>0.0004</td>
</tr>
<tr>
<td>MAPRE1</td>
<td>NM_012325</td>
<td>30900447</td>
<td>0.03</td>
</tr>
<tr>
<td>-</td>
<td>BM476468</td>
<td>30910136</td>
<td>0.0002</td>
</tr>
<tr>
<td>Gene symbol</td>
<td>GenBank accession nr.</td>
<td>Location (bp position)</td>
<td>FDR</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>------</td>
</tr>
<tr>
<td>E2F1</td>
<td>NM_005225</td>
<td>31727874</td>
<td>0.09</td>
</tr>
<tr>
<td>LOC744651</td>
<td>BC014056</td>
<td>32285332</td>
<td>0.03</td>
</tr>
<tr>
<td>AHCY</td>
<td>NM_000687</td>
<td>32342287</td>
<td>0.05</td>
</tr>
<tr>
<td>DNCL2A</td>
<td>NM_014183</td>
<td>32592165</td>
<td>0.008</td>
</tr>
<tr>
<td>CDC91L1</td>
<td>NM_080476</td>
<td>32612361</td>
<td>0.005</td>
</tr>
<tr>
<td>TPS3INP2</td>
<td>BC035639</td>
<td>32764711</td>
<td>0.06</td>
</tr>
<tr>
<td>NCOA6</td>
<td>NM_014071</td>
<td>32766396</td>
<td>0.0006</td>
</tr>
<tr>
<td>FLJ37008</td>
<td>AK094327</td>
<td>32848568</td>
<td>0.007</td>
</tr>
<tr>
<td>ACAS2</td>
<td>NM_018677</td>
<td>32973263</td>
<td>0.05</td>
</tr>
<tr>
<td>ITGB4BP</td>
<td>NM_002212</td>
<td>33331206</td>
<td>0.003</td>
</tr>
<tr>
<td>C20orf44</td>
<td>NM_018244</td>
<td>33354960</td>
<td>0.02</td>
</tr>
<tr>
<td>SDBCAG84</td>
<td>NM_015966</td>
<td>33599691</td>
<td>0.03</td>
</tr>
<tr>
<td>FERIL4</td>
<td>NM_025206</td>
<td>33609998</td>
<td>0.006</td>
</tr>
<tr>
<td>RBM12</td>
<td>NM_006047</td>
<td>33704926</td>
<td>0.04</td>
</tr>
<tr>
<td>NFS1</td>
<td>NM_021100</td>
<td>33725686</td>
<td>0.008</td>
</tr>
<tr>
<td>C20orf104</td>
<td>NM_016436</td>
<td>33992323</td>
<td>0.08</td>
</tr>
<tr>
<td>FLJ23314</td>
<td>AK026967</td>
<td>34304803</td>
<td>0.02</td>
</tr>
<tr>
<td>C20orf4</td>
<td>NM_015511</td>
<td>34308105</td>
<td>0.07</td>
</tr>
<tr>
<td>KIAA0964</td>
<td>NM_014902</td>
<td>34587733</td>
<td>0.02</td>
</tr>
<tr>
<td>MYRL2</td>
<td>NM_006097</td>
<td>34606826</td>
<td>0.001</td>
</tr>
<tr>
<td>TGIF2</td>
<td>NM_021809</td>
<td>34653775</td>
<td>2E-05</td>
</tr>
<tr>
<td>C20orf24 (RIPS)</td>
<td>NM_018840</td>
<td>34673829</td>
<td>2E-05</td>
</tr>
<tr>
<td>NDRG3</td>
<td>NM_032013</td>
<td>34714406</td>
<td>0.06</td>
</tr>
<tr>
<td>RBL1</td>
<td>NM_002895</td>
<td>35084572</td>
<td>0.001</td>
</tr>
<tr>
<td>RPN2</td>
<td>NM_002951</td>
<td>35260905</td>
<td>0.06</td>
</tr>
<tr>
<td>BLCAP</td>
