Expression profiling of budding cells in colorectal cancer reveals an EMT-like phenotype and molecular subtype switching

Linde De Smedt[1], Sofie Palmans[1], Daan Andel[1], Olivier Govaere[1], Bram Boeckx[2,3], Dominiek Smeets[2,3], Eva Galle[2,3], Jasper Wouters[1], David Barras[4], Madeleine Suffiotti[4], Jeroen Dekervel[5], Thomas Tousseyn[1,6], Gert De Hertogh[1,6], Hans Prenen[7], Sabine Tejpar[7], Diether Lambrechts[2,3] and Xavier Sagaert*[1,6]

[1]Translational Cell and Tissue Research Unit, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium; [2]Vesalius Research Center, Vlaams instituut voor Biotechnologie (VIB), Leuven, Belgium; [3]Laboratory for Translational Genetics, Department of Oncology, KU Leuven, Leuven, Belgium; [4]Bioinformatics Core Facility, Swiss Institute for Bioinformatics, Lausanne, Switzerland; [5]Laboratory of Hepatology, Department of Clinical and Experimental Medicine, University Hospitals Leuven, Leuven, Belgium; [6]Department of Pathology, University Hospitals Leuven, Leuven, Belgium and [7]Digestive Oncology Unit, Department of Oncology, University Hospitals Leuven, Leuven, Belgium

Background: Tumour budding, described as the presence of single cells or small clusters of up to five tumour cells at the invasive margin, is established as a prognostic marker in colorectal carcinoma. In the present study, we aimed to investigate the molecular signature of tumour budding cells and the corresponding tumour bulk.

Methods: Tumour bulk and budding areas were microdissected and processed for RNA-sequencing. As little RNA was obtained from budding cells, a special low-input mRNA library preparation protocol was used. Gene expression profiles of budding as compared with tumour bulk were investigated for established EMT signatures, consensus molecular subtype (CMS), gene set enrichment and pathway analysis.

Results: A total of 296 genes were differentially expressed with an FDR <0.05 and a twofold change between tumour bulk and budding regions. Genes that were upregulated in the budding signature were mainly involved in cell migration and survival while downregulated genes were important for cell proliferation. Supervised clustering according to an established EMT gene signature categorised budding regions as EMT-positive, whereas tumour bulk was considered EMT-negative. Furthermore, a shift from CMS2 (epithelial) to CMS4 (mesenchymal) was observed as tumour cells transit from the tumour bulk to the budding regions.

Conclusions: Tumour budding regions are characterised by a phenotype switch compared with the tumour bulk, involving the acquisition of migratory characteristics and a decrease in cell proliferation. In particular, most tumour budding signatures were EMT-positive and switched from an epithelial subtype (CMS2) in the tumour bulk to a mesenchymal subtype (CMS4) in budding cells.

Colorectal cancer (CRC) remains a major cause of cancer-related death. The most robust prognostic indicator for CRC is TNM-staging. This classification system includes depth of invasion (T), extent of nodal involvement (N) and presence of distant metastasis (M; Edge and Compton, 2010). Yet, there are significant survival differences for patients in the same disease stage. Therefore, the quest for additional prognostic parameters remains a hot topic in CRC-related research. Various publications report tumour

budding, which is defined as single tumour cells or small groups up to five tumour cells at the invasive margin, as an additional prognostic factor (Hase et al, 1993; Ueno et al, 2002; Okuyama et al, 2003; Ha et al, 2005; Nakamura et al, 2008; Ohtsuki et al, 2008; Wöhlke et al, 2011; Betge et al, 2012; Horcic et al, 2013; Karamitopoulou et al, 2013). Depending on the study, tumour budding is observed in around 40% of the diagnosed CRC cases (Okuyama et al, 2002). Besides in CRC, tumour budding is also reported in a variety of other tumours like oral squamous cell carcinoma or oesophageal cancer (Roh et al, 2004; Jensen et al, 2015).

Tumour budding is considered to be the first step in cancer metastasis, as budding cells are thought to migrate through the extracellular matrix, invade lymphovascular structures and form metastatic tumour colonies in lymph nodes and at distant sites (Lugli et al, 2012). Although budding cells are considered to be the histologic reflection of epithelial–mesenchymal transition (EMT), this hypothesis is not validated so far and the mechanisms by which budding cells detach from the main tumour are not clear (Zlobec and Lugli, 2010). Indeed, so far, only immunohistochemical studies have been performed in an attempt to unravel the biological basis of tumour budding (Zlobec et al, 2007; Zlobec and Lugli, 2010). High-throughput experiments on tumour samples that reveal the complete molecular fingerprint of tumour budding cells are lacking, but nevertheless may be of great value in understanding the invasive dynamics of migration, invasion and metastasis of cancer.

