Colorectal adenoma to carcinoma progression is accompanied by changes in gene expression associated with ageing, chromosomal instability, and fatty acid metabolism

Beatriz Carvalho · Anke H. Sillars-Hardebol · Cindy Postma · Sandra Mongera · Jochim Terhaar Sive Droste · Askar Obulkasim · Mark van de Wiel · Wim van Criekinge · Bauke Ylstra · Remond J. A. Fijneman · Gerrit A. Meijer

Accepted: 28 November 2011 / Published online: 26 January 2012 © International Society for Cellular Oncology 2012

Abstract

Background Colorectal cancer develops in a multi-step manner from normal epithelium, through a pre-malignant lesion (so-called adenoma), into a malignant lesion (carcinoma), which invades surrounding tissues and eventually can spread systemically (metastasis). It is estimated that only about 5% of adenomas do progress to a carcinoma.

Aim The present study aimed to unravel the biology of adenoma to carcinoma progression by mRNA expression profiling, and to identify candidate biomarkers for adenomas that are truly at high risk of progression.

Methods Genome-wide mRNA expression profiles were obtained from a series of 37 colorectal adenomas and 31 colorectal carcinomas using oligonucleotide microarrays. Differentially expressed genes were validated in an independent colorectal gene expression data set. Gene Set Enrichment Analysis (GSEA) was used to identify altered expression of sets of genes associated with specific biological processes, in order to better understand the biology of colorectal adenoma to carcinoma progression.

Results mRNA expression of 248 genes was significantly different, of which 96 were upregulated and 152 downregulated in carcinomas compared to adenomas. Classification of adenomas and carcinomas using the expression of these genes showed to be very accurate, also when tested in an independent expression data set. Gene-sets associated with ageing (which is related to senescence) and chromosomal instability were upregulated, and a gene-set associated with fatty acid metabolism was downregulated in carcinomas compared to adenomas. Moreover, gene-sets associated with chromosomal location revealed chromosome 4q22 loss and chromosome 20q gain of gene-set expression as being relevant in this progression.

Concluding remark These data are consistent with the notion that adenomas and carcinomas are distinct biological entities. Disruption of specific biological processes like senescence (ageing), maintenance of chromosomal instability and altered metabolism, are key factors in the progression from adenoma to carcinoma.

Keywords Colorectal cancer · Progression · Gene expression · Ageing · Chromosomal instability · Fatty acid metabolism
1 Introduction

Colorectal cancer (CRC) is a major health concern worldwide. In Europe, CRC ranks second in terms of cancer incidence and mortality, after lung cancer [1]. Most often the disease is only diagnosed at a late stage, when clinical symptoms arise.

CRC develops in a multi-step manner from normal epithelium, through a pre-malignant lesion (so-called adenoma), into a malignant lesion (carcinoma), which invades surrounding tissues and eventually can spread systemically (metastasis). While pre-malignant lesions, adenomas, can easily be detected by colonoscopy, most of these adenomas have a low risk of progressing into cancer. It is estimated that this happens in about 5% of adenomas [2, 3]. Unravelling the biology of adenoma to carcinoma progression is crucial for identifying biomarkers for adenomas that are truly at high risk of progression.

By genome-wide analysis of DNA copy number changes we have shown that adenomas harbouring a focus of carcinoma (progressed adenomas), already contain in their phenotypically still benign components most chromosomal aberrations that are present in their invasive counterpart [4, 5]. This indicates that these chromosomal regions harbour genes that play a role in the adenoma-to-carcinoma progression.

Many genes have been implicated in the development of CRC. Initially, APC, KRAS, TP53 and DCC have been described in the progression model introduced by Fearon and Vogelstein [6]. More recently, by sequencing a panel of 11 colorectal cancers, Wood et al [7] identified 140 candidate cancer genes (CAN genes) mutated in CRC. As gene expression can be affected in multiple ways, including DNA copy number dosage and epigenetic silencing, studies integrating both copy number changes and point mutations [8, 9] as well as mutations and hypermethylation of promoter regions of genes [10] have enabled the discovery of additional genes and pathways relevant to CRC development. One example of a clinically relevant gene involved in CRC progression is AURKA, which is both amplified and over-expressed in the majority of CRCs [5, 11, 12] and is a target for anti-cancer drugs [13].

Microarray expression profiling is a robust technique to analyse the expression of thousands of genes simultaneously. Gene expression in CRC has been widely studied with microarrays, either comparing carcinomas to normal colon tissue [14–18] or by comparing microsatellite instable with microsatellite stable CRCs [19, 20]. In addition, the mechanisms of metastasis [21–25], prediction of recurrence risk of stage II and stage III CRCs [26–29] and response to treatment in advanced CRC patients [30] have been investigated using expression profiling. To date, a limited number of studies have focused on the differential expression between colorectal adenomas and carcinomas and in most of them the number of samples analysed was rather limited, especially concerning the number of adenomas [31–34].

After having studied differential mRNA expression of designated cancer pathways between colorectal adenomas and carcinomas [35], we set out to identify in an unbiased approach which genes show altered expression during progression from adenoma to carcinoma, using the same data set of 37 colorectal adenomas and 31 colorectal carcinomas analysed by oligonucleotide microarrays.

Gene Set Enrichment Analysis (GSEA) was used to identify altered expression of sets of genes associated with specific biological processes in order to identify the molecular mechanisms that drive colorectal adenoma to carcinoma progression.

2 Material and methods

2.1 Samples data set

Expression microarray data available from 68 snap-frozen colorectal tumour samples (37 adenomas and 31 carcinomas) prospectively collected at the VU University medical center (VUmc), Amsterdam, the Netherlands [5] were analysed in the present study. The study was carried out in accordance with the ethical guidelines of our institution concerning informed consent about the use of patient’s material after endoscopic or surgical procedures. The 68 frozen specimens corresponded to 31 females and 34 males (3 patients had multiple tumours). The mean age was 69 (range 47–89). Supplementary Table 1 shows all relevant clinical information about the samples used. Evaluation of tumour content was done by a pathologist (G.A.M.) on 4 μm haematoxylin and eosin-stained cryo sections obtained before and after (i.e. sandwich method) the tissue part from which RNA was isolated. Only cases containing at least 70% of tumour cells in both 4 μm sections were considered for further analysis [5]. All expression microarray data are available at Gene Expression Omnibus (GEO) http://www.ncbi.nlm.nih.gov/geo/ [36], accession number GSE8067.

2.2 Microarray data analysis

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

Supervised analysis for comparing carcinomas to adenomas was done using the Wilcoxon signed rank test. Genes were considered to be differentially expressed in a genome-wide setting when the p-value was less than 1e-5 (False discovery rate; FDR<0.1).
2.3 Prediction Analysis for Microarrays (PAM)

PAM [37] was used to classify our samples based on the gene expression signatures. Leave One Out Cross Validation (LOOCV) was used to evaluate the classification accuracy of the entire data set. The cross validation process was run 100 times. This strategy was also applied on an independent microarray expression data set (GEO database accession number GSE20916) [34] containing 45 adenomas and 36 carcinomas.

2.4 Gene Set Enrichment Analysis (GSEA)

Association of the differential expression with biological processes was performed using Gene Set Enrichment Analysis (GSEA v2.0, http://www.broad.mit.edu/gsea/) [38]. GSEA makes use of the fact that changes in biological characteristics require coordinate variation in expression of groups of genes, i.e. gene-sets, which regulate biological activity. While absolute changes in expression levels of the majority of individual genes are often modest to small and do not reach statistical significance, GSEA allows to estimate group-wise variation in expression of pre-defined gene-sets, indicative for pathway activity levels. In brief, GSEA performs a competitive analysis of pre-defined gene-sets, whereby all genes analysed by expression arrays are ranked according to their differential expression between two sample groups (i.e. colorectal adenomas and carcinomas). GSEA calculates a pathway Enrichment Score (ES) that estimates whether genes from pre-defined gene-sets are enriched among the highest- (or lowest-) ranked genes or whether they are distributed randomly. Default settings were used. Thresholds for significance were determined by permutation analysis (1000 permutations). False Discovery Rate (FDR), i.e. correction for multiple testing, q-values < 0.1 were considered significant.

For GSEA analysis, a total of 3,732 pre-defined gene-sets were available. We analysed the data using three different types of gene-sets, namely ontology gene-sets (n=1454; of which 1,057 were selected using the default setting - microarray results for at least 15 genes per gene-set), curated gene-sets (n=1892; of which 1,434 were selected using the default setting) and gene-sets organized by chromosomal position (n=386; of which 257 were selected using the default setting).

2.5 Identification of key genes for pathway activity

Individual genes from cancer-related gene-sets for which mRNA expression levels differed most between colorectal adenomas and CRCs, were considered key genes for pathway activity. For individual genes within cancer related gene-sets that contributed to the enrichment score, Wilcoxon-ranking test p-values < 1e-5, which in our setting implied FDR < 0.1, of the differential expression at the single gene level were considered significant.

### Table 1  Top 20 significantly upregulated and downregulated genes in colorectal carcinomas in comparison with adenomas

| Gene Symbol | Refseq    | P-value          | Wilcoxon test |
|-------------|-----------|------------------|---------------|
| Upregulated |           |                  |               |
| SPON2       | NM_012445 | 2.46E-10         |               |
| RGS16       | NM_002928 | 8.02E-10         |               |
| LY6E        | NM_002346 | 3.16E-09         |               |
| SFRP4       | NM_003014 | 8.27E-09         |               |
| MFAP2       | NM_017459 | 9.57E-09         |               |
| AURKA       | NM_003600 | 1.28E-08         |               |
| LOC283658   | AL833463  | 2.11E-08         |               |
| TIE1        | NM_005424 | 2.27E-08         |               |
| CTHRC1      | NM_138455 | 6.03E-08         |               |
| BMP7        | AK094784  | 6.46E-08         |               |
| C200QF20    | NM_018270 | 9.72E-08         |               |
| ISG15       | NM_015101 | 1.56E-07         |               |
| SLC27A5     | NM_012254 | 1.56E-07         |               |
| SSSCA1      | NM_006396 | 1.56E-07         |               |
| TROAP       | NM_005480 | 1.66E-07         |               |
| CCL21       | NM_002899 | 1.78E-07         |               |
| TCF15       | NM_006602 | 1.78E-07         |               |
| CENPH       | NM_022909 | 1.90E-07         |               |
| PLSRC3      | NM_020360 | 2.03E-07         |               |
| Downregulated |         |                  |               |
| FAM55D      | NM_152315 | 1.48E-09         |               |
| PRKACB      | NM_002731 | 1.87E-09         |               |
| CLIC3       | NM_004669 | 2.17E-09         |               |
| ENTPD5      | NM_001249 | 2.72E-09         |               |
| ATOH8       | NM_032828 | 3.67E-09         |               |
| SLITRK6     | NM_032229 | 4.59E-09         |               |
| PTGER4      | NM_000958 | 4.59E-09         |               |
| CACNA2D2    | NM_006030 | 5.32E-09         |               |
| TBX3        | NM_016569 | 6.17E-09         |               |
| RETNLB      | NM_032579 | 6.64E-09         |               |
| GLYATL1     | NM_080661 | 6.64E-09         |               |
| –           | AW975332  | 8.90E-09         |               |
| ID4         | NM_001546 | 1.19E-08         |               |
| –           | AK095147  | 1.28E-08         |               |
| UGT1A6      | NM_001072 | 1.28E-08         |               |
| LOC645904   | BM545871  | 1.70E-08         |               |
| VSIG2       | NM_014312 | 2.27E-08         |               |
| –           | BC039491  | 2.43E-08         |               |
| KLHL18      | BC015962  | 2.61E-08         |               |
| –           | AK026378  | 2.80E-08         |               |
3 Results

Microarray expression analysis was performed on snap-frozen samples from 37 adenomas and 31 carcinomas. Expression of 248 genes were significantly different between adenomas and carcinomas (Wilcoxon rank test, p<1e-5), after the analysis of the whole data set (28830 probes). Of these, 96 were significantly upregulated in carcinomas compared to adenomas and 152 were significantly downregulated in carcinomas compared to adenomas. The top 20 of both upregulated and downregulated genes are shown in Table 1. A complete list of the differentially expressed genes, upregulated and downregulated in carcinomas is presented in Supplementary Tables 2 and 3, respectively. The most upregulated genes in carcinomas in comparison to adenomas were Spondin 2 (SPON2) and Regulator of G-Protein signalling 16 (RGS16), located at chromosome arms 4p16.3 and at 1q25.3, respectively. The most downregulated genes in carcinomas in comparison to adenomas were family with sequence similarity 55, member D (FAM55D) and protein kinase, cAMP-dependent, catalytic, beta (PRKACB), located at chromosome arms 11q23.1 and 1p36.1, respectively.

PAM analysis using LOOCV to accurately classify the tumours as adenomas or carcinomas, on the whole gene data set, revealed a set of eight genes that were frequently selected in the 100 runs (Table 2). All eight genes were included in the 248 differentially expressed genes, namely the top upregulated (SPON2) and the top downregulated (FAM55D) genes.

Next, we wanted to check if the expression of the differentially expressed genes (n=248 genes) would be accurate enough to classify adenomas and carcinomas, in an independent adenomas and carcinomas expression dataset. We used the expression dataset of Skrzypczak and collaborators [34], from which we were able to match 169 genes out of 248. From this analysis, we observed that this set of genes was able to classify adenomas and carcinomas with 71% accuracy (Table 3). The same analysis using only the set of upregulated genes (n=96) was performed, as these are suitable candidates to be used as biomarkers in clinical practice.

Table 2 Frequently selected features (in 100 runs) with Leave One Out Cross Validation (LOOCV) applied to the whole data set

| Accession number | Gene Symbol | Nr. of times selected |
|------------------|-------------|-----------------------|
| AK095147         | FLJ37828    | 100                   |
| BM545871         | LOC645904   | 100                   |
| NM 004669        | CLIC3       | 100                   |
| NM 012445        | SPON2       | 2                     |
| NM 032229        | SLITRK6     | 100                   |
| NM 032579        | RETNLB      | 100                   |
| NM 080661        | GLYATL1     | 100                   |
| NM 152315        | FAM55D      | 100                   |

Table 3 Average classification percentage (100 runs) of the 169 matched genes (out of 248 differentially expressed genes) on the independent expression data set from Skrzypczak et al. [34] containing 45 adenomas and 36 carcinomas

|                | Predicted* | Total |
|----------------|------------|-------|
|                | Adenoma    | Carcinoma |
| True           | 45         | 0      | 45    |
| Carcinoma      | 23.41      | 12.59  | 36    |

*Overall misclassification 28.90

From this analysis, we observed that this set of upregulated genes, from which we were able to match 71 genes out of 96, was able to classify adenomas and carcinomas with 89% accuracy (Table 4).

To compare mRNA expression between colorectal adenomas and carcinomas at the pathway level, we performed gene set enrichment analysis (GSEA). Data were analysed using three different types of gene-sets; i.e. ontology gene-sets (gene-sets organised by biological process, molecular function and cellular component), curated gene-sets (gene-sets collected from various sources such as online pathway databases, publications in PubMed, and knowledge of domain experts) and gene-sets organised by chromosomal position (gene-sets corresponding to each human chromosome and each cytogenetic band that has at least one gene).

Within these biological processes, pathways or locations, those genes whose expression was also significantly different between adenomas and carcinomas at an individual level were considered as biomarkers or key genes reflecting the changes occurring within these pathways during progression from adenoma to carcinoma.

Of the 1,057 ontology gene-sets analysed, twenty one were significantly differentially expressed between adenomas and carcinomas (FDR < 0.1) (Table 5). These were all upregulated in carcinomas compared to adenomas and most were involved in chromosome binding, chromosome segregation, mitosis or cell cycle. Key genes recurrently observed within these pathways are Structural maintenance of chromosomes 1A (SMC1A), Aurora kinase A (AURKA), TPX2,
### Table 5 Key genes in biological processes associated with colorectal adenoma to carcinoma progression (FDR < 0.10)

| Gene set (FDR value) | Carcinoma vs adenoma | Gene Symbol | P-value (Wilcoxon signed rank test) |
|----------------------|----------------------|-------------|-------------------------------------|
| **Ontology gene-sets** |                      |             |                                     |
| Regulation of mitosis (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| Mitosis (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| Cell cycle phase (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| **POLD1** | 1.22E-06 |
| Mitotic cell cycle (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| **POLD1** | 1.22E-06 |
| Cell cycle process (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| **SSCA1** | 4.12E-06 |
| **M-phase (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| **POLD1** | 1.22E-06 |
| Collagen (FDR=0.03) | Upregulated | **LUM** | 1.22E-06                           |
| M-phase of mitotic cell cycle (FDR=0.03) | Upregulated | **AURKA** | 1.28E-08                           |
| PLK1 | 3.67E-07 |
| SSSCA1 | 1.56E-07 |
| **TPX2** | 5.53E-06 |
| SMC1A | 2.00E-06 |
| **SPO11** | 1.77E-06 |
| Chromosomal part (FDR=0.06) | Upregulated | **RFC3** | 4.12E-06                           |
| SMC1A | 2.00E-06 |
| Chromatin binding (FDR=0.06) | Upregulated | **POLD1** | 1.22E-06                           |
| **SUV39H1** | 4.45E-07 |
| SMC1A | 2.00E-06 |
| Chromosome segregation (FDR=0.07) | Upregulated | **SMC1A** | 2.00E-06                           |

#### Table 5 (continued)

| Gene set (FDR value) | Carcinoma vs adenoma | Gene Symbol | P-value (Wilcoxon signed rank test) |
|----------------------|----------------------|-------------|-------------------------------------|
| Chromosome pericentric region (FDR=0.07) | Upregulated | **SMC1A** | 2.00E-06                           |
| Chromosome cycle checkpoint (FDR=0.08) | Upregulated | **SMC1A** | 2.00E-06                           |
| Protease inhibitor activity (FDR=0.08) | Upregulated | **AURKA** | 1.28E-08                           |
| Spindle checkpoint (FDR=0.08) | Upregulated | **TPX2** | 5.53E-06                           |
| Chromatin organization and biogenesis (FDR=0.08) | Upregulated | **SMC1A** | 2.00E-06                           |
| Microtubule organization and biogenesis (FDR=0.08) | Upregulated | **SMC1A** | 2.00E-06                           |
| Collagen metabolism (FDR=0.08) | Upregulated | **ADH1C** | 4.12E-06                           |
| Progeria (FDR=0.02) | Upregulated | **PTGS2** | 1.67E-06                           |
| Old age (FDR=0.09) | Upregulated | **PLK1** | 3.67E-07                           |
| Fatty acid metabolism (FDR=0.09) | Down regulated | **ADH1C** | 4.12E-06                           |

#### Curated gene-sets

| Gene set (FDR value) | Carcinoma vs adenoma | Gene Symbol | P-value (Wilcoxon signed rank test) |
|----------------------|----------------------|-------------|-------------------------------------|
| Progeria (FDR=0.02) | Upregulated | **PTGS2** | 1.67E-06                           |
| Old age (FDR=0.09) | Upregulated | **PTGS2** | 1.67E-06                           |
| Fatty acid metabolism (FDR=0.09) | Down regulated | **ADH1C** | 4.12E-06                           |

#### Positional gene-sets

| Gene set (FDR value) | Carcinoma vs adenoma | Gene Symbol | P-value (Wilcoxon signed rank test) |
|----------------------|----------------------|-------------|-------------------------------------|
| 20q11 (FDR=0.09) | Upregulated | **C20orf24** | 1.67E-06                           |
| 20q13 (FDR=0.02) | Upregulated | **AURKA** | 1.28E-08                           |
| 4q22 (FDR=0.05) | Downregulated | – | –                                   |

*Within the three differentially expressed processes involved in the adenoma to carcinoma progression (FDR<0.10), genes that contribute to the GSEA pathway enrichment score were ranked according to the difference in mRNA expression between adenomas and CRCs (Wilcoxon test p-values). Genes with a p-value<1e-5 were considered significant at the genome-wide level and selected as key genes within pathways (depicted)*
In the present study we compared genome-wide gene expression in a collection of 37 colorectal adenomas and 31 carcinomas. Previously, this data set has been studied for different types of predefined gene-sets, i.e. ontology gene-sets, curated gene-sets, and gene-sets organized by chromosomal position. In addition, the effect of this altered expression on tumour biology was evaluated using GSEA pathway analysis with three different types of predefined gene-sets, i.e. ontology gene-sets, curated gene-sets, and gene-sets organized by chromosomal position.

4.1 Differentially expressed genes

mRNA expression was significantly different between adenomas and carcinomas \((p<1e-5)\) for 248 genes, of which 96 were upregulated and 152 downregulated in carcinomas compared to adenomas. Of these, some have previously been found to be upregulated in colon cancer, like, Frizzled-related protein 4 \((SFRP4)\) [39], AURKA [40] Bone morphogenetic protein 7 \((BMP7)\) [41] and Centromere protein H \((CENPH)\) [42] or downregulated in cancer like, prostatic acid phosphatase \((P 5 A P)\) [43] and inhibitor of DNA binding 4 \((ID4)\) [44], supporting the validity of the arrays and analyses procedures used in the present study. However, most genes were not been described in previous expression microarray studies comparing colorectal adenomas to carcinomas [31–34]. This discrepancy could be explained by the fact that these studies used different platforms (either Affymetrix arrays or cDNA microarrays) and different strategies for gene selection compared to the present study. Nonetheless, 8 of the 248 differentially expressed genes observed in the present study, including 2 of the top downregulated genes, are also described to be differentially expressed through CRC progression by Skrzypczak and collaborators [34].

The most upregulated genes in carcinomas in comparison to adenomas were SPON2 and RGS16. SPON2 encodes for an adhesion molecule present in the extra cellular matrix. RGS16 encodes for a protein that inhibits signal transduction by increasing the GTPase activity of G protein alpha subunits. The most downregulated genes in carcinomas in comparison to adenomas were FAM55D and PRKACB. Not much is known about FAM55D while PRKACB is involved in the regulation of several cellular processes like proliferation, cell cycle and microtubule dynamics.

Of the top 20 most upregulated genes, 12 have already been found to be overexpressed in epithelial cancers namely, colon, prostate or breast cancer, i.e. SPON2 [45], RGS16 [46], SFRP4 [39], AURKA [5, 40] TIE1 [47], CTHRC1 [48], BMP7 [41], C20orf20 [5, 49], ISG15 [50], CCL21 [51], TCFL5 [5, 52] and CENPH [42]. Of the top 20 most downregulated genes, five have been associated with epithelial cancer before (including colon cancer), i.e. ENTPD5 [53], PTGER4 [43], CACNA2D2 [54], ID4 [44] and UGT1A6 [55]. Downregulation of UGT1A6 as well as PRKACB genes was also observed by others [34]. In addition, ten of the genes differentially expressed between adenomas and carcinomas \((TPX2, \text{TCFL5}, \text{SLC26A10} \text{ and CENPH upregulated and } SPTBN2, \text{SOCS6}, \text{SLITRK6}, \text{HR, ETFDH and RGS16, BMP7, C20orf11, CENPH, SFRP4, and TPX2.}}\)
APOB48R downregulated) have previously been described to be mutated in CRC [7]. This overlap with existing data corroborated the validity of the approach that was used.

Some of these genes have already been quite extensively studied and in the case of AURKA even used as target for anticancer therapy. Very recently, in a census of amplified and overexpressed human cancer genes, AURKA was described as one of the genes for which good evidence of involvement in the development of human cancer was shown [56]. In CRC in particular, AURKA was shown to be amplified [11] and overexpressed [40]. We also have shown that overexpression of AURKA in CRC was correlated with the copy number increase of 20q13.2 genomic region and was implicated in the chromosomal instability-related adenoma to carcinoma progression [5]. AURKA has a role in centrosome duplication as well as proper chromosome alignment and segregation and its abnormal expression leads to centrosome anomalies and chromosomal instability [57]. Moreover, overexpression of AURKA induces malignant transformation [58]. Another well studied gene is BMP7. It has been associated with e.g. lung, prostate, breast, cervical and colorectal cancer [41, 59–62]. BMP7 can act as suppressor of tumour development by e.g. inhibiting cellular motility in lung cancer cells [59] or promoting growth arrest by repressing telomerase activity in cervical cancer cells [62]. In colorectal cancer, BMP7 has been shown to promote invasive potential [41] and to be associated with liver metastasis and poor prognosis of CRC patients [63].

The PAM analysis revealed a set of 8 genes (all included in the 248 differentially expressed genes), which helped the classifier to achieve best performance. Moreover, when looking at the performance of differentially expressed genes in an independent expression data set, we observed that these genes classified adenomas and carcinomas with an accuracy of 71%, which was improved to 89%, if using only the upregulated genes. This result is very promising, given the fact that upregulated genes are suitable to be used as biomarkers in a clinical setting.

4.2 Upregulated pathways

Analysis of the differences in gene expression between adenomas and carcinomas at the level of specific biological processes showed that ageing (defined by the curated gene-sets progeria and old age), ontology based gene-sets associated with chromosomal instability and the chromosomal position gene-sets 20q11 and 20q13 were upregulated in colon adenoma to carcinoma progression. Not surprisingly, several of the differentially expressed genes discussed above emerged as key genes in differentially expressed pathways.

The notion that cancer is associated with ageing is not new, as in both conditions DNA damage plays a crucial role [64]. One of the cellular features associated with ageing is cellular senescence, a process by which cells enter a state of permanent cell cycle arrest [65]. DNA alterations associated with early neoplasia, including mutations in oncogenes like KRAS can promote senescence and may thus contribute to the oncogene-induced senescence (OIS) [66, 67]. Genes that provoke bypass of OIS will drive oncogenic-mediated transformation enhancing tumor progression [68].

Senescence is a potent tumour suppressive mechanism but it can also contribute to malignant progression, as senescent cells secrete pro-inflammatory cytokines which may be harmful for the surrounding microenvironment [65, 69]. Interestingly, in the present study one of the key genes in the activity of both ageing-related gene-sets is PTGS2 (COX-2). PTGS2 is responsible for the biosynthesis of prostanooids involved in inflammation and mitogenesis and it is known to be overexpressed in colorectal cancer [70].

Analysis of the ontology gene-sets revealed that the genes contributing most to the difference between adenomas and carcinomas were involved in chromosome binding, chromosome segregation, mitosis and cell cycle. These functions are frequently disrupted in chromosomal instable tumours [71]. Interestingly, these genes overlap with upregulation of key genes within the gene-sets located on 20q (20q11 and 20q13). This result is consistent with the observation that gain of the long arm of chromosome 20 is the most frequent chromosomal event in colon cancer. Moreover, we have previously reported on an association between 20q gain and progression from colorectal adenoma to carcinoma [4, 5]. Within these locations eight key genes were identified, expression of which contributed most to the difference between adenomas and carcinomas. Among those, four have previously been described by us to have a putative role in progression, namely C20orf24, AURKA, C20orf20 and RBM38 [5]. The other four were BMP7, TPX2, BCA54 and C20orf11.

Another gene ontology set that was overexpressed in carcinomas in comparison to adenomas was the collagen gene set. This emphasizes the role of the stroma in the progression from premalignant adenomas to invasive cancer, in which the tumor stroma response, which involved proliferation of fibroblasts, is a key factor. We previously showed that gain of 8q24 was associated with increased stroma percentage in CRCs with 20q gain, which is a key step in adenoma to carcinoma progression [72].

4.3 Downregulated pathways

Fatty acid metabolism (curated set) and the chromosome position set 4q22 were the only gene-sets that were downregulated in carcinomas when compared to adenomas.

The role of fatty acid metabolism in cancer development is rather complex. On one hand, fatty acid metabolism is involved in cancer development and cell growth for the
necessary synthesis of e.g. cell and organelle membranes [73]. On the other hand, fatty acid metabolism, through beta oxidation in complex interaction with glucose metabolism and hypoxia, plays a role in the energy supply for the cancer cells [74]. These metabolic changes are stimulated or inhibited by several oncogenes or tumour suppressor genes respectively. These genes are known to be involved in proliferative and survival signalling pathways [75].

One of the top genes contributing to the differential expression of the fatty acid metabolism pathway is the alcohol dehydrogenase ADH1C, which is also significantly downregulated in colorectal carcinomas compared to adenomas at the single gene level and therefore considered a key gene in the activity of this pathway. Alcohol dehydrogenase produces carcinogenic acetaldehyde through ethanol oxidation. Polymorphisms in ADH1C have been associated with CRC risk in combination with alcohol consumption [76]. The exact role of ADH1C in adenoma to carcinoma progression, however, is not immediately obvious from these data.

GSEA analysis revealed the chromosome 4q22 gene-set as significantly downregulated in colorectal carcinomas compared to adenomas. However, in the 4q22 gene-set no key gene (significantly downregulated in carcinomas compared to adenomas) was identified and no 4q22 genes are in the top 248 list of differentially expressed genes. This indicates that rather subtle changes at the individual gene level may contribute to the difference observed for the overall set of genes. Deletions of 4q occur in colorectal carcinomas but not in adenomas [5, 77] and no significant gene dosage effects have been identified for genes at this locus, which could mean that other mechanisms like mutations and/or epigenetic regulation are involved. Nonetheless, we have shown that 4q22.1–4q35.2 deletion is associated with a worse outcome in stage II colon cancer patients [78].

4.4 ‘Vogelgram’ genes

In the present study, no significant differences in mRNA expression between adenomas and carcinomas were found for genes traditionally attributed to CRC development like APC, KRAS, TP53 and DCC [6]. It is not surprising that initial events like altered expression of APC or KRAS do not appear in this analysis because expression of these genes is already altered in many adenomas. On the other hand, to some extent this may be due to technical reasons. When base line expression levels of genes are not very high, loss of expression would result in only a relatively minor signal that could remain undetected.

In summary, whole genome mRNA expression patterns of colorectal adenomas and carcinomas were compared and a set of 248 genes were significantly up or downregulated in carcinomas compared to adenomas, suggesting a functional role of these genes in the progression from colorectal adenoma to carcinoma. Gene-sets at chromosomes 4q22 and 20q and biological processes like ageing (which is associated with senescence), chromosomal instability and fatty acid metabolism were differentially expressed in colorectal adenoma to carcinoma progression.

These data are consistent with the notion that adenomas and carcinomas are distinct biological entities. Disruption of specific biological processes like senescence, maintenance of chromosomal instability and altered metabolism are key factors in the progression from adenoma to carcinoma.

Declarations The study was carried out in accordance with the ethical guidelines of our institution concerning informed consent about the use of patient’s material after endoscopic or surgical procedures. The authors declare that they have no conflict of interest.

References

1. J. Ferlay, P. Autier, M. Boniol, M. Heanue, M. Colombet, P. Boyle, Estimates of the cancer incidence and mortality in Europe in 2006. Ann. Oncol., 18, 581–592 (2007)
2. T. Muto, H.J. Bussey, B.C. Morson, The evolution of cancer of the colon and rectum. Cancer 36, 2251–2270 (1975)
3. H. Shinya, W.J. Wolff, Morphology, anatomic distribution and cancer potential of colonic polyps. Ann. Surg., 190, 679–683 (1979)
4. M. Hermsen, C. Postma, J. Baak, M. Weiss, A. Papallo, A. Scuotto, G. Roemen, J.W. Arends, R. Williams, W. Giaretti, A. De Goey, G. Meijer, Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. Gastroenterology 123, 1109–1119 (2002)
5. B. Carvalho, C. Postma, S. Mongera, E. Hopmans, S. Diskin, M.A. van de Wiel, C.W. van, O. Thas, A. Matthai, M.A. Cuesta, J.S. Terhaar Sive Droste, M. Craanen, E. Schrock, B. Ylstra, G.A. Meijer, Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. Gut 58, 79–89 (2009)
6. E.R. Fearon, B. Vogelstein, A genetic model for colorectal tumorigenesis. Cell 61, 759–767 (1990)
7. L.D. Wood, D.W. Parsons, S. Jones, J. Lin, T. Sjoblom, R.J. Leary, D. Shen, S.M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezsö, V. Ustyansky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P. A. Wilson, J.S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J.K. Willson, S. Sukumar, K. Polyak, B.H. Park, C.L. Pethiyagoda, P.V. Pant, D.G. Ballinger, A.B. Sparks, J.S. Hartigan, D.R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S. D. Markowitz, G. Parmigiani, K.W. Kinzler, V.E. Velculescu, B. Vogelstein, The genomic landscapes of human breast and colorectal cancers. Science 318, 1108–1113 (2007)
8. R.J. Leary, J.C. Lin, J. Cummins, S. Boca, L.D. Wood, D.W. Parsons, S. Jones, T. Sjoblom, B.H. Park, R. Parsons, J. Willis, D. Dawson, J.K. Willson, T. Nikolskaya, Y. Nikolsky, L. Kopelovich, N. Papadopoulos, L.A. Pennacchio, T.L. Wang, S.D. Markowitz, G. Parmigiani, K.W. Kinzler, B. Vogelstein, V.E. Velculescu, Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc. Natl. Acad. Sci. U. S. A. 105, 16224–16229 (2008)
18. M. Bianchini, E. Levy, C. Zucchini, V. Pinski, C. Macagno, S.P. Min, N.K. Kim, S.W. Nam, H. Kim, Differential gene expression profiles of metastases in paired primary and metastatic colorectal carcinomas. Oncology 75, 92–101 (2008)

19. J.R. Bischoff, L. Anderson, Y. Zhu, K. Mossie, L. Ng, B. Souza, B. Arango, K. Birkenkamp-Demtroder, F.B. Sorensen, L.L. Christensen, A new target for anticancer drug. Recent Pat. Anticancer Drug Discov. 3, 114–122 (2008)

20. O. Kitahara, Y. Furukawa, T. Tanaka, K. Kihara, K. Ono, R. Yanagawa, M.E. Nita, T. Takagi, Y. Nakamura, T. Tsunoda, Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res. 61, 3544–3549 (2001)

21. T.T. Zou, F.M. Salaru, Y. Xu, V. Shustova, J. Yin, Y. Mori, D. Shibata, F. Sato, S. Wang, A. Olaru, E. Deacu, T.C. Liu, J.M. Abraham, S.J. Meltzer, Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 21, 4855–4862 (2002)

22. K. Birkenkamp-Demtroder, L.L. Christensen, S.H. Olesen, C.M. Frederiksen, P. Laiho, L.A. Aaltenon, S. Lauberg, F.B. Sorensen, R. Hagemann, T.F. ORntoft, Gene expression in colorectal cancer. Cancer Res. 62, 4352–4363 (2002)

23. G.A. Ganepola, R.M. Mazzotti, D. Weeresinghe, G.A. Corner, C. J. Parish, D.H. Chang, N.C. Tebbutt, C. Murone, N. Ahmed, L.H. Augenlicht, J.M. Mariadason, Gene expression profiling of primary and metastatic colon cancers identifies a reduced proliferative rate in metastatic tumors. Clin. Exp. Metastasis. 27, 1–9 (2010)

24. K.H. Koh, H. Rhee, H.J. Kang, E. Yang, K.T. You, H. Lee, B.S. Min, N.K. Kim, S.W. Nam, H. Kim, Differential gene expression profiles of nonneoplastic mucosa may predict clinical outcome of colon cancer patients. Dis. Colon Rectum 48, 2238–2248 (2005)

25. M. Grade, P. Hornmann, S. Becker, A.B. Hummon, D. Wangsa, S. Varma, R. Simon, T. Liersch, H. Becker, M.J. Dillipiantonio, B. M. Ghadimi, T. Ried, Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. Cancer Res. 67, 41–56 (2007)

26. A. Barrier, P.Y. Boelle, A. Lemoine, C. Tse, D. Brault, F. Chiappini, F. Lacaine, S. Houry, M. Huguier, A. Flahault, S. Dudoit. Gene expression profiling of colorectal adenocarcinoma: correlations with clinicopathological features, p16 expression, and telomerase activity. Hum. Pathol. 39, 599–604 (2008)

27. D. Arango, P. Laiho, A. Kokko, P. Alhopuro, H. Sammalkorpi, R. Salovaara, D. Nicorici, S. Hautaniemi, H. Alazzouzi, J.P. Mecklin, H. Jarvinen, A. Hemminki, J. Astola, Schwartz S J, L.A. Aaltenon, Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. Gastroenterology 129, 874–884 (2005)

28. E. Bandrés, R. Malumbres, E. Cubedo, B. Honorato, R. Zarate, A. Labarga, U. Gabisu, J.J. Sola, J. Garcia-Foncillas, A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes’ B colon cancer patients. Oncol. Rep. 17, 1089–1094 (2007)

29. A. Barrier, P.Y. Boelle, F. Roser, J. Gregg, C. Tse, D. Brault, F. Lacaine, S. Houry, M. Huguier, B. Frane, A. Flahault, A. Lemoine, S. Dudoit. Stage II colon cancer prognosis prediction by tumor gene expression profiling. J. Clin. Oncol. 24, 4685–4691 (2006)

30. M. Del Rio, F. Molina, C. Bascul-Mollevi, V. Copois, F. Bibeau, P. Chalbos, C. Bareil, A. Kramar, N. Salvetat, C. Fraslon, E. Conseiller, V. Granci, B. Leblanc, B. Pau, P. Martineau, M. Ychou. Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. J. Clin. Oncol. 25, 773–780 (2007)

31. D.A. Noterman, U. Alon, A.J. Sierk, A.J. Levine, Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res. 65, 3124–3130 (2001)

32. Y.M. Lin, Y. Furukawa, T. Tsunoda, C.T. Yue, K.C. Yang, Y. Nakamura. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 21, 4120–4128 (2002)

33. O. Galamb, B. Gyorffy, F. Sipos, S. Spisak, A.M. Nemeth, P. Miheller, Z. Tulassay, E. Dinya, B. Molnar, Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. Dis. Markers. 25, 1–16 (2008)

34. M. Skrzypczak, K. Goryca, T. Rubel, A. Paziewska, M. Mikula, D. Jarosz, J. Pachlewski, J. Olezdk, J. Ostrowski, Modeling onco- genic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. PLoS. One. 5 (2010)

35. A.H. Sillars-Hardebol, B. Carvalho, M. de Wit, C. Postma, P.M. Delis-van Diemen, S. Mongera, B. Ylstra, M.A. van de Wiel, G.A. Meijer, R.I. Fijneman, Identification of key genes for carcinogenic pathways associated with colorectal adenoma-to-carcinoma progression. Tumour. Biol. 31, 89–96 (2010)

36. R. Edgar, M. Domrachev, A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210 (2002)
genes in head and neck squamous cell carcinoma: Correlation with progression and prognosis. Int. J. Cancer 123, 2594–2604 (2008)
55. R.A. Hubner, K.R. Muir, J.F. Liu, R.F. Logan, M. Grainge, N. Armitage, V. Shepherd, S. Popat, R.S. Houlston, Genetic variants of UGT1A6 influence risk of colorectal adenoma recurrence. Clin. Cancer Res. 12, 6585–6589 (2006)
56. T. Santarius, J. Shipley, D. Brewer, M.R. Stratton, C.S. Cooper, A census of amplified and overexpressed human cancer genes. Nat. Rev. Cancer 10, 59–64 (2010)
57. S. Sen, H. Katayama, K. Sasai, Functional significance of Aurora kinase A in centrosome amplification and genomic instability. Adv. Exp. Med. Biol. 617, 99–108 (2008)
58. H. Zhou, J. Kuang, L. Zhong, W.L. Kuo, J.W. Gray, A. Sahin, B.R. Brinkley, S. Sen, Tumour amplified kinase STK15/BTK induces centrosome amplification, aneuploidy and transformation. Nat. Genet. 20, 189–193 (1998)
59. J. Chen, L. Ye, F. Xie, Y. Yang, L. Zhang, W.G. Jiang, Expression of bone morphogenetic protein 7 in lung cancer and its biological impact on lung cancer cells. Anticancer. Res. 30, 1113–1120 (2010)
60. C. Morrissey, L.G. Brown, T.E. Pitts, R.L. Vessella, E. Corey, Bone morphogenetic protein 7 is expressed in prostate cancer metastases and its effects on prostate tumor cells depend on cell phenotype and the tumor microenvironment. Neoplasia 12, 192–205 (2010)
61. E.L. Alarjo, M. Pardissin, J.M. Kotelainen, K. Savinainen, R. Karhu, A. Kallioniemi, BMP7 influences proliferation, migration, and invasion of breast cancer cells. Cancer Lett. 275, 35–43 (2009)
62. L. Casar, H. Li, A.R. Pinto, C. Nicholls, S. Bayne, J.P. Liu, Bone morphogenetic protein-7 inhibits telomerocyte activity, telomere maintenance, and cervical tumor growth. Cancer Res. 68, 9157–9166 (2008)
63. K. Motoyama, F. Tanaka, Y. Kosaka, K. Mimori, H. Uetake, H. Inoue, K. Sugihara, M. Mori, Clinical significance of BMP7 in human colorectal cancer. Ann. Surg. Oncol. 15, 1530–1537 (2008)
64. J.H. Hoeijmakers, DNA damage, aging, and cancer. N. Engl. J. Med. 361, 1475–1485 (2009)
65. T. Kuilman, C. Michaloglou, W.J. Mooi, D.S. Peeker, The essence of senescence. Genes Dev. 24, 2463–2479 (2010)
66. W.J. Mooi, D.S. Peeker, Oncogene-induced cell senescence—halting on the road to cancer. N. Engl. J. Med. 355, 1037–1046 (2006)
67. H.T. Hemann, M. Narita, Oncogenes and senescence: breaking down in the fast lane. Genes Dev. 21, 1–5 (2007)
68. L.C. Vredeveld, B.D. Rowland, S. Douma, R. Bernards, D.S. Peeker, Functional identification of LRF as an oncogene that bypasses RASV12-induced senescence via upregulation of CYCLIN E. Carcinogenesis 31, 201–207 (2010)
69. A.R. Davalos, J.P. Coppe, J. Campisi, P.Y. Desprez, Senescent cells as a source of inflammatory factors for tumor progression. Cancer Metastasis Rev. 29, 273–283 (2010)
70. D. Wang, R.N. DuBois, Pro-inflammatory prostaglandins and progression of colorectal cancer. Cancer Lett. 267, 197–203 (2008)
71. H. Rajagopalan, M.A. Nowak, B. Vogelstein, C. Lengauer, The significance of unstable chromosomes in colorectal cancer. Nat. Rev. Cancer 3, 695–701 (2003)
72. R.J. Fijneman, B. Carvalho, C. Postma, S. Mongera, V.W. van Hinsbergh, G.A. Meijer, Loss of 1p36, gain of 8q24, and loss of 9q34 are associated with stroma percentage of colorectal cancer. Cancer Lett. 258, 223–229 (2007)
73. R.J. DeBerardinis, N. Sayed, D. Ditsworth, C.B. Thompson, Brick by brick: metabolism and tumor cell growth. Curr. Opin. Genet. Dev. 18, 54–61 (2008)
74. I. Samudio, R. Harmancey, M. Fiegel, H. Kantarjian, M. Konopleva, B. Korchin, K. Kaluvarachhi, W. Bornmann, S. Duvvuri, H. Taegtmeyer, M. Andreeff, Pharmacologic inhibition of fatty
acids oxidation sensitizes human leukemia cells to apoptosis induction. J. Clin. Invest. 120, 142–156 (2010)
75. V. Fritz, L. Fajas, Metabolism and proliferation share common regulatory pathways in cancer cells. Oncogene 29, 4369–4377 (2010)
76. N. Homann, I.R. Konig, M. Marks, M. Benesova, F. Stickel, G. Millonig, S. Mueller, H.K. Seitz, Alcohol and colorectal cancer: the role of alcohol dehydrogenase 1 C polymorphism. Alcohol. Clin. Exp. Res. 33, 551–556 (2009)
77. E.J. Douglas, H. Fiegler, A. Rowan, S. Halford, D.C. Bicknell, W. Bodmer, I.P. Tomlinson, N.P. Carter, Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. Cancer Res. 64, 4817–4825 (2004)
78. R.P. Brosens, E.J. Belt, J.C. Haan, T.E. Buffart, B. Carvalho, H. Grabsch, P. Quirke, M.A. Cuesta, A.F. Engel, B. Ylstra, G.A. Meijer, Deletion of chromosome 4q predicts outcome in stage II colon cancer patients. Cell. Oncol. 34, 215–223 (2011)