<td>NM_006698</td>
<td>35579832</td>
<td>0.03</td>
</tr>
<tr>
<td>CTNNBL1</td>
<td>NM_030877</td>
<td>35827021</td>
<td>0.001</td>
</tr>
<tr>
<td>C20orf102</td>
<td>NM_080607</td>
<td>36007097</td>
<td>0.09</td>
</tr>
<tr>
<td>KIAA0406</td>
<td>NM_014657</td>
<td>36045374</td>
<td>0.03</td>
</tr>
<tr>
<td>C20orf77</td>
<td>NM_021215</td>
<td>36127986</td>
<td>0.0006</td>
</tr>
<tr>
<td>TGM2</td>
<td>NM_004613</td>
<td>36191453</td>
<td>0.03</td>
</tr>
<tr>
<td>FLJ12683</td>
<td>AK022745</td>
<td>36482793</td>
<td>0.01</td>
</tr>
<tr>
<td>U71b</td>
<td>Y11166</td>
<td>36487314</td>
<td>0.0006</td>
</tr>
<tr>
<td>LOC128439</td>
<td>NM_139016</td>
<td>36510727</td>
<td>0.03</td>
</tr>
<tr>
<td>KIAA1219</td>
<td>AB033045</td>
<td>36640313</td>
<td>0.01</td>
</tr>
<tr>
<td>FLJ12785</td>
<td>NM_024855</td>
<td>36812237</td>
<td>0.01</td>
</tr>
<tr>
<td>C20orf129</td>
<td>NM_030919</td>
<td>37014104</td>
<td>0.01</td>
</tr>
<tr>
<td>DDX35</td>
<td>NM_021931</td>
<td>37101039</td>
<td>0.04</td>
</tr>
<tr>
<td>FLJ11409</td>
<td>AK021471</td>
<td>39209554</td>
<td>0.05</td>
</tr>
<tr>
<td>PLCG1</td>
<td>NM_002660</td>
<td>39229924</td>
<td>0.03</td>
</tr>
<tr>
<td>KIAA0395</td>
<td>AB007855</td>
<td>39246999</td>
<td>0.07</td>
</tr>
<tr>
<td>Gene symbol</td>
<td>GenBank accession</td>
<td>Location (bp position)</td>
<td>FDR</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>C20orf9</td>
<td>NM_016004</td>
<td>41675968</td>
<td>0.002</td>
</tr>
<tr>
<td>MYBL2</td>
<td>NM_002466</td>
<td>41754244</td>
<td>0.01</td>
</tr>
<tr>
<td>C20orf111</td>
<td>NM_016470</td>
<td>42259525</td>
<td>0.002</td>
</tr>
<tr>
<td>-</td>
<td>BC019858</td>
<td>42287375</td>
<td>0.02</td>
</tr>
<tr>
<td>C20orf142</td>
<td>BC029662</td>
<td>42364982</td>
<td>0.05</td>
</tr>
<tr>
<td>-</td>
<td>AW386993</td>
<td>42494154</td>
<td>0.0002</td>
</tr>
<tr>
<td>FLJ31616</td>
<td>AK056178</td>
<td>42556305</td>
<td>0.05</td>
</tr>
<tr>
<td>TEDE1</td>
<td>NM_006811</td>
<td>42568967</td>
<td>0.04</td>
</tr>
<tr>
<td>YWHAB</td>
<td>NM_003404</td>
<td>42949715</td>
<td>0.01</td>
</tr>
<tr>
<td>GW128</td>
<td>NM_014052</td>
<td>42969000</td>
<td>0.001</td>
</tr>
<tr>
<td>FLJ30809</td>
<td>AK055371</td>
<td>43001197</td>
<td>0.09</td>
</tr>
<tr>
<td>TOMM34</td>
<td>NM_006809</td>
<td>43004464</td>
<td>0.01</td>
</tr>
<tr>
<td>RBP5UHL</td>
<td>NM_014276</td>
<td>43374284</td>
<td>0.02</td>
</tr>
<tr>
<td>SDC4</td>
<td>BQ183780</td>
<td>43407546</td>
<td>0.07</td>
</tr>
<tr>
<td>C20orf35</td>
<td>NM_018478</td>
<td>43472553</td>
<td>0.01</td>
</tr>
<tr>
<td>C20orf167</td>
<td>NM_052951</td>
<td>43857408</td>
<td>0.0008</td>
</tr>
<tr>
<td>UBE2C</td>
<td>NM_007019</td>
<td>43876451</td>
<td>0.01</td>
</tr>
<tr>
<td>PTE1</td>
<td>NM_005469</td>
<td>43903875</td>
<td>0.02</td>
</tr>
<tr>
<td>C20orf164</td>
<td>NM_080752</td>
<td>43941017</td>
<td>0.01</td>
</tr>
<tr>
<td>C20orf162</td>
<td>NM_080603</td>
<td>43945063</td>
<td>0.04</td>
</tr>
<tr>
<td>C20orf165</td>
<td>NM_080608</td>
<td>43948747</td>
<td>0.10</td>
</tr>
<tr>
<td>OVC01</td>
<td>NM_015945</td>
<td>44417131</td>
<td>0.05</td>
</tr>
<tr>
<td>C20orf64</td>
<td>NM_033550</td>
<td>44748895</td>
<td>0.0003</td>
</tr>
<tr>
<td>NCOA3</td>
<td>NM_006534</td>
<td>45717814</td>
<td>0.003</td>
</tr>
<tr>
<td>ARFGEF2</td>
<td>NM_006420</td>
<td>47083407</td>
<td>0.03</td>
</tr>
<tr>
<td>CSE1L</td>
<td>NM_001316</td>
<td>47134042</td>
<td>0.01</td>
</tr>
<tr>
<td>DDX27</td>
<td>NM_017895</td>
<td>47293760</td>
<td>0.0006</td>
</tr>
<tr>
<td>B4GALT5</td>
<td>NM_004776</td>
<td>47684169</td>
<td>0.06</td>
</tr>
<tr>
<td>KIAA0939</td>
<td>AB023156</td>
<td>47940156</td>
<td>0.01</td>
</tr>
<tr>
<td>KIAA0757</td>
<td>NM_006038</td>
<td>47956096</td>
<td>0.03</td>
</tr>
<tr>
<td>ZNF313</td>
<td>NM_018683</td>
<td>47995349</td>
<td>0.07</td>
</tr>
<tr>
<td>CEBPB</td>
<td>NM_005194</td>
<td>48241161</td>
<td>0.04</td>
</tr>
<tr>
<td>PAR-6 beta</td>
<td>AB044555</td>
<td>48787877</td>
<td>0.05</td>
</tr>
<tr>
<td>DPM1</td>
<td>NM_003859</td>
<td>48985006</td>
<td>2E-05</td>
</tr>
<tr>
<td>MOCS3</td>
<td>NM_014484</td>
<td>49010133</td>
<td>0.0007</td>
</tr>
<tr>
<td>-</td>
<td>AF095854</td>
<td>49148492</td>
<td>0.08</td>
</tr>
<tr>
<td>FLJ22105</td>
<td>AK025758</td>
<td>49438313</td>
<td>0.02</td>
</tr>
<tr>
<td>KIAA0611</td>
<td>AB014511</td>
<td>49648670</td>
<td>0.04</td>
</tr>
<tr>
<td>ZFP64</td>
<td>NM_022088</td>
<td>50201382</td>
<td>0.01</td>
</tr>
<tr>
<td>ZNF217</td>
<td>NM_006526</td>
<td>51617455</td>
<td>0.03</td>
</tr>
<tr>
<td>AURKA (STK6)</td>
<td>NM_003600</td>
<td>54394854</td>
<td>0.006</td>
</tr>
<tr>
<td>Gene symbol</td>
<td>GenBank accession nr.</td>
<td>Location (bp position)</td>
<td>FDR</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------</td>
<td>------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>CSTF1</td>
<td>NM_001324</td>
<td>54411973</td>
<td>0.01</td>
</tr>
<tr>
<td>C20orf43</td>
<td>NM_016407</td>
<td>54526586</td>
<td>0.002</td>
</tr>
<tr>
<td>FLJ37465</td>
<td>AK094784</td>
<td>55177288</td>
<td>0.04</td>
</tr>
<tr>
<td>RAE1</td>
<td>NM_003610</td>
<td>55386523</td>
<td>0.01</td>
</tr>
<tr>
<td>RNPC1</td>
<td>NM_017495</td>
<td>55417321</td>
<td>0.02</td>
</tr>
<tr>
<td>TMEPAI</td>
<td>NM_020182</td>
<td>55657093</td>
<td>0.02</td>
</tr>
<tr>
<td>VAPB</td>
<td>NM_004738</td>
<td>56452680</td>
<td>0.03</td>
</tr>
<tr>
<td>FLJ90166</td>
<td>CA312715</td>
<td>56459415</td>
<td>0.03</td>
</tr>
<tr>
<td>GNAS</td>
<td>NM_080425</td>
<td>56863685</td>
<td>0.007</td>
</tr>
<tr>
<td>-</td>
<td>BU682808</td>
<td>56902873</td>
<td>0.09</td>
</tr>
<tr>
<td>TH1L</td>
<td>NM_016397</td>
<td>57001935</td>
<td>0.0001</td>
</tr>
<tr>
<td>ATP5E</td>
<td>NM_006886</td>
<td>57037199</td>
<td>0.001</td>
</tr>
<tr>
<td>C20orf45</td>
<td>NM_016045</td>
<td>57043406</td>
<td>0.0004</td>
</tr>
<tr>
<td>EDN3</td>
<td>NM_000114</td>
<td>57334347</td>
<td>0.01</td>
</tr>
<tr>
<td>PPP1R3D</td>
<td>NM_006242</td>
<td>57947113</td>
<td>0.003</td>
</tr>
<tr>
<td>C20orf177</td>
<td>AL137442</td>
<td>57956376</td>
<td>0.02</td>
</tr>
<tr>
<td>TAF4</td>
<td>NM_003185</td>
<td>59983947</td>
<td>0.002</td>
</tr>
<tr>
<td>FLJ25473</td>
<td>NM_144703</td>
<td>60139052</td>
<td>0.09</td>
</tr>
<tr>
<td>C20orf40</td>
<td>NM_014054</td>
<td>60143137</td>
<td>0.02</td>
</tr>
<tr>
<td>PSMA7</td>
<td>NM_152255</td>
<td>60148148</td>
<td>0.004</td>
</tr>
<tr>
<td>ADRM1</td>
<td>NM_007002</td>
<td>60317216</td>
<td>0.05</td>
</tr>
<tr>
<td>FLJ25011</td>
<td>AK057740</td>
<td>60394790</td>
<td>0.03</td>
</tr>
<tr>
<td>CABLES2</td>
<td>BC003122</td>
<td>60398884</td>
<td>0.09</td>
</tr>
<tr>
<td>SLC21A12</td>
<td>NM_016354</td>
<td>60773742</td>
<td>0.001</td>
</tr>
<tr>
<td>C20orf20</td>
<td>NM_018270</td>
<td>60901606</td>
<td>0.0004</td>
</tr>
<tr>
<td>COL9A3</td>
<td>NM_001853</td>
<td>60938880</td>
<td>0.06</td>
</tr>
<tr>
<td>TCFL5</td>
<td>NM_006602</td>
<td>60943254</td>
<td>0.001</td>
</tr>
<tr>
<td>DATF1</td>
<td>NM_022105</td>
<td>61011623</td>
<td>0.0007</td>
</tr>
</tbody>
</table>