Tumour cells that undergo EMT are characterised by loss of epithelial and acquisition of mesenchymal properties (Bhangu et al, 2012; Talbot et al, 2012). One of the hallmarks of EMT is downregulation of E-cadherin, a transmembrane cell–cell adhesion molecule required for the formation of adherens junctions (Xu et al, 2009). The cytoplasmic tail of E-cadherin binds with β-catenin, forming a link with the cytoskeleton. During EMT loss of E-cadherin results in breakdown of cell junctions and nuclear translocation of β-catenin is observed, where it functions as an oncogenic transcription factor inducing upregulation of EMT-related transcription factors (e.g., SNAIL, ZEB and TWIST transcription factors). However, β-catenin is not the sole inducer of SNAIL, ZEB and TWIST expression (Peinado et al, 2007; Xu et al, 2009; Dawson and Lugli, 2015). TWIST, SNAIL and ZEB transcription factors are important inhibitors of epithelial differentiation markers (e.g., membranous E-cadherin expression) and induce the acquisition of a mesenchymal phenotype through upregulation of fibronectin, N-cadherin, matrix metalloproteinases and so on (Talbot et al, 2012; Tan et al, 2014; Isella et al, 2015).

In the present study, we investigated gene expression profiles in spatially defined CRC tumour areas, that is, in the tumour bulk as well as in tumour budding cells isolated by laser microdissection. Although the invasion front has been molecularly analysed before on the tissue (Hlubek et al, 2007; Horst et al, 2009), we are the first to investigate the molecular expression signature of individual budding cells and manage to increase our insights into the mechanisms of tumour budding.

MATERIALS AND METHODS

Patient selection. Eight cases of CRC were selected from the archive of the pathology department of the University Hospitals of Leuven (Belgium). Following inclusion criteria were set: (1) diagnosis of CRC was established according to the criteria of the World Health Organisation (WHO); (2) diagnosis was made between 2003 and 2012; (3) patients were free from pre-operative radio or chemotherapy to avoid interference with the grade of tumour budding and gene expression patterns; (4) high levels of

tumour budding were observed in both formalin-fixed, paraffin-embedded (FFPE) and fresh frozen (FF) tissues; and (5) a well or moderately differentiated morphologic pattern (as budding in poorly differentiated carcinomas is more difficult to assess because of the often non-cohesive growth pattern; Figure 1A). This study was approved by the institutional ethics commission of the University Hospitals Leuven, Belgium

Immunohistochemistry. Five-micron-thick FFPE sections were immunostained for cytokeratin (Clone AE1/AE3, Dako, Glostrup, Denmark) in an automated manner using the Bond-Max autostainer (Leica, Nussloch, Germany) according to the manufacturer's protocol with heat-induced epitope retrieval at pH 9.

Assessment of tumour budding. Tumour budding was quantified on prekeratin-immunostained sections. The slides were scanned on low magnification ($\times 100$) for the most dense budding region. In this region, tumour budding foci were counted on a $\times 400$ high-power field (HPF). Sections with more than 10 budding foci per HPF were graded as budding-high, otherwise tumours were classified as budding-low. Haematoxylin and eosin (H&E) stained sections of matched FF tissue were scored for degree of tumour budding as well. Score 1 was given to cases with low-grade budding, score 2 for cases with moderate budding and score 3 for cases presenting with extensive budding (Figure 1B). Only cases with high-grade budding in FFPE and FF material were included in the study.

Laser capture microdissection. Eight-micrometre-thick FF sections were cut and placed on RNase-free metal-framed poly-ethylene terephthalate (PET) slides (Leica) followed by a rapid nuclear staining protocol. Therefore, the slides were rehydrated in a series of ethanol (95%, 75%, 50%) and incubated in cresyl violet acetate (Sigma-Aldrich, St Louis, MO, USA) for 1 min. Afterwards, the slides were dehydrated in an ethanol series (50%, 75%, 95%, 100%, 100%) for 15 s each. Using a laser-microdissection device (LMD6500 and DFC310 FX, Leica), tumour budding cells were carefully microdissected. The budding cells were pooled until a microdissected surface area of at least 186 789 μm^2 was reached. The cells from the tumour bulk were also microdissected to get a similar dissected area of budding and bulk cells. The identification of tumour budding cells was solely based on morphology and stromal admixture was avoided as much as possible. A mean surface area of 407 815 μm^2 of tumour bulk (range: 358 845–533 538 μm^2) and a mean surface area of 268 382 μm^2 of tumour budding (range: 186 789–414 521 μm^2) was dissected (Figure 1C).

RNA isolation. Total RNA was extracted from the microdissected main tumour and budding samples using the RNeasy Plus Micro Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's protocol with additional DNase digestion. Concentration and RNA integrity were determined by running the extracted RNA on a BioAnalyzer 2100 instrument (Agilent Technologies, Santa Clara, CA, USA) using a picochip under Eukaryote total RNA settings. Total yield ranged from 4088 to 17 780 pg with mean yield of 9310 pg for the main tumour and a range from 826 to 4326 pg with a mean yield of 2170 pg for tumour budding samples.

RNA sequencing, analysis and annotation. RNA sequencing was performed on eight budding and their eight matched tumour bulk samples using the TruSeq RNA access Library Prep Kit (Illumina) following the manufacturer's instructions. In short, RNA was fragmented using divalent cations at 94 °C for 8 min. The cleaved RNA fragments were then converted to cDNA using random priming during first and second strand synthesis. An A-base was added to the resulting double-stranded cDNA fragments and sequencing adaptors were ligated. The resulting library was amplified using PCR. Subsequently, two rounds of 90 min hybridisation at 58 °C with sequence-specific probes were used to
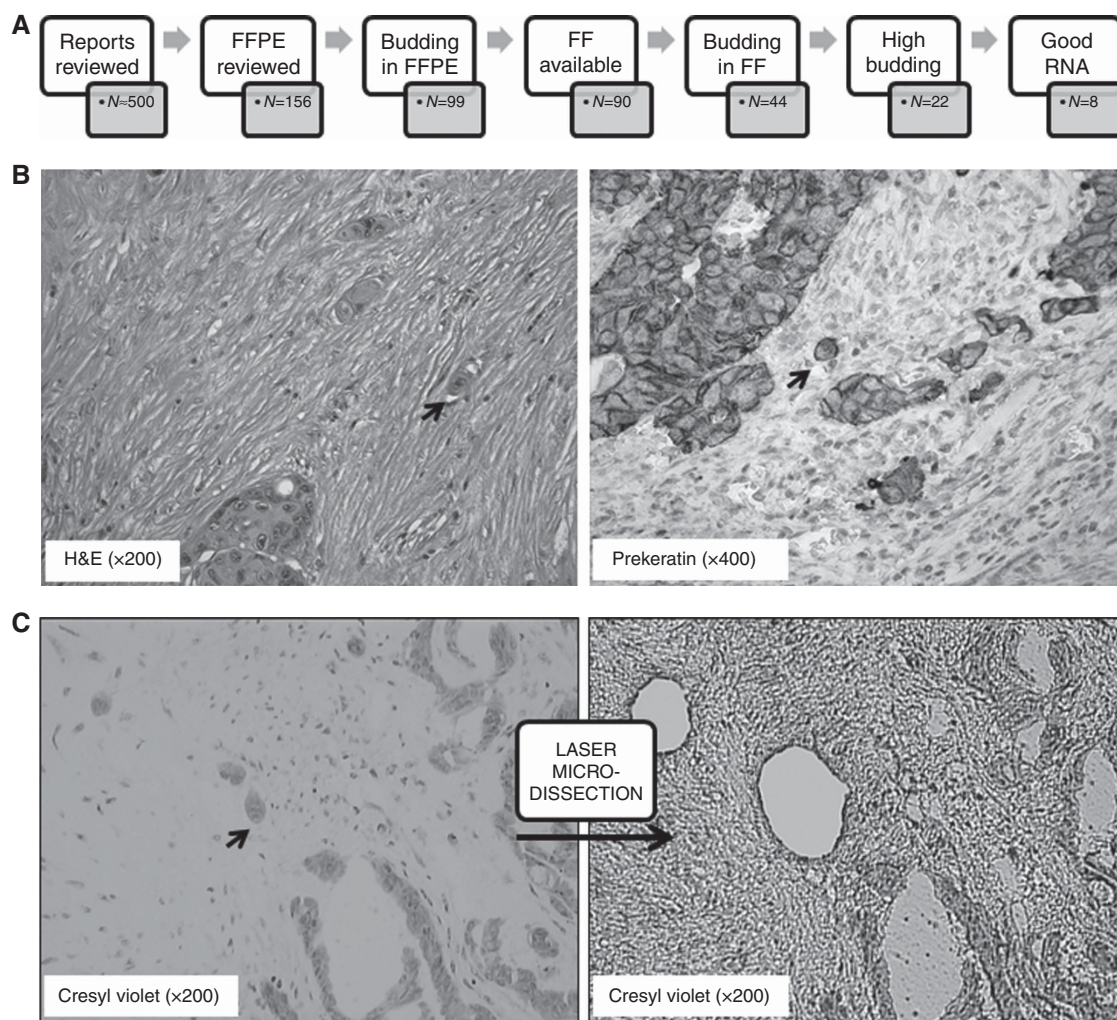
Figure 1. Sample selection and processing. (A) Flow chart of selection process. (B) H&E (left) and prekeratin (right) stained section, tumour budding is indicated with a black arrow. (C) Fresh frozen slide stained with cresyl violet before (left) and after (right) laser microdissection; tumour budding is indicated with a black arrow. A full colour version of this figure is available at the *British Journal of Cancer* journal online.

capture the coding regions of the transcriptome and bind these to streptavidin magnetic beads. Two heated washes were performed to remove non-specifically bound fragments from the beads after which the captured fragments were again enriched with PCR. The resulting libraries were quantified with a qPCR using the KAPA Library Quantification for Illumina (Kapa Biosystems Inc., Wilmington, MA, USA) and sequenced on a HiSeq2500 (Illumina) using a V4 flowcell generating 1×50 bp reads. Sequencing reads were processed and analysed as previously described (Nassar *et al*, 2015). Briefly, raw reads were mapped after adaptor-trimming to the human transcriptome and reference genome (GRCh37.65/ hg19) using TopHat 2.0 (Kim *et al*, 2013) and Bowtie2.0. Reads were assigned to ensemble gene IDs with the HTSeq software package (Risso *et al*, 2011; Langmead and Salzberg, 2012; Kim *et al*, 2013; Nassar *et al*, 2015). Two samples (main tumour six and budding sample three) with less than 10 million reads were excluded. The reads of the other samples were normalised with EDASeq (Risso *et al*, 2011) resulting in an average of 13 785 080 counts per sample. Differentially expressed genes between budding and bulk were identified with the DESeq software package (Anders and Huber, 2010). Raw sequencing reads are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-4065.

Pathway analysis. All differentially expressed genes (FDR < 0.05 and at least a twofold change) were analysed through the use of

QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity) to identify the relevant canonical pathways, upstream regulators, networks and molecular and cellular functions. Upstream regulators were also identified using i-*cis* Target (www.gbiomed.kuleuven.be/apps/lcb/i-cisTarget; Herrmann *et al*, 2012; Imrichová *et al*, 2015).

Clustering and CMS subtyping. Gene clustering was performed in R using the ward.D2 clustering algorithm on the logarithmic transformed reads per kilobase per million mapped reads (RPKM) normalised reads. Cluster analyses were performed using correlation distance metrics and average linkage agglomeration algorithm and visualised as heatmaps (R package nclust version 1.9.0, http:// bcf.isb-sib.ch/Resources.html). Genes with a zero variance across the data set were removed from the main clustering heatmap. The thousand most differentially expressed genes between main and buds were selected to be plotted in the heatmap of Figure 2B. The CMS calls of our samples were determined using the single sample options and the nearest neighbours classifier, which is part of the CMS-classifier package (Guinney *et al*, 2015). For the CMS heatmap, CMS-specific genes were first identified by linear regression using the limma R package by assessing one CMS *vs* all the others and by selecting the twenty most-associated genes with positive t-statistics (to select genes that are more expressed in each CMS) in three of the data sets (Budinska *et al*, Marisa *et al* and TCGA) published in Guinney *et al* (2015).
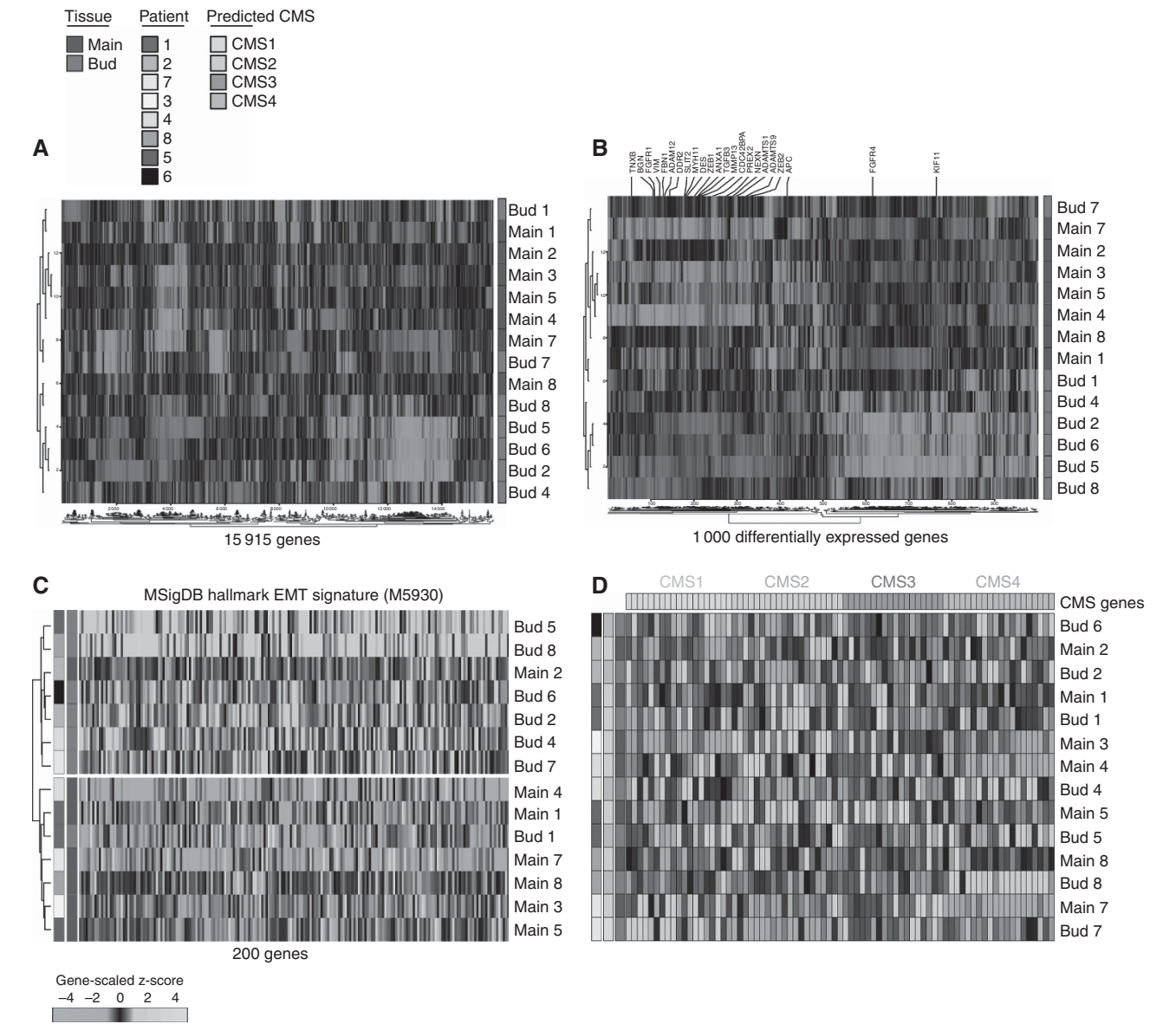
Figure 2. Gene expression profiles of tumour bulk and budding areas plotted on a heatmap. (A) Tumour bulk and budding profiles were roughly clustered into two groups. (B) Heatmap of the 1000 most differentially expressed genes. (C) Tumour budding areas were clustered as EMT-postive and tumour bulk samples as EMT-negative when plotted according to an established EMT gene signature taken from the MSigDB database. (D) A shift from CMS2 (epithelial) to CMS4 (mesenchymal) was seen when tumour cells transited from bulk to budding regions for the majority of the samples. A full colour version of this figure is available at the *British Journal of Cancer* journal online.

The heatmap was drawn using the pheatmap R package using the CMS-specific genes.

RESULTS

Tumour characteristics. Out of a series of 156 CRCs that were reviewed, only eight were withheld for this study as only these met our inclusion criteria as described above (Figure 1A). From the eight cases selected, five originated from the left colon, three from the right colon. In six cases, tumour-positive lymph nodes were present, whereas five cases presented with haematogenous metastasis at the time of diagnosis. In six cases, the tumours were characterised by microsatellite stability while the two other cases, clinically both presenting in the right colon and without metastasis, were microsatellite unstable (MSI; Table 1).

Expression profiles of tumour buds and main tumour. Transcriptome analysis by RNA-seq was performed on eight samples containing tumour budding cells and eight samples with cells from the corresponding tumour bulk. The novelty of this study design relies on the fact that the gene expression profile of the tumour bulk was compared with the gene expression profile of single detaching cells (= tumour budding cells) at the invasive border of the tumour. For two samples (main six and budding three), there were too few sequencing reads and therefore these samples were excluded from further analysis. Differential gene expression analysis comparing all tumour bulk samples with all tumour budding samples revealed that 296 genes, each of which mapped to an official Human Gene Nomenclature gene ID, were differentially expressed between tumour bulk and tumour budding samples after applying an FDR-correction ($q < 0.05$). Of these, 193 genes were characterised by increased expression in the tumour budding cells

Table 1. Tumour characteristics

Sample number	Tumour location	Age	Gender	Degree of differentiation	TNM	MSI/MSS	KRAS/NRAS/BRAF
1	Sigmoid	70	M	Moderate–Poor	pT3aN1aM1	MSS	WT
2	Sigmoid	61	F	Moderate	pT3N1bM1	MSS	KRAS p.G13D
3	Sigmoid	60	M	Moderate–Poor	pT3N2bM1	MSS	WT
4	Colon ascendens	85	F	Moderate	pT4aN2bM1	MSS	WT
5	Sigmoid	57	M	Moderate	pT3N1bM1	MSS	WT
6	Rectum	64	F	Moderate	pT4aN0M0	MSS	WT
7	Caecum	32	M	Poor	pT3N0M0	MSI	BRAFV600E
8	Colon ascendens	61	M	Poor	pT4aN1bM0	MSI	BRAFV600E

Table 2. Top 10 upregulated and downregulated genes

Up	Function	FC (FDR)	Down	Function	FC (FDR)
GSTM2	Detoxification of toxins and products of oxidative stress	3.739 ($q = 0.0228$)	CELP	—	-6.476 ($q = 0.0029$)
CYP1B1	Catalysation of metabolites	3.535 ($q = 0.0002$)	GRAMD2	—	-5.865 ($q = 0.0025$)
ANXA1	Inhibits activation of NFKB	3.489 ($q = 0.0074$)	CHP2	Regulation of cell proliferation and tumour growth	-5.160 ($q = 0.0086$)
ABI3BP	Glycosaminoglycan binding	2.561 ($q = 0.0037$)	CDHR1	Calcium-dependent cell adhesion molecule	-5.123 ($q = 0.0020$)
DES	Regulation of sarcomere composition	3.311 ($q = 0.0005$)	LRRC31	Regulates apoptosis and NFKB activation	-5.104 ($q = 0.0006$)
ABCA6	Regulates ATPase activity	3.240 ($q = 0.0101$)	ISM2	Involved in complement pathway	-4.753 ($q = 0.0375$)
NAV3	Contributes to growth, differentiation and apoptosis	3.209 ($q = 0.0351$)	KLHL32	Involved in protin ubiquitination	-4.668 ($q = 0.0422$)
ADAMTS1	Interacts with VEGFA	1.839 ($q = 0.0438$)	SLC19A3	Transmembrane receptor	-4.495 ($q = 0.0289$)
ADAMTS9	Organ shape during development and angiogenesis	2.255 ($q = 0.0098$)	C10orf99	—	-4.442 ($q < 0.0001$)
HDAC9	Regulates histone deacetylation	3.061 ($q = 0.0385$)	NXPE4	—	-4.372 ($q = 0.0438$)

compared with the tumour bulk. Conversely, 103 genes were unique for the tumour bulk signature. An overview of the top 10 genes that were most differentially expressed can be found in Table 2 (Supplementary Table 1 summarises the differential expression profile). Unsupervised clustering demonstrated that tumour bulk cells and tumour budding cells cluster separately in two groups that roughly correspond to tumour bulk *vs* tumour budding, as shown in Figure 2A and B. It has to be emphasised that matched pairs of tumour bulk and budding cells did not cluster together due to the strong differences between tumour budding compared with tumour bulk. Furthermore, clustering based on an established EMT gene signature taken from the MSigDB database revealed that six out of seven budding cells were classified as highly expressing EMT genes, whereas most bulk samples clustered as low EMT gene expression (Figure 2C). In samples containing tumour budding cells, following markers related to the process of EMT were higher expressed compared with cells derived from the tumour bulk: *SLIT2* (Log2FC = 2.819, $q < 0.001$), *ZEB1* (Log2FC = 1.962, $q < 0.001$), *ZEB2* (Log2FC = 2.022, $q < 0.001$), *DES* (Log2FC = 1.422, $q = 0.0212$) and *VIM* (Log2FC = 3.311, $q < 0.001$). *TGFβ3* (Log2FC = 2.389, $q = 0.004$) and *APC* (Log2FC = 1.483, $q = 0.010$) were upregulated in tumour budding together with a shift from *FGFR4* (Log2FC = -1.340, $q = 0.040$) to *FGFR1* (Log2FC = 1.578, $q = 0.010$).

Thorough investigation of the differentially expression profile identified *PREX2*, an interesting marker upregulated in the tumour budding cells that is known to be involved in cellular migration. Furthermore, significant loss of histone variant 2A family member X (H2A.X) expression was observed in the budding cells as compared with the tumour bulk.

Pathway analysis. Next, we applied IPA to the 296 differentially expressed genes. The biological functions and processes that were identified involved specific molecular and cellular functions such as cell movement, cell death, cell survival, cell growth, cell proliferation, cellular assembly, cellular organisation, cell function and cell maintenance. When restricting the analysis to genes significantly upregulated in budding areas, enrichment was seen in processes such as cellular movement, cell morphology and cell survival. Genes downregulated in budding areas were significantly involved in cell growth and proliferation. Subsequent analysis for enriched pathways showed that differentially expressed genes between the tumour bulk and budding cells are involved in the following canonical pathways: hepatic fibrosis ($P < 0.001$), actin cytoskeleton remodelling ($P = 0.001$), epithelial adherens junction signalling ($P = 0.001$), axonal guidance signalling ($P = 0.001$) and signalling through the family of Rho GTPases ($P = 0.002$). Furthermore, IPA identified a set of upstream regulators that can explain the differential expression profile between cells derived from the tumour bulk and tumour budding cells. EMT-associated genes such as: *SMAD3, TGFB3, NF-KB, CTNNB1, TWIST1, TGFB1, JUNB, EGF* and *NOTCH4* were identified as potential upstream regulators. Interestingly, *FOXA2*, significantly downregulated in budding cells, and *RUNX2*, significantly higher expressed in

budding cells, were validated as possible upstream regulators of the differentially expressed profile. As they are differentially expressed between main tumour mass and budding regions, *RUNX2* and *FOXA2* are promising genes potentially involved in regulating EMT that need to be further investigated during tumour budding and early metastasis. Analysis via i-*cis* Target, that predicts regulatory features and *cis*-regulatory modules, indicated *POU2F1, JUND, FOSL2, SNAIL, CDX2, TWIST1, TWIST2, SNAIL, RUNX2, TEAD4, TCF4* and *TCF3* as potential upstream regulators of the differentially expressed profile.

Consensus molecular subtypes. Next, we applied the recently reported consensus molecular subtypes (CMS) to the gene signatures identified by RNA-seq in the tumour bulk and budding cells. This revealed that seven tumour bulk profiles were characterised by CMS2, whereas on the other hand the five budding cell profiles closely matched the CMS4 subtype (Figure 2D). For one sample, the tumour budding signature was classified as CMS1, which is the molecular subtype that is known to be enriched for MSI-positive tumours (Guinney *et al*). Notably, this tumour was indeed diagnosed as an MSI-positive tumour by immunohistochemistry and genetic analysis according to the Bethesda guidelines (instability was observed in six out of eight repeats and hypermethylation was observed).

DISCUSSION

Tumour budding, defined as single cells or small groups of up to five tumour cells at the invasive border of the tumour, is a well-documented histological phenomenon in CRC and is considered to be the first step in cancer metastasis (Ueno *et al*, 2002). However, its biological mechanisms and pathways remain mainly unknown. In the present study, tumour budding areas in CRC were carefully microdissected, avoiding stromal admixture as much as possible, in pair with matched tumour bulk cells followed by RNA-sequencing. Previous gene expression studies on the tissue level have shown that the tumour invasion front showed higher expression of Wnt/ β-catenin target genes and upregulation of NF-κB target genes as compared with the tumour centre (Horst *et al*, 2009; Hlubek *et al*, 2007). We took it to the next level and studied the gene expression profile on the cellular level by microdissecting tumour budding and centre cells. As such, we demonstrated that the expression profiles of tumour bulk and budding cells in CRC are significantly different and that budding profiles cluster together separately from tumour bulk signatures. This unique signature demonstrates that tumour budding cells and tumour bulk cells have a different behaviour and that they are involved in distinct biological processes. Indeed, our data reveal that genes involved in cell growth and proliferation are more frequently expressed in the tumour bulk while genes involved in cell survival, morphology and migration are dominantly expressed in the budding cells. These results indicate that tumour budding cells undergo phenotype switching while detaching from the main tumour: budding cells tend to slow down their proliferation and acquire a more migratory phenotype with resistance to apoptosis. This feature of a low proliferative index in tumour budding cells is also described by Dawson (Galván *et al*, 2014; Dawson and Lugli, 2015).

It is generally hypothesised but not yet proven that tumour budding is the histologic reflection of EMT. The EMT is a process, identified in cell culture experiments, during which cells tend to lose their epithelial hallmarks and acquire a more mesenchymal phenotype allowing cells for migration and invasion (Zlobec and Lugli, 2010). This mechanism is already validated in animal models, but data in human samples remain sparse. The analysis of the differential gene expression pattern between bulk and buds with IPA demonstrated enrichment in the genetic footprint of

EMT. *ZEB1* and *ZEB2* are well known inducers of EMT and those genes were significantly upregulated in budding on RNA-level (Galván *et al*, 2014). Furthermore, a significant decrease in Histone 2A family member X (H2A.X) was observed in the tumour budding cells as compared with the tumour bulk. H2A.X is a histone variant known to be involved in DNA repair and cell growth, and loss of H2A.X was recently found to be an important regulator of EMT by inducing EMT transcription factors (Weyemi *et al*, 2016). Signalling through Wnt, TGFβ or receptor tyrosine kinases are described to affect ZEB1 and ZEB2 expression, which are important repressors of epithelial markers and inducers of mesenchymal proteins, in a positive manner. Active Wnt signalling in the tumour buds was demonstrated by increased expression of *APC*. *TGFβ3*, a member of the TGFβ family, showed increased expression in budding compared with tumour bulk. So far, there are only few evidences that TGFβ3 is associated with an EMT-like signature, but our data suggest that TGFβ3 is one of the main inducers of the TGFβ family for an EMT-like expression profile on RNA level (Xu *et al*, 2009). In 2015, Jensen *et al* identified TGFβ signalling as a key player in the mechanism of budding in oral squamous cell carcinoma (Jensen *et al*, 2015). Our study further supports the hypothesis of a close relation between tumour budding and EMT, because we found that budding regions showed an upregulation of genes involved in migration, cytoskeleton and extracellular matrix remodelling, all processes which are well known to have a key role during EMT (Lamouille *et al*, 2014). These events require a complex interplay between the extracellular matrix and the cytoskeleton mediated by proteins on the plasma membrane of the cell.

Regarding the predicted upstream regulators, some interesting and EMT-related markers came up: *TGFβ1, NFκB, TWIST1, SMAD3* and *CTNNB1*. β-catenin, encoded by the *CTNNB1* gene, had a predicted active state in the tumour budding. Therefore β-catenin should be relocated from cytoplasm to nucleus where it can fulfil its role as oncogenic transcription factor. Active signalling through the Wnt pathway is a possible initiator of this event (Xu *et al*, 2009; Lugli *et al*, 2012). It makes sense that expression levels of β-catenin are similar in tumour bulk and budding cells, because the function of β-catenin is in concordance with its cellular location. Furthermore, different groups already described nuclear β-catenin in tumour buds with cytoplasmic β-catenin in the rest of the tumour as shown by immunohistochemistry (García-Solano *et al*, 2012; Lamouille *et al*, 2014).

Owing to the high-throughput design of our study, we identified some potential additional markers, such as *FOXA2, RUNX2, TCF3, PREX2* and *H2A.X*, that may have a major role in the mechanism of early metastasis. Expression of *FOXA2* was significantly decreased in budding compared with bulk regions. FOXA2 is a transcriptional regulator of *ACTG2, CDX2, DDC, MAP1B, MYH11, PPARG, SCD* and *TAGLN*, genes which are differentially expressed between bulk and buds in our series. Furthermore, FOXA2 positively regulates cell–cell adhesion mediated by cadherins, explaining how loss of FOXA2 can induce detachment from the main tumour (Song *et al*, 2010). *RUNX2*, another transcription factor, was significantly upregulated in budding compared with bulk cells. Genes whose transcription is known to be regulated by RUNX2 – such as *BIRC5, CIIT1, DPYSL3, MMP134, MYH11, SERPINE1, SVIL, TAGLN* and *TLN* – were all differentially expressed in our budding cells compared with their matched tumour bulk. So far, RUNX2 is mainly described in osteoblast differentiation and bone metastasis, however, our data indicate an important role for RUNX2 in tumour budding and metastasis initiation (Cohen-Solal *et al*, 2015). PREX2 is an important regulator of cell migration and H2A.X is involved in the process of early metastasis in colorectal cell lines (Weyemi *et al*, 2016).

Remarkably, the classification of budding and bulk expression profiles revealed that five of the bulk profiles matched with CMS2, although the samples were randomly selected based on presence of prominent budding as demonstrated by microscopy. The CMS2 group contains 37% of the CRCs and is characterised by higher levels of copy number alterations in oncogenes and tumour-suppressor genes, without an increase in *KRAS* mutations (Guinney *et al*, 2015). Four budding samples were clustered in the CMS4 group, which accounts for 23% of all CRCs. CMS4 is characterised by high stromal infiltration, TGFβ activation and poor relapse-free and overall survival (Guinney *et al*, 2015). The expression profiles budding sample 7 (proven to be MSI by immunohistochemistry and genetic testing according to the Bethesda guidelines) was classified as CMS1. This CMS1 subgroup contains mostly MSI tumours characterised by hypermutation, BRAF mutation, immune infiltration and worse survival after relapse. Our results demonstrate that subtype switching from one molecular subtype to another is a common event in spatially defined regions in CRC and that tumour heterogeneity should be taken into account when classifying CRCs according to established molecular criteria. As the method of tissue sampling (e.g., size of tumour piece, spatial location within the tumour, tumour cell percentage, multiple sampling, etc.) has obviously high impact on molecular subtyping, it is advisable that new, yet to be determined standards should come in place to integrate all variables when it comes to interpreting molecular results in CRC. In addition, further research is needed to clarify how this molecular subtype switching is induced: gain of tumour function or gain of function of the tumour microenvironment (Isella *et al*, 2015).

In conclusion, gene expression profiles of spatially defined bulk and budding regions are significantly different indicating that budding and main tumour cells are involved in distinct biological processes. Moreover, the expression profile of dissected budding areas reflects an EMT-like signature with activation of both TGFβ and Wnt signalling. Therefore, we validate for the first time that tumour budding areas reflect an EMT-like signature on RNA level in human CRC tissue. The function of FOXA2, RUNX2, TEAD4 and TCF4 should be investigated in the process of early metastasis as they reflect important regulators of migratory and invasive dynamics on the RNA level. Finally, we are the first to suggest molecular subtype switching from CMS2 to CMS4 in bulk to budding regions in CRC indicating that the molecular background is not constant throughout the tumour and the location of the analysed area within the tumour highly influences the molecular subtyping.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Betge J, Kornprat P, Pollheimer MJ, Lindtner RA, Schlemmer A, Rehak P, Vieth M, Langner C (2012) Tumor budding is an independent predictor of outcome in AJCC/UICC stage II colorectal cancer. *Ann Surg Oncol* **19**: 3706–3712.

Bhangu A, Wood G, Mirnezami A, Darzi A, Tekkis P, Goldin R (2012) Epithelial mesenchymal transition in colorectal cancer: seminal role in promoting disease progression and resistance to neoadjuvant therapy. *Surg Oncol* **21**: 316–323.

Cohen-Solal KA, Boregowda RK, Lasfar A (2015) RUNX2 and the PI3K/AKT axis reciprocal activation as a driving force for tumor progression. *Mol Cancer* **14**: 137.

Dawson H, Lugli A (2015) Molecular and pathogenetic aspects of tumor budding in colorectal cancer. *Front Med (Lausanne)* **2**: 11.

Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* **17**: 1471–1474.

Galván JA, Astudillo A, Vallina A, Crespo G, Folgueras MV, González MV (2014) Prognostic and diagnostic value of epithelial to mesenchymal transition markers in pulmonary neuroendocrine tumors. *BMC Cancer* **14**: 855.

García-Solano J, Conesa-Zamora P, Trujillo-Santos J, Torres-Moreno D, Mäkinen MJ, Pérez-Guillermo M (2012) Immunohistochemical expression profile of β-catenin, E-cadherin, P-cadherin, laminin-5γ2 chain, and SMAD4 in colorectal serrated adenocarcinoma. *Hum Pathol* **43**: 1094–1102.

Guinney J, Dienstmann R, Wang X, De Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S (2015) The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**(11): 1350–1356.

Ha SS, Choi HJ, Park KJ, Kim JM, Kim SH, Roh YH, Kwon HC, Roh MS (2005) Intensity of tumor budding as an index for the malignant potential in invasive rectal carcinoma. *Cancer Res Treat* **37**: 177–182.

Hase K, Shatney C, Johnson D, Trollope M, Vierra M (1993) Prognostic value of tumor 'budding' in patients with colorectal cancer. *Dