

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

An integrative framework identifies alternative splicing events in colorectal cancer development

Andrea Bisognin^{a,1}, Silvia Pizzini^{b,1}, Lisa Perilli^b, Giovanni Esposito^c,
Simone Mocellin^d, Donato Nitti^d, Paola Zanovello^{b,c}, Stefania Bortoluzzi^{a,*},
Susanna Mandruzzato^{b,c}

^aDepartment of Biology, University of Padova, Padova, Italy

^bOncology and Immunology Section, Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy

^cIstituto Oncologico Veneto (IOV), IRCCS, Padova, Italy

^dSurgery Section, Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy

ARTICLE INFO

Article history:

Received 25 June 2013

Received in revised form

17 September 2013

Accepted 8 October 2013

Available online 19 October 2013

Keywords:

Colorectal cancer

Alternative splicing

Exon arrays

ABSTRACT

Alternative splicing (AS) is a common mechanism which creates diverse RNA isoforms from a single gene, potentially increasing protein variety. Growing evidence suggests that this mechanism is closely related to cancer progression. In this study, whole transcriptome analysis was performed with GeneChip Human exon 1.0 ST Array from 80 samples comprising 23 normal colon mucosa, 30 primary colorectal cancer and 27 liver metastatic specimens from 46 patients, to identify AS events in colorectal cancer progression. Differentially expressed genes and exons were estimated and AS events were reconstructed by combining exon-level analyses with AltAnalyze algorithms and transcript-level estimations (MMBGX probabilistic method). The number of AS genes in the transition from normal colon mucosa to primary tumor was the most abundant, but fell considerably in the next transition to liver metastasis. 206 genes with probable AS events in colon cancer development and progression were identified, that are involved in processes and pathways relevant to tumor biology, as cell–cell and cell–matrix interactions. Several AS events in VCL, CALD1, B3GNT6 and CTHRC1 genes, differentially expressed during tumor development were validated, at RNA and at protein level. Taken together, these results demonstrate that cancer-specific AS is common in early phases of colorectal cancer natural history.

© 2013 Federation of European Biochemical Societies.

Published by Elsevier B.V. All rights reserved.

1. Introduction

Colorectal cancer (CRC) development is a model of cancerogenesis with a complex multistep process, involving

accumulation of a significant number of genetic alterations of genes regulating key cellular processes (Fearon and Vogelstein, 1990; Sheffer et al., 2009). About 20–25% of colorectal cancer patients present with distant metastatic disease

Abbreviations: ncRNAs, non-coding RNAs; N, normal colon mucosa; T, primary colorectal cancer; M, liver metastasis; ASEs, alternatively spliced exons.

* Corresponding author. Tel.: +39 (0)49 8276502; fax: +39 (0)49 8276209.

E-mail address: stefania.bortoluzzi@unipd.it (S. Bortoluzzi).

¹ These authors contributed equally to this study.

1574-7891/\$ – see front matter © 2013 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

<http://dx.doi.org/10.1016/j.molonc.2013.10.004>

(mainly in the liver) at diagnosis, and unresectable liver metastasis is associated with short survival (Gleisner et al., 2008; Jemal et al., 2005; Ogino and Goel, 2008).

Several studies have evaluated gene expression and genomic profiling of CRC (Cardoso et al., 2007; Habermann et al., 2008), focusing mainly on differential gene expression for disease phenotype classification (e.g. neoplastic vs normal tissue), but also on gene expression variability according to anatomical regions in normal colon (Birkenkamp-Demtroder et al., 2005; LaPointe et al., 2008). A recent comprehensive study involved exome sequence, DNA copy number, promoter methylation, mRNA and microRNA expression. As well as identifying recurrent mutated genes and methylation patterns, this study identified recurrent alterations in several pathways (e.g., WNT, PI3K, TGF- β , MAPK, p53) and found molecular signatures associated with tumor aggressiveness (Cancer Genome Atlas (2012)), thus reinforcing the concept that multiple genetic events are required to unleash the malignant progression of CRC and also that only a genome-wide approach can interpret complex scenarios in the biology of tumors.

Ample evidence now shows that most human genes undergo alternative splicing (AS) to express many transcript isoforms. The resulting protein isoforms play distinct roles, which contribute to increase the functional diversity of cells while maintaining a limited number of genes encoded by the genome.

Although splicing events do not always have functional consequences, it is clear that this process has great potential to produce a significant biological effect in several cell processes. For example, there are now several studies demonstrating that splicing alterations occur in many cancers (Pal et al., 2012). A quantitative estimate of splicing disruption in cancer has also been attempted, and the expression of normal splice variants was found to be widely and significantly disrupted in at least half the cancers studied (Ritchie et al., 2008). Until now, a relatively small number of studies have addressed the role of AS in tumors from breast, brain, lung and colorectal cancer, from which it appears that splicing alterations are quite common in cancer (Germann et al., 2012). In this regard, alternatively spliced proteins are particularly important in oncology, since they may contribute to the etiology of cancer, be involved in the metastatic process (Gutschner et al., 2013), serve as prognostic biomarkers (Brinkman, 2004) and provide tumor selective drug targets.

In addition, transcriptome complexity has gradually become appreciated in the last few years. Several classes of non-coding RNAs (ncRNAs) control expression at multiple levels, acting as epigenetic (Taft et al., 2010), transcriptional (Zardo et al., 2012) and post-transcriptional regulators (Bisognin et al., 2012; Lionetti et al., 2009) in normal development, physiology and, when dysfunctional, disease conditions.

Control of RNA processing is currently recognized as an essential component of gene expression regulation. For instance, alternative cleavage and polyadenylation play important roles in CRC development (Morris et al., 2012). RNA-based processes are definitely involved, either as causative entities, modulating influences, or as compensatory responses to disease (Ward and Cooper, 2010).

More than 95% of human genes encode splice isoforms, some of which exert antagonistic functions (Miura et al., 2012; Pan et al., 2008; Wang et al., 2008). Alternative splicing (AS) increases the diversity of both ncRNAs and coding transcripts, reflecting protein isoforms and directly influencing protein–protein interaction networks (Ellis et al., 2012). AS is accurately controlled, both spatially and temporally, by the interplay of cis-acting signals of trans-acting elements (Kornblihtt et al., 2013; Miura et al., 2011). The latter may comprise splicing machinery components as well as trans-acting ncRNAs with regulatory roles.

Over 50% of disease-causing mutations affect splicing (Tazi et al., 2009). Several splice variants are commonly found to be enriched in cancer tissue compared with normal surrounding tissue. Splicing changes may result from mutations within intronic or exonic splicing elements in cancer genes. However, aberrant splicing often involves transcripts from non-mutated genes, indicating defects in splicing effectors or regulators (Ward and Cooper, 2010). It should be noted that the metastasis-associated ncRNA MALAT-1 modulates AS, controlling the ratios among various isoforms through its interaction with the serine/arginine-rich (SR) family of nuclear phosphoproteins of the splicing machinery (Tripathi et al., 2010). Other advances in understanding molecular mechanisms and pathways modulating the AS of transcripts encoding key cancer proteins (caspase 9 AS regulation by phosphoinositide 3-kinase/Akt pathway (Goehle et al., 2010); SLC39A14 CRC regulation by the Wnt pathway (Thorsen et al., 2011)) have emphasized the importance of AS deregulation in all aspects of cancer aetiology. Aberrant splicing in cancer may increase the oncogenic potential of protein variants and promote cancer progression (Germann et al., 2012). Specific splicing events products may be signatures identifying cancer subtypes, predicting clinical outcomes or indicating treatment choices. It was the knowledge that the AS of Bcl-2 generates both pro- and anti-apoptotic proteins, whose combination regulates the apoptosis machinery critical for cell fate (Akgul et al., 2004), which paved the way for the development of novel anticancer drugs affecting the AS of Bcl-x and other human apoptotic genes (Shkreta et al., 2008). Modulation of splicing through small and anti-sense RNA molecules is also a powerful approach against disease causes and effects.

In this view, identification of cancer-associated splicing events and differential isoform expression between cancerous and normal tissues is a key issue in ongoing cancer research.

A few studies have investigated AS in colon cancer by means of exon arrays. Early in this field, Gardina et al. examined 20 paired tumor-normal colon cancer samples from 10 patients, predicting and partially validating specific splicing events mostly affecting cytoskeletal, extracellular and cell–cell interaction proteins (Gardina et al., 2006). Thorsen et al. studied tissue- and tumor-specific AS in various types of tumors, and validated six out of 23 detected colon cancer-specific AS events, only two of which were not reported by Gardina et al. (Thorsen et al., 2008). Mojica and Hawthorn reported exon array-based data from 13 non-neoplastic colonic epithelial cells from 10 patients, and directly compared them with previously mentioned matched samples, showing the complexity of AS and highlighting the limitations of current transcript annotations (Mojica and Hawthorn, 2010). More

Table 1 – Patient and tumor characteristics.

No. of patients	46
Age (years, mean \pm s.d.)	60.7 \pm 10.2
Gender	
Female	17 (37%)
Male	29 (63%)
Tumor site	
Cecum, ascending colon, transverse colon	13 (28%)
Splenic [left] flexure, descending colon, sigmoid colon	20 (44%)
Rectum	13 (28%)
TNM stage	IV
Liver metastasis	
Synchronous	39 (85%)
Metachronous	7 (15%)

recently, Thorsen et al. showed that the AS of the divalent cation transporter SLC39A14 in CRC is regulated by the Wnt pathway, through regulation of the SRSF1 splicing factor and its regulatory kinase SRPK1 (Thorsen et al., 2011).

Overall, previous exon array-based studies have demonstrated the scarce specificity of available AS detection methods. Indeed, identification of cancer AS events with RNA-seq data is still an open issue. In the present study, we used exon chips to obtain exon expression profiles in a large set of biopsies from normal colon mucosa, primary CRC and liver metastases, partially matched by individual patient. We implemented an integrative framework, combining detection of exons involved in cancer-related AS events, and identification of transcript isoforms differentially expressed during CRC development and progression, to identify a restricted set of AS events, some of which were experimentally validated.

2. Material and methods

2.1. Patient samples and RNA extraction

For this study, 46 patients with colorectal adenocarcinoma, who underwent surgery at the Department of Surgery, Oncology and Gastroenterology, University of Padova, between March 1994 and September 2008, were retrospectively selected from the institutional prospectively maintained colorectal database. Patients with known history of hereditary CRC syndrome were excluded. The Ethics Committee of the University Hospital of Padova approved the study, and all patients provided their written informed consent. Enrolled patients did not receive neo-adjuvant treatment. We selected samples of normal mucosa (N), primary colorectal cancer (T) and liver metastases (M). Table 1 lists the main patient and tumor characteristics. Normal mucosa samples were taken at a minimum distance of 10 cm from the tumor site. All samples, after excision, were immediately snap-frozen in liquid nitrogen and stored at -80° until use.

From each tissue sample, 7 μ m sections were prepared in a Leica CM 1950 cryostat (Leica Microsystems, Wetzlar, Germany). Haematoxylin- and eosin-stained sections of each specimen were evaluated by an experienced pathologist, and only samples with more than 80% of tumor tissue were

considered for RNA extraction. Total RNA was extracted from cryostat sections with 1 mL Trizol reagent (Invitrogen Life Technology Inc., Carlsbad, CA, USA) according to the manufacturer's instructions. RNA concentration was quantified on a NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Waltham, MA, USA). RNA quality was evaluated by RNA 6000 Nano LabChip (Agilent Technologies, Santa Clara, CA, USA) on an Agilent 2100 Bioanalyzer. Samples with RNA integrity number <6 were excluded. Total extracted RNA was purified according to the 'RNA Cleanup' protocol, with column DNase digestion to remove residual genomic DNA and to deliver high-quality total RNA (Qiagen, Crawley, West Sussex, UK).

2.2. Microarray hybridization and raw data processing

GeneChip Human Exon 1.0 ST (Affymetrix) was used to analyze both gene and exon expression in T, N and M samples. RNA isolated from tissue samples was labeled and hybridized according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA). Briefly, 100 ng of total RNA from each sample was labeled with the Ambion WT expression kit (Ambion Inc, Austin, TX, USA) as provided by the manufacturer. End-labeling, hybridization, washing and scanning were performed according to the user manual of the GeneChip Whole Transcript (WT) Sense Target Labeling Assay (Affymetrix), and using an Affymetrix GCS 3000 7G scanner.

Initial quality control was performed with Affymetrix[®] Expression Console[™] software (v.1.0) to determine the success of hybridizations. Probe-level signals were summarized in parallel, to estimate both exon and gene expression levels in the samples. Raw data were processed by RMAExpress, a GUI program to compute gene expression summary values for Affymetrix GeneChip[®] data with the Robust Multichip Average expression summary and to carry out quality assessment with probe-level metrics. Quality control of samples was based on two main PLM-based quality statistics, Normalized Unscaled Standard Error (NUSE) and Relative Log Expression (RLE), to assess the overall quality of signals in each array, according to the distribution of standard errors and relative logs of expression of single probesets. MA plots before and after RMA were evaluated to identify biases associated with specific intensity classes. The IQR limits option was used to visualize control limits ($1.5 \times \text{IQR}$ above the upper quartile and $1.5 \times \text{IQR}$ below the lower quartile), derived from normal boxplot outlier identification rules. Genes and exons were ranked according to expression profile variability across samples with the CV (coefficient of variation). For descriptive analysis of samples, the agglomerative clustering procedure was used, with Pearson's correlation distance and complete linkage clustering.

2.3. Data analysis for AS identification

AltAnalyze version 1.0 Beta (Emig et al., 2010) was applied to exon-level analysis. A 0.05 maximum detection above background p -value and a probeset intensity greater than 70 were required to preselect probesets, which were then used to identify AS exons (ASEs; p -value 0.05), i.e., those exons with expression variability of which in a sample contrast is not attributable to whole gene expression variability. Both

MiDAS (Microarray Detection of Alternative Splicing) and FIRMA (Finding Isoforms using Robust Multichip Analysis) algorithms were applied to identify ASEs in three contrasted conditions/groups of samples. For each contrast, two post-analysis filters were applied: a fold change of at least 2 for alternative spliced exons, and a fold change of at most 3 for genes.

MMBGX (Multi-Mapping Bayesian Gene eXpression) (Turro et al., 2010) version 0.99.9, was used for probabilistic estimation of the expression level of each known Ensembl transcript and gene interrogated by the set of probes considered, from the GeneChip Human Exon platform. The p_t statistic (the posterior probability that a transcript is more up-regulated than its corresponding gene in condition 1 relative to condition 2) was calculated for each transcript and each sample group contrast. In this way, transcripts with p_t values of <0.05 or >0.95 were considered to be differentially expressed (differentially expressed transcripts, DETs). For each contrast, genes with at least one DET were considered putatively differentially spliced in that comparison (differentially spliced genes, DSGs).

2.3.1. Functional enrichment

GO terms and KEGG pathway functional enrichment was carried out with DAVID. Only enrichment results with FDR $<1\%$ were judged to be significant.

2.3.2. Selection of candidates of AS events for experimental validation

Results from ASEs were then complemented with information from the xmapcore R package, version 1.2.8, which remaps Affymetrix probes to Ensembl v.56. A UCSC genome custom track was built for each candidate AS event, to facilitate visual inspection of probes in the genomic context. Probes were color-coded according to the specificity of alignment, to give priority to AS occurring in sets of probes which do not cross-hybridize to other genomic positions.

In addition to the p_t statistics implemented in the MMBGX, differential expression of transcripts was also assessed by evaluating the posterior probability distribution plots of gene and transcript expression levels. Priority was given to transcripts for which expression estimate plots show an opposite trend compared with that of the gene.

2.4. Reverse transcription of mRNA, quantitative PCR and data analysis

One μg of total RNA was used for first-strand cDNA synthesis using the SuperScript™ II Reverse Transcriptase kit (Invitrogen by Life Technologies Inc., USA) in a total volume of 20 μl according to the manufacturer's instructions. Reverse transcription was performed as follows: a volume of 12 μl containing oligo (dT)_{12–18} primers (Invitrogen), dNTP mixture and RNA was incubated at 65 °C for 5 min, then were added 5X First-Strand Buffer and 0.1 M DTT up to 19 μl and incubated at 42 °C for 2 min; finally Superscript II RT (200 U/ μl) was added to a final volume of 20 μl . Final reaction mixture was incubated at 42 °C for 50 min and terminated at 70 °C for 15 min. The cDNA was stored at –20 °C until further use.

In order to validate transcript array expression data, RT-qPCR was carried out by LightCycler 480 II using LightCycler

480 multiwell white plates (Roche Applied Sciences, Indianapolis, USA) and SYBR Green I Master (Roche), with specifically designed primers (Table S3A in Supplementary File 3) or Probe Masters with RealTime Ready Single Assay (Roche) (Table S3B in Supplementary File 3). Primers were designed using Primer3Plus software. Primer-Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) was used to check primer specificity for transcripts and genomic targets. Primers were designed to detect and distinguish at least among the transcripts that are differentially expressed in our integrative analysis, unless otherwise specified. Amplicons were tested for potential secondary structure using OligoAnalyzer 3.1 (Integrated DNA Technologies, Belgium). PCR products were 423 bp for transcript isoforms –001 (ENST00000361901), 132 bp for the isoform –005 (ENST00000361675) and 235 bp for the –012 isoform (ENST00000436461) all belonging to CALD1; for VCL-001 (ENST00000372755) and for VCL-201 (ENST00000211998), PCR products were 197 and 274 bp, respectively.

Optimal reaction conditions for RT-qPCR with SYBR Green I Master were obtained using 0.5 μM forward primer, 0.5 μM reverse primer, RNase/DNase-free water and cDNA template up to a final volume of 20 μl . Amplifications were performed starting with a 5 min enzyme activation at 95 °C, followed by 45 cycles of denaturation at 95 °C for 10 s, annealing at 60 °C for 15 s and extension at 72 °C for 15 s. At the end of each run a melting curve analysis was performed from 65 to 95 °C.

Instead, in RT-qPCR performed with RealTime Ready Single Assay each custom assay was already distributed into the plates and includes transcript specific primers and a Universal Probelibrary (UPL) probe, which is a short FAM-labeled hydrolysis probe containing locked nucleic acid (LNA). Then Probe Master and cDNA template were added up to a final volume of 20 μl . Amplifications were performed starting with a 10 min enzyme activation at 95 °C, followed by 45 cycles of denaturation at 95 °C for 10 s, annealing at 60 °C for 30 s and extension at 72 °C for 1 s.

Cycle of quantification (C_q) values over 36 were excluded from further mathematical calculations. A sample without cDNA was used as negative control.

Experiments were performed on at least 70 samples included in the final exon array dataset of 80 samples. Each sample was measured independently three times in triplicate to assess the repeatability and reproducibility of results and the data were analysed according to the $\Delta\Delta C_q$ method against the internal reference gene DACT1, selected as a trustworthy expressed control with minimum variability (measured as expression profile Shannon entropy) using LightCycler 480 Software, Version 1.5 (Roche). Data were expressed as mean values \pm SE. Statistical significance was assessed with t-test.

2.5. Western blot analysis

Total proteins were extracted from 3 triplets of matched frozen tissues (N, T, M) in lysis buffer containing protease inhibitors. 40 μg of protein from each sample was denatured, fractionated by 10% SDS-PAGE, and transferred to PVDF membranes (Immobilon-P Transfer Membranes, Millipore, Milan, Italy). After blocking of non-specific antigens with 5% BSA solution, blots were incubated overnight at 4 °C with primary

mouse monoclonal antibody against caldesmon (clone 1.B.638:sc-70479, 1:100 working dilution, Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) in 5% BSA 0.05% PBS-Tween 20 buffer. Antibody binding to the membrane was detected with a secondary antibody (sheep anti-mouse IgG 1:5000, GE Healthcare, Milan, Italy) conjugated to horseradish peroxidase and visualized by enzyme-linked chemiluminescence (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific, Rockford, IL, USA) with the Chemidoc XRS System (Bio-Rad). After immunodetection, the PVDF membrane was stained with 0.1% Coomassie G-250 as loading control (Welinder and Ekblad, 2011). To normalize the signal of h-caldesmon and l-caldesmon, densitometric analysis was performed with ImageJ software, and the intensity of the two bands was normalized against the signal of the entire lane.

3. Results

3.1. Genome-wide gene and exon expression variation during CRC development

To understand the role of AS in CRC progression, we performed a genome-wide analysis in a group of patients, not only to find specific cases of substantial deregulation of AS in tumor progression, but also to describe the whole effect of AS deregulation in colon cancer development and in the metastatic process. To identify tumor-specific changes taking place from normal colon mucosa (N) to primary colorectal cancer (T) and subsequently to liver metastasis (M), we estimated gene and exon expression using RNA isolated from 80 samples comprising 23 N, 30 T and 27 M, obtained from 46 patients (for patient and tumor characteristics, see Table 1). This dataset included 27 samples belonging to 9 patients with three matched samples (T, N and M from the same patient) as detailed in Table 2. We used GeneChip Human Exon 1.0, containing about 5.4 million probes grouped into 1.4 million exon-level probesets, interrogating over one million exon clusters, classified according to the annotation level of the genomic probe selection regions. Exon-level expression measures were based on an average on 4 probes per exon and gene expression measures on an aggregation of all probes belonging to the same gene, with an average of 41 probes per gene.

To analyze both gene and exon expression, samples were tagged both according to sample type (N, T, M) and patient matching of samples, primary tumor topological site (left, right, rectum), type of metastasis (synchronous, metachronous) and gender. After ranking gene- and exon-level expression profiles by variability, descriptive cluster analyses of samples was performed, as described in Methods, and both gene- and exon-level expression information was used.

Figure 1 shows heatmaps with sample classification obtained in parallel according to the 1000 most variable genes (Figure 1A) and 1000 most variable exons (Figure 1B). Interestingly, our data show that the expression profiles of the most variable exons allow better separation of sample groups than those of the most variable genes. In addition, exon profiles are able not only to separate properly normal samples

from tumors and metastasis, but also to classify samples on the basis of possibly subtle sub-class differences. That is, normal colon tissue samples are almost perfectly classified according to patient gender. We studied this result further by analyzing differentially expressed genes and transcripts, and identified 10 genes with transcripts down-regulated in male patients. It was interesting to observe that the genes are mapped in chromosome X, whereas most of the genes with transcripts over-expressed in males are mapped in chromosome Y (See in Supplementary File 1, text and Table S1 for results). Conversely, some of the sample misclassifications appear to be due to similarity across samples from the same patient, i.e., a metastasis may be more similar to the primary tumor of the same patient than to other metastases from other patients. Cluster analysis does not clearly classify samples according to type of metastasis or topological site of primary tumor.

3.2. A robust set of candidate genes is involved in AS events during CRC progression

To identify AS events in CRC progression, we estimated exon-level expression with AltAnalyze and at transcript level with MMBGX, and then identified events with significant differential expression during tumor progression. MMBGX directly estimates isoform-level expression and therefore can be used to compare the expression of variants between conditions. Moreover, it takes into account the one-to-many mapping between probes and probesets and the complex pattern of exon sharing between different isoforms. The software has built-in knowledge of array structure and mapping between probes and Ensembl genes and transcripts and so is limited to available annotation (whereas exon level methods do not use any prior knowledge). MMBGX expects as input raw probe-level exon array data and produces, for each gene and transcript, a thousand samples from the posterior probability distribution of concentration using a Monte Carlo Markov Chain algorithm. Probability density plots of concentration of each molecular species can be obtained and a statistics called *pt* is implemented as a measure of the probability of differential splicing of each transcript.

To guarantee the possibility of comparing and integrating the results, AltAnalyze was applied to a customized set of

Table 2 – Description of sample groups for exon array dataset (N = normal colon mucosa, T = primary tumor, M = liver metastasis).

Sample characteristics			
Match type	No. of patients	Tissue type	No. of samples
N-T-M	9	N	23
N-T	5		
T-M	8	T	30
M-N	3		
N	6	M	27
T	8		
M	7	Total	80
Total	46		

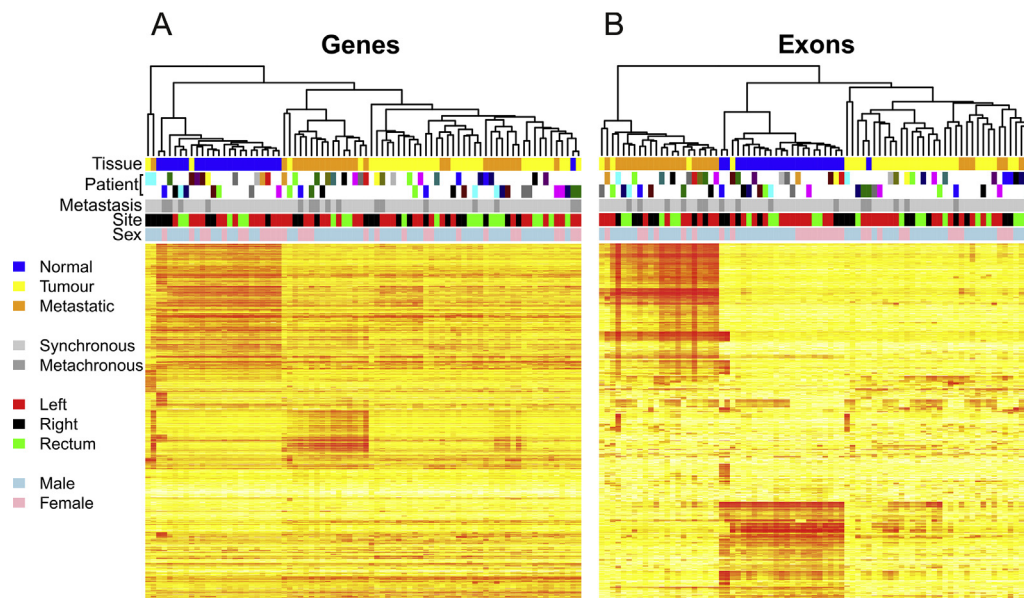


Figure 1 – Sample classification and heat map based on 1000 most variable genes (A) and exons (B) across sample set. Top: color map indicates tissue type, metastasis type, primary tumour site and patient gender. Second and third lines of top color map: patient matching of samples, showing triplets and pairs of samples from each patient in same color. Left: color code for top color map.

expression signals of 449,810 probesets, coinciding with the set available to MMBGX software annotation. The group of probesets in question comprises 97% of the Affymetrix core set (278,418 probesets), plus 171,392 probesets aligning to Ensembl exons, with 33,740 genes represented (Table 3).

3.2.1. Genes with significant AS exons during tumor progression

For the same sample group comparison, we used the FIRMA and MiDAS algorithms of AltAnalyze and detected various

numbers of AS events, with FIRMA identifying a higher number of AS events. MiDAS and FIRMA are exon level methods designed to infer splicing events looking at differential inclusion of an exon when two groups of samples are compared. Direct detection is not per se possible since Affymetrix exon arrays do not include probes spanning more than one exon as junction arrays do. MiDAS in particular compares via a classical one-way ANOVA test a full splicing model, which takes into consideration an interaction term between sample and exon, and a simplified one (no alternative splicing) where,

Table 3 – Functional enrichment of 206 genes involved in AS events in colon cancer development identified by integrated analysis.

KEGG ID	Term	P value	Gene Total	Fold Enrichment	FDR
hsa04510	Focal adhesion	5.89E-005	VWF, LAMA4, COL1A2, IGF1, ACTN1, RELN, VTN, THBS2, AKT3, VCL	5.584	0.067
hsa04512	ECM-receptor interaction	8.73E-005	VWF, LAMA4, COL1A2, RELN, VTN, AGRN, THBS2	9.352	0.099
ID GO-BP	Term BP	P value	Genes	Fold enrichment	FDR
GO:0007155	Cell adhesion	3,87E-004	SDK1, LMO7, ACTN1, VTN, SSPN, MMRN1, TINAG, VCL, NCAM1, VWF, COL17A1, LAMA4, FAT4, CD22, DSC2, SGCE, RELN, AMICA1, THBS2, DST, HABP2	2.420	0.627
GO:0022610	Biological adhesion	3,94E-004	SDK1, LMO7, ACTN1, VTN, SSPN, MMRN1, TINAG, VCL, NCAM1, VWF, COL17A1, LAMA4, FAT4, CD22, DSC2, SGCE, RELN, AMICA1, THBS2, DST, HABP2	2.420	0.638
ID GO-MF	Term MF	P value	Genes	Fold enrichment	FDR
GO:0005509	Calcium ion binding	3,37E-006	F13A1, UTRN, TPD52, CALB2, TMEM37, ATP2B4, FAT4, GSN, ITIH1, DMD, PLS1, AMY2B, ANO7, THBS2, PLA2G10, SCUBE2, ACTN1, MMP11, FBLN1, PLCE1, FBLN2, ATP2A3, CAPN13, SULF1, DSC2, RELN, SGCE, DST, LCP1	2.636	0.005
GO:0051015	Actin filament binding	3,71E-004	UTRN, PLS1, ACTN1, NEXN, DST, LCP1	9.633	0.508
GO:0003779	Actin binding	3,94E-004	CALD1, UTRN, ACTN1, LMO7, PALLD, NEXN, VCL, GSN, DMD, PLS1, DST, LCP1, TMOD1	3.435	0.540

the difference between the logged signal for the exon and its gene is expected to be a constant across all samples. If the full model fits better the data then exon is considered ASE. FIRMA does not try to estimate explicitly the discrepancy between observed and expected exon expression value in each sample to infer alternative inclusion. It rather fits the standard RMA model for the exon array and if large residuals for probes of a given exon are detected an alternative splicing event is assumed.

Independently of the specific measure considered, the number of genes with AS events in the transition from N to T was the most abundant (about 4% and 1% of genes, according to FIRMA and MIDAS measures, respectively), whereas the subsequent transition to liver metastasis accounted for one order of magnitude fewer AS (0.4% and 0.1%, respectively) (Figure 2A). About half genes giving rise to significant AS events during tumor progression overlap with known protein domains and motifs (data not shown). The numbers and proportions of ASEs identified by the various algorithms in the same transitions (NT, TM) follow the same trend as the AS genes (Figure 2B).

3.2.2. Differentially expressed transcripts

Using MMBGX, we were able to make probabilistic estimates of the expression level of transcripts in sample groups and select those genes in which at least one transcript was significantly more up-regulated in one condition with respect to the other, compared with the corresponding gene. As expected, and matching the results obtained with AltAnalyze, transcript-based analysis also identified the highest number of genes with AS, in comparisons of normal colon tissue with liver metastases (619), whereas more comparable numbers of AS genes were identified when the comparisons involved T vs M and N vs T tissues (Figure 3).

3.2.3. Candidate genes involved in AS events

To identify a more restricted but perhaps more robust set of genes involved in AS events, we combined genes with ASEs with genes with at least one DET, and obtained a list of 206 non-redundant candidate genes, of possible importance for colon cancer biology (See Figure 3 and the table in

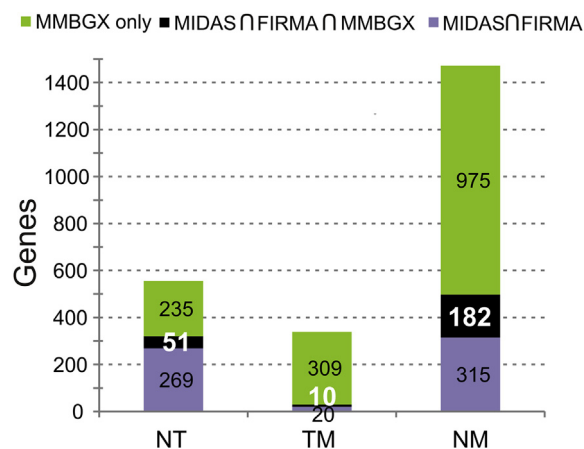


Figure 3 – Number of genes associated to alternative splicing in the considered sample contrasts. Green bars show genes with differentially expressed transcripts identified by MMBGX. Purple bars show the overlap (\cap = intersection) between the set of genes with exons deemed as alternatively spliced by FIRMA and MIDAS algorithms of AltAnalyze. Black bars show the overlap between the two previously described gene sets.

Supplementary File 2). For each sample comparison, html tables accessible online (http://compgen.bio.unipd.it/suppl_materials/molonc/2013/) report AltAnalyze results about differentially expressed probesets/exons, and intersections of AltAnalyze and MMBGX results. Exon probesets are hyper-linked to UCSC Genome Browser custom tracks color-coded according to specificity (See Methods).

The final list of genes for which we predicted significant differential splicing for each contrast comprises those genes interrogated by the exon-level probesets recognized as alternatively spliced by both MIDAS and FIRMA statistics and with at least one transcript with significantly variable expression in the sample groups considered, according to MMBGX.

For each contrast, Figure 3 shows genes interrogated by probesets recognized as alternatively regulated by both MIDAS and FIRMA statistics and deemed alternatively spliced

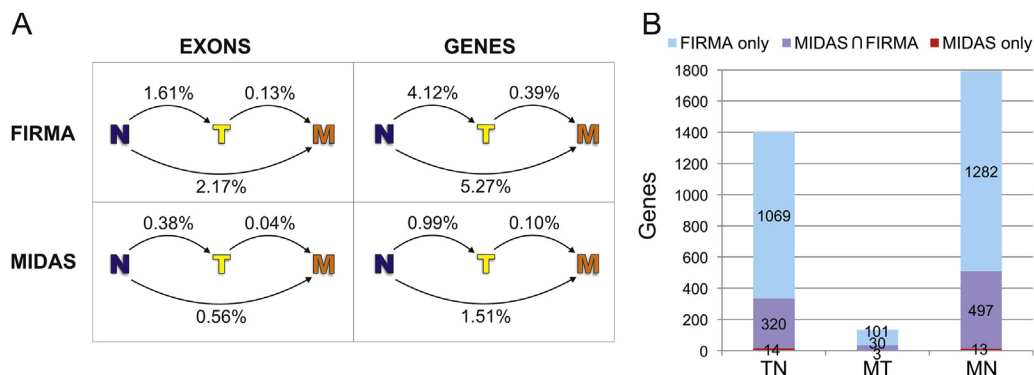


Figure 2 – Identification of exons and genes involved in contrast-specific AS events by FIRMA and MIDAS algorithms of AltAnalyze. (A) Percentages of genes and exons identified in pairwise sample contrasts, between normal colon mucosa (N), primary tumour (T) and liver metastasis. A total of 33,740 genes were considered in all contrasts; 171,128, 171,186 and 174,399 exons were considered for NT, TM and NM contrasts, respectively. (B) Overlap (\cap = intersection) between gene sets with exons deemed alternatively spliced by FIRMA and MIDAS algorithms, in different sample contrasts.

Table 4 – Additional information on alternative transcripts and their protein products considered for validation experiments.

Official gene symbol	Transcript ID	Protein ID: Ensembl identifier Uniprot identifier	Characteristics of protein isoforms	Length (aa)	Mass (Kda)
B3GNT6	B3GNT6-201	ENSP00000346256 Q6ZMB0-1	It synthesizes the core 3 structure of the O-glycan, an important precursor in the biosynthesis of mucin-type glycoproteins.	384	43
CALD1	CALD1-005	(Also known as H-CAD) ENSP00000354826 Q05682-1	It is an actin- and myosin-binding protein implicated in the regulation of actomyosin interactions in smooth muscle cells. Binds tightly and specifically to actin, calmodulin, tropomyosin, and myosin.	793	93
	CALD1-001	(Also known as WI-38 L-CAD II 1-CAD) ENSP00000354513 Q05682-4	It is expressed in endothelial cells and is involved in migration. Plays a role in cytoskeletal architecture and dynamics in non smooth muscle cells. Missing 254 aa after myosin and calmodulin-binding region in repeat regions (exclusion of exon 1).	538	63
	CALD1-012	ENSP00000411476C9J813	The residue at the extremity of the sequence is not the actual terminal residue in the complete protein sequence. No experimental confirmation available.	459	54
CTHRC1	CTHRC1-001	ENSP00000330523 Q96CG8-1	May act as a negative regulator of collagen matrix deposition. Can be tethered to the cell surface to promote actin polymerization and cell polarity.	243	26
VCL	VCL-201	(Also known as: Metavinculin, meta-VCL) ENSP00000211998 P 18206-2	It is actin filament (F-actin)-binding protein involved in cell-matrix adhesion and cell–cell adhesion. It is a muscle-specific isoform and is co-expressed with vinculin in muscle tissues. Regulates cell-surface E-cadherin expression and potentiates mechanosensing by the E-cadherin complex. May also play important roles in cell morphology and locomotion. It should be important for force transduction. Metavinculin seems unable to form actin filament bundles Contains an exon (exon 19) that alters the biochemical properties of the five-helix bundle in the tail domain.	1134	124
	VCL-001	(Also known as: Vinculin, VCL) ENSP00000361841 P 18206-1	It is actin filament (F-actin)-binding protein controls focal adhesion formation, strength, and migration. It regulates the structural integrity of cell–cell adhesions by mediating the mechano-response of E-cadherin. Vinculin is capable of bundling F-actin into thick bundles. Missing 67 aa in C-terminal tail.	1066	117

according to MMBGX. In this reduced list, five times more genes alternatively spliced genes in the N-T and T-M transitions (51 and 10 genes, respectively), whereas the larger difference was observed when normal tissue was compared with metastases (182 genes). Overall, our analysis of the total influence of AS events in CRC progression indicated that the highest variability is mainly restricted to transformation to primary tumor, while a reduced number of AS events takes place in progression to metastatic tissue.

The group of 206 genes involved in AS events in CRC development was significantly enriched in genes belonging to Focal Adhesion (10 genes) and ECM-receptor interaction (7) KEGG Pathways, playing roles in adhesion processes (21) and with molecular functions mainly related to calcium ion binding (29), actin filament binding (6) and actin binding (13) (Table 3).

3.3. Validation of AS transcripts differentially expressed during tumor progression

Given the complexity of the computational analysis required to reconstruct alternative transcripts (AT) expression, we validated the robustness of results obtained with the genome-wide approach by quantitative RT-PCR (RT-qPCR). Genes with AS events were reviewed manually to take into

account probe cross-hybridization to multiple exons/genes, and probe alignments to transcripts and ESTs, identifying 18 best candidates (Table S3 in Supplementary File 3). Also taking into consideration available biological information on these genes and previous results about pathway and functional enrichment of the largest gene set, we selected for experimental validation of 5 candidate DETs belonging to 4 genes, CALD1, VCL, B3GNT6 and CTHRC1 (Table 4). These genes are specifically involved in cell adhesion, cell–cell and cell-matrix junctions and cytoskeleton organization (CALD1 and VCL), vascular remodeling (CTHRC1) and biosynthetic/metabolic processes (B3GNT6), all being processes important for colon biology.

For the transcripts belonging to VCL gene, known to express multiple transcripts, we designed sets of primers to amplify the differentially expressed transcripts (Table S4A in Supplementary File 4). Regarding CALD1 gene, associated to 13 protein coding isoforms (Table S3), we used primers specific only for the CALD1-012 transcript. For transcripts B3GNT6-201 and CTHRC1-001, we used commercially available probes ensuring transcript specificity (Table S4B in Supplementary File 4). RT-qPCR was performed with total RNA extracted from at least 70 of the 80 tissue samples used for the exon chip (Table 2).

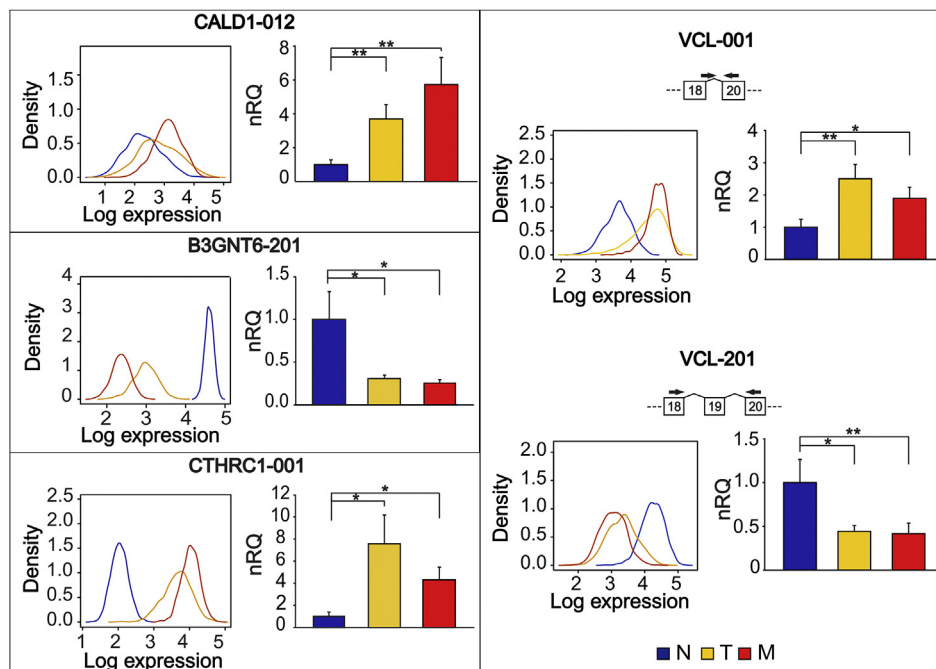


Figure 4 – Validation of microarray-detected alternative splicing events in N, T and M tissues from CRC patients. For each transcript, expression in N (blue), T (yellow) and M (red) samples was estimated from exon array data (density curves with probabilistic estimation of transcript expression; left panel), and measured by RT-qPCR (right panel). RT-qPCR data are shown as mean \pm SE of 3 experiments performed in triplicate. * $P < 0.05$ vs normal colon mucosa. ** $P < 0.01$ vs normal colon mucosa. nRQ: normalized Relative Quantity. For VCL alternative transcripts, sequence-validated exon structures are shown and arrows indicate primer positions.

Of the selected genes, CALD1 and CTHRC1 are particularly interesting, since they are among 48 genes significantly associated with risk of recurrence in four independent studies of patients with stage II/III colon cancer, treated with surgery alone or surgery plus chemotherapy (O’Connell et al., 2010).

In the case of CALD1-012 transcript isoform, results of RT-qPCR are in agreement with those obtained by MMBGX estimate and show that this transcript is not only significantly increased in the primary tumor, but it is further up-regulated in metastases (Figure 4).

VCL gene encodes an actin filament (F-actin)-binding protein involved in focal adhesion and migration. As a scaffolding protein, vinculin binds to many different ligands. We considered two isoforms VCL-001 (vinculin) and VCL-201 (metavinculin), the latter containing a region encoded an additional exon (exon 19). This is a relatively poorly conserved sequence, whose inclusion alters the structural and biochemical properties of the tail domain. As shown in Figure 4, we confirmed by RT-qPCR that both isoforms are differentially expressed when comparing normal colon mucosa with primary tumour and with metastasis, showing an opposite behavior: VCL-001 is up regulated whereas VCL-201 is down regulated in T and M tissues compared to the normal counterpart, with no difference observed between T and M for both transcripts (Thompson et al., 2013).

CTHRC1 codes for a secreted protein, which is considered a potential biomarker for diagnosis, since its expression level in CRC is increased. Our results clearly indicate that the expression of transcript CTHRC1-001 is significantly increased both

in primary and metastatic CRC, compared to normal tissue (Figure 4).

Gene B3GNT6 codes for a glycosyltransferase, which adds stepwise carbohydrates to form the core 3 O-glycan structure, restricted in its occurrence to mucins present in specialized tissues such as colon. It has been reported that the core 3 structure in colon cancer tissues is reduced as the activity of core 3 synthase is lower (Brockhausen, 1999; Kim, 1998), and also that the expression of the protein gradually disappears as the grade of the tumor progresses (Iwai et al., 2005). Accordingly, our genomic analysis showed that transcript B3GNT6-201 was down-regulated both in tumour and in metastasis, compared with normal tissue, and RT-qPCR did confirm this result (Figure 4).

We used Western blot (WB) analysis for studying the protein isoforms encoded by the different caldesmon transcripts in an independent set of three new patients with matched samples (N, T, M). According to a search of the Ensembl database, AS of the gene encoding CALD1 results in 26 transcripts of which 13 actually code for a protein product (Table S3 in Supplementary File 3) involved in cell motility and actin cytoskeleton remodeling. Of these, CALD-005 is the longest transcript, containing an extended form of exons 5 and 6, giving rise to a high molecular weight isoform of 793 aa (h-Caldesmon), mainly expressed by smooth muscle (Lin et al., 2009). By MMBGX analysis we estimated that only three transcripts belonging to the CALD1 gene (CALD1-001, CALD1-005 and CALD1-012) are differentially expressed (Figure 5A and Table 4). Of these, CALD1-005 showed a predominant expression

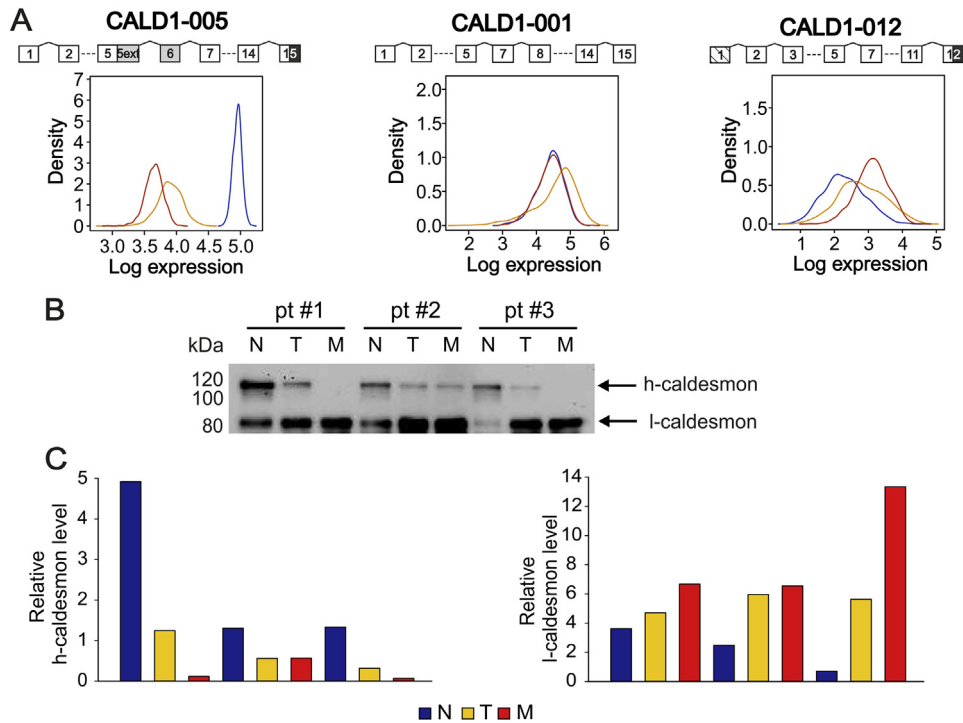


Figure 5 – Detection of caldesmon isoforms expression in matched N, T and M tissue. (A) For each CALD1 transcript, the plots show the expression estimations obtained by MMBGX probabilistic analysis of exon array data in N (blue), T (yellow) and M (red) samples from 46 patients. The exon structure of alternative transcripts is reported above, with boxes indicating the differences among isoforms: the exon and the region included only in CALD1-005 are shown in grey, the shorter exons 12 and 15 are shown in dark grey and the alternative exon 1 is shown by the dashed texture. (B) Western blot analysis with primary monoclonal antibody detecting h-caldesmon protein band at 120 kDa and l-caldesmon protein band at 80 kDa, performed in 9 matched samples from 3 patients. (C) Quantification of h-caldesmon and l-caldesmon protein level by densitometric analysis in the same samples considered for Western blot.

in normal colon mucosa, compared with tumor tissue, probably reflecting the relatively higher expression of smooth muscle in this tissue (Figure 5A). This result was also supported at protein level, since WB analysis showed that h-Caldesmon was mainly expressed in N tissue from all 3 patients, and progressively decreased in T and M tissues (Figure 5). All the other isoforms, collectively called l-caldesmon, are ubiquitously expressed in non-muscle cells, but their significance remains to be determined. Because of their high content of glutamine residues, during SDS-PAGE h-caldesmon and l-caldesmon isoforms migrate to a seeming molecular weight of 120 and 70–80 kDa, respectively. We show that the expression of an 80-kDa band increases from N to T and from T to M tissue, thus confirming results observed with transcripts CALD-001 and CALD-012 in RT-qPCR and exon analysis (Figures 4 and 5).

4. Discussion

In principle, arrays or sequencing-based technologies are useful for transcriptome characterization. With both exon arrays and RNA-seq, information supplied by short elements (probes or sequencing reads) directly refers to gene exons, whereas inference of the quality and expression of alternative

transcripts is technically complicated by the existence of many isoforms per gene, with complex patterns of exon sharing between isoforms. Several problems still limit the potential effect of AS data on the ability of generating fruitful information. First of all, the complexity of the data requires new computational methods; another challenge is to understand whether these transcriptional changes effectively translate into different transcripts. Exon arrays are not designed for direct observation of a splicing event, as junction arrays are, but they do allow the inference of the occurrence of splicing events. Various computational methods have been developed with this in mind, mainly based on detecting large-scale changes in the expression of individual exons relative to the gene level signal. We selected two different methods implemented in the AltAnalyze package. MiDAS is based on a sound statistical model and is quite conservative. FIRMA instead is designed to perform well also in case of high intra-group variability (Purdom et al., 2008).

However, a significant statistic at exon level is not really enough to deduce how each of the multiple gene transcripts vary in sample groups, because the exon of interest may be shared by many transcripts, as the probes designed to bind it cross-hybridize with other exons of the same gene or even with other genes containing similar sequences. In these cases, it is difficult to determine how much of the signal change in

the probeset is attributable to the expression change of the exon of interest or to other cross-hybridizing exons. Since exon-level statistics can neither detect nor quantify the abundance of each individual isoform, and may thus miss the most biologically significant information embedded in exon array data, in the present study we used an integrated procedure based on a combination of results obtained with AltAnalyze with those of MMBGX, which identifies differentially expressed transcripts according to the probabilistic estimation of individual transcript expression levels in sample groups. This integration phase presented some challenges, e.g., for establishing homogeneous annotations and sets of probes used by different methods, but it is important, since exon- and transcript-level methods are orthogonal and have diverse advantages and disadvantages. As exon-level methods are characterized by low specificity, we applied ancillary methods to increase specificity by filtering out probesets prior to analysis. MMBGX can directly estimate isoform-level expression and can therefore be used to compare the expression of variants between conditions and even within a single sample. For probabilistic reconstruction MMBGX relies on transcript structures, as reported in Ensembl annotation.

Our results obtained from unsupervised hierarchical clustering showed that exon-level expression can classify tissues more efficiently than gene expression, at least as regards normal colon mucosa from primary CRC and liver metastases. To the best of our knowledge, exon expression has never been used to classify tumor samples, and our data indicate that this tool can be exploited as a subclass discovery tool, potentially able to exploit fine differences between seemingly homogeneous tumor samples. It is not surprising that classification at exon level is superior than that obtained at gene level, considering that a measure at the so-called gene level is actually the average value of a number of transcript values, whereas an exon-level measure evaluates the actual transcribed exons and therefore reflects the transcriptome more accurately.

We have recently observed that, during CRC progression, the main changes in miRNA expression take place in primary tumors, with small variation in metastatic lesion (Pizzini et al., 2013). The results presented here extend our previous findings and also confirm that, at the level of AT, the main changes follow the same pattern, with few changes during the metastatic process. Consistently, hierarchical clustering of genes or exons clearly distinguishes between normal colon and tumors, whereas primary tumors and metastases are partly mixed and display a lesser degree of dissimilarity. Collectively, these results stress the importance of alterations occurring during malignant transformation into primary tumors, and establish a phenotype which is almost stable in metastasis and which has a profound effect on the outcome of the tumor.

In this study an integrative framework allowed us to define with high confidence a set of 206 genes showing significant AS events during tumor development and/or progression. This group of genes was found enriched in biological processes and pathways highly relevant for cancer progression, as adhesion and ECM-receptor interaction. We selected 18 most trustworthy candidate genes with significant AS events. We further investigated four genes among them, associated to some aspects of cancer development, and confirmed the differential

expression of at least one transcript per gene in colorectal cancer progression.

The group of 18 genes selected includes relevant cancer genes. Moreover, alternative splicing events important in cancer development are reported for experimentally validated genes (see below) as well as for at least other three genes. CD79a gene (ENSG00000105369), that encodes an Ig-alpha protein of the B-cell antigen component, has three different transcripts. An alternatively spliced transcript variant (DeltaCD79b) is involved in B-chronic lymphocytic leukemia (Cragg et al., 2002). PDE4B gene (ENSG00000184588), a cAMP-specific phosphodiesterase, presents 19 transcripts, 11 of which are protein coding. It is known that the genes (PDE4A, 4B, 4C, 4D) of the PDE4 gene family are associated to multiple splicing variants. Transcripts variants from PDE4B gene were found in melanoma (Narita et al., 2007). SLC39A14 gene (ENSG00000104635) encodes a divalent cation transporter that control gene transcription, growth, development, and differentiation. Alternative splicing of SLC39A14 is involved in colorectal cancer it is under the control of the Wnt signalling pathway (Sveen et al., 2012; Thorsen et al., 2011).

Among the genes with validated AS events there is the VCL gene. It encodes an F-actin binding cytoskeletal protein associated with cell–cell and cell–matrix junctions. We found that the two isoforms VCL-001 (vinculin) and VCL-201 (metavinculin, containing a region encoded an additional exon 19) are differentially expressed when comparing normal colon mucosa with primary tumour and with metastasis. The opposite behavior of the two transcripts indicates that the exon 19 skipping tend to increase in cancer cells. Even if much remains to be discovered concerning the differences in function between vinculin and metavinculin, it is accepted that the two forms have differential affinity for diverse ligands and different oligomerisation properties, supporting the biological relevance of the observed AS event (Thompson et al., 2013).

Furthermore, our findings related to the transcripts belonging to CALD-1 gene and differentially expressed in CRC progression are of particular interest. Caldesmon is an actin-linked regulatory protein, found in smooth muscle and non-muscle cells, with important functions in cell motility, including migration, invasion and proliferation. AS events of caldesmon have already been reported between CRC and normal colon tissue, probably reflecting the involvement of AT with specific roles in these important processes (Gardina et al., 2006; Thorsen et al., 2008). It has also been demonstrated that caldesmon isoforms, encoded by the CALD-1 gene, are differentially expressed in tumor tissue and stroma embedded in colon adenocarcinoma and metastatic tissue (Kohler, 2011). Specifically, it has been shown, by immunohistochemistry with three different antibodies against human caldesmon isoforms, that the longest h-caldesmon isoform is mainly expressed in smooth muscle cells and also in pericryptal fibroblasts in the colon. In colorectal adenocarcinomas, h-caldesmon expression is markedly reduced in large areas of the stroma and cancer epithelium did not stain for h-caldesmon; lymph-node metastases also displayed little h-caldesmon immunoreactivity of the stroma (Kohler, 2011). Our results match those of this study, as this long isoforms down-regulated at both mRNA and protein level (Figures 4 and 5). As regards the shortest isoforms, the results of exon chips

were validated at mRNA level (Figure 4), but the antibody we used could not discriminate among the shortest isoforms, although an increase could be documented when liver metastases from colon mucosa were compared. It thus appears that many isoforms of CALD1 are modulated during CRC progression, although also in this case the significant transition is from normal tissue to primary tumor.

Considering that AS is a common mechanism used by more than 95% of human genes, and that AT may play different roles in many important processes, knowledge of AT in cancer may have important consequences. In particular, splice isoforms can be used as markers of tumor progression and thus may represent the objective of therapeutic interventions targeting a single isoform in a key regulatory pathway, a set of isoforms, or their regulatory network. Therefore, research in this field is only beginning to yield results.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by funds from the Italian Ministry of Health to SB (2010NYKNS7_002), the Italian Association for Cancer Research (AIRC) to PZ (Regional Research Program 2008 grant # 6421) and to SB (Special Program Molecular Clinical Oncology 5x1000 to AGIMM, AIRC-Gruppo Italiano Malattie Mieloproliferative, project number #1005) and the Fondazione Cassa di Risparmio di Padova e Rovigo (Excellence project 2011/12) to SB. SP was an AIRC fellow. We thank Prof. Roberto Turolla for providing access to computational facilities.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.molonc.2013.10.004>.

REFERENCES

- Akgul, C., Moulding, D.A., Edwards, S.W., 2004. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell. Mol. Life Sci. CMLS* 61, 2189–2199.
- Birkenkamp-Demtroder, K., Olesen, S.H., Sorensen, F.B., Laurberg, S., Laiho, P., Aaltonen, L.A., Orntoft, T.F., 2005. Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid. *Gut* 54, 374–384.
- Bisognin, A., Sales, G., Coppe, A., Bortoluzzi, S., Romualdi, C., 2012. MAGIA(2): from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). *Nucleic Acids Res.* 40, W13–W21.
- Brinkman, B.M., 2004. Splice variants as cancer biomarkers. *Clin. Biochem.* 37, 584–594.
- Brockhausen, I., 1999. Pathways of O-glycan biosynthesis in cancer cells. *Biochim. biophys. acta* 1473, 67–95.
- Cancer Genome Atlas, N, 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.
- Cardoso, J., Boer, J., Morreau, H., Fodde, R., 2007. Expression and genomic profiling of colorectal cancer. *Biochim. biophys. acta* 1775, 103–137.
- Cragg, M.S., Chan, H.T., Fox, M.D., Tutt, A., Smith, A., Oscier, D.G., Hamblin, T.J., Glennie, M.J., 2002. The alternative transcript of CD79b is overexpressed in B-CLL and inhibits signaling for apoptosis. *Blood* 100, 3068–3076.
- Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., Wrana, J.L., Blencowe, B.J., 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46, 884–892.
- Emig, D., Salomonis, N., Baumbach, J., Lengauer, T., Conklin, B.R., Albrecht, M., 2010. AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res.* 38, W755–W762.
- Fearon, E.R., Vogelstein, B., 1990. A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767.
- Gardina, P.J., Clark, T.A., Shimada, B., Staples, M.K., Yang, Q., Veitch, J., Schweitzer, A., Awad, T., Sugnet, C., Dee, S., Davies, C., Williams, A., Turpaz, Y., 2006. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* 7, 325.
- Germann, S., Gratadou, L., Dutertre, M., Auboeuf, D., 2012. Splicing programs and cancer. *J. Nucleic Acids* 2012, 269570.
- Gleisner, A.L., Choti, M.A., Assumpcao, L., Nathan, H., Schulick, R.D., Pawlik, T.M., 2008. Colorectal liver metastases: recurrence and survival following hepatic resection, radiofrequency ablation, and combined resection-radiofrequency ablation. *Arch. Surg.* 143, 1204–1212.
- Goehe, R.W., Shultz, J.C., Murudkar, C., Usanovic, S., Lamour, N.F., Massey, D.H., Zhang, L., Camidge, D.R., Shay, J.W., Minna, J.D., Chalfant, C.E., 2010. hnRNP L regulates the tumorigenic capacity of lung cancer xenografts in mice via caspase-9 pre-mRNA processing. *J. Clin. Invest.* 120, 3923–3939.
- Gutschner, T., Hammerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Gross, M., Zornig, M., MacLeod, A.R., Spector, D.L., Diederichs, S., 2013. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189.
- Habermann, J.K., Bader, F.G., Franke, C., Zimmermann, K., Gemoll, T., Fritzsche, B., Ried, T., Auer, G., Bruch, H.P., Roblick, U.J., 2008. From the genome to the proteome—biomarkers in colorectal cancer. *Langenbeck's arch. surg./Deutsche Gesellschaft Chirurgie* 393, 93–104.
- Iwai, T., Kudo, T., Kawamoto, R., Kubota, T., Togayachi, A., Hiruma, T., Okada, T., Kawamoto, T., Morozumi, K., Narimatsu, H., 2005. Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci. U S A* 102, 4572–4577.
- Jemal, A., Ward, E., Hao, Y., Thun, M., 2005. Trends in the leading causes of death in the United States, 1970–2002. *JAMA J. Am. Med. Assoc.* 294, 1255–1259.
- Kim, Y.S., 1998. Mucin glycoproteins in colonic neoplasia. *Keio J. Med.* 47, 10–18.
- Kohler, C., 2011. Histochemical localization of caldesmon isoforms in colon adenocarcinoma and lymph node metastases. *Virchows Arch.* 459, 81–89.
- Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., Munoz, M.J., 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature reviews. Mol. Cell Biology* 14, 153–165.

- LaPointe, L.C., Dunne, R., Brown, G.S., Worthley, D.L., Molloy, P.L., Wattoo, D., Young, G.P., 2008. Map of differential transcript expression in the normal human large intestine. *Physiol. Genomics* 33, 50–64.
- Lin, J.J., Li, Y., Eppinga, R.D., Wang, Q., Jin, J.P., 2009. Chapter 1: roles of caldesmon in cell motility and actin cytoskeleton remodeling. *Int. Rev. Cell Mol. Biol.* 274, 1–68.
- Lionetti, M., Biasiolo, M., Agnelli, L., Todoerti, K., Mosca, L., Fabris, S., Sales, G., Delilieri, G.L., Bicciato, S., Lombardi, L., Bortoluzzi, S., Neri, A., 2009. Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma. *Blood* 114, e20–e26.
- Miura, K., Fujibuchi, W., Sasaki, I., 2011. Alternative pre-mRNA splicing in digestive tract malignancy. *Cancer Sci.* 102, 309–316.
- Miura, K., Fujibuchi, W., Unno, M., 2012. Splice isoforms as therapeutic targets for colorectal cancer. *Carcinogenesis* 33, 2311–2319.
- Mojica, W., Hawthorn, L., 2010. Normal colon epithelium: a dataset for the analysis of gene expression and alternative splicing events in colon disease. *BMC Genomics* 11, 5.
- Morris, A.R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A.S., Carvalho, B., Meijer, G.A., Agami, R., 2012. Alternative cleavage and polyadenylation during colorectal cancer development. *Clin. Cancer Res.* 18, 5256–5266.
- Narita, M., Murata, T., Shimizu, K., Nakagawa, T., Sugiyama, T., Inui, M., Hiramoto, K., Tagawa, T., 2007. A role for cyclic nucleotide phosphodiesterase 4 in regulation of the growth of human malignant melanoma cells. *Oncol. Reports* 17, 1133–1139.
- O'Connell, M.J., Lavery, I., Yothers, G., Paik, S., Clark-Langone, K.M., Lopatin, M., Watson, D., Baehner, F.L., Shak, S., Baker, J., Cowens, J.W., Wolmark, N., 2010. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. *J. Clin. Oncol.* 28, 3937–3944.
- Ogino, S., Goel, A., 2008. Molecular classification and correlates in colorectal cancer. *J. Mol. Diag.* 10, 13–27.
- Pal, S., Gupta, R., Davuluri, R.V., 2012. Alternative transcription and alternative splicing in cancer. *Pharmacol. Ther.* 136, 283–294.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415.
- Pizzini, S., Bisognin, A., Mandruzzato, S., Biasiolo, M., Faccioli, A., Perilli, L., Rossi, E., Esposito, G., Rugge, M., Pilati, P., Mocellin, S., Nitti, D., Bortoluzzi, S., Zanovello, P., 2013. Impact of microRNAs on regulatory networks and pathways in human colorectal carcinogenesis and development of metastasis. *BMC Genomics* 14, 589.
- Purdom, E., Simpson, K.M., Robinson, M.D., Conboy, J.G., Lapuk, A.V., Speed, T.P., 2008. FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics* 24, 1707–1714.
- Ritchie, W., Granjeaud, S., Puthier, D., Gautheret, D., 2008. Entropy measures quantify global splicing disorders in cancer. *PLoS Comput. Biol.* 4, e1000011.
- Sheffer, M., Bacolod, M.D., Zuk, O., Gardina, S.F., Pincas, H., Barany, F., Paty, P.B., Gerald, W.L., Notterman, D.A., Domany, E., 2009. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc. Natl. Acad. Sci. U S A* 106, 7131–7136.
- Shkreta, L., Froehlich, U., Paquet, E.R., Toutant, J., Elela, S.A., Chabot, B., 2008. Anticancer drugs affect the alternative splicing of Bcl-x and other human apoptotic genes. *Mol. Cancer Ther.* 7, 1398–1409.
- Sveen, A., Bakken, A.C., Agesen, T.H., Lind, G.E., Nesbakken, A., Nordgard, O., Brackmann, S., Rognum, T.O., Lothe, R.A., Skotheim, R.I., 2012. The exon-level biomarker SLC39A14 has organ-confined cancer-specificity in colorectal cancer. *Int. J. Cancer. J. Int. Cancer* 131, 1479–1485.
- Taft, R.J., Simons, C., Nahkuri, S., Oey, H., Korbie, D.J., Mercer, T.R., Holst, J., Ritchie, W., Wong, J.J., Rasko, J.E., Rokhsar, D.S., Degnan, B.M., Mattick, J.S., 2010. Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nature Struct. Mol. Biol.* 17, 1030–1034.
- Tazi, J., Bakkour, N., Stamm, S., 2009. Alternative splicing and disease. *Biochim. Biophys. Acta* 1792, 14–26.
- Thompson, P.M., Tolbert, C.E., Campbell, S.L., 2013. Vinculin and metavinculin: oligomerization and interactions with F-actin. *FEBS Letters* 587, 1220–1229.
- Thorsen, K., Mansilla, F., Schepeler, T., Oster, B., Rasmussen, M.H., Dyrskjot, L., Karni, R., Akerman, M., Krainer, A.R., Laurberg, S., Andersen, C.L., Orntoft, T.F., 2011. Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway. *Mol. Cell. Prot. MCP* 10, M110 002998.
- Thorsen, K., Sorensen, K.D., Brems-Eskildsen, A.S., Modin, C., Gaustadnes, M., Hein, A.M., Kruhoffer, M., Laurberg, S., Borre, M., Wang, K., Brunak, S., Krainer, A.R., Topping, N., Dyrskjot, L., Andersen, C.L., Orntoft, T.F., 2008. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol. Cell. Prot. MCP* 7, 1214–1224.
- Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., Blencowe, B.J., Prasanth, S.G., Prasanth, K.V., 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell.* 39, 925–938.
- Turro, E., Lewin, A., Rose, A., Dallman, M.J., Richardson, S., 2010. MMBGX: a method for estimating expression at the isoform level and detecting differential splicing using whole-transcript Affymetrix arrays. *Nucl. Acids Res.* 38, e4.
- Wang, E.T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
- Ward, A.J., Cooper, T.A., 2010. The pathobiology of splicing. *J. Pathol.* 220, 152–163.
- Welinder, C., Ekblad, L., 2011. Coomassie staining as loading control in Western blot analysis. *J. Prot. Res.* 10, 1416–1419.
- Zardo, G., Ciolfi, A., Vian, L., Billi, M., Racanicchi, S., Grignani, F., Nervi, C., 2012. Transcriptional targeting by microRNA-polycomb complexes: a novel route in cell fate determination. *Cell Cycle* 11, 3543–3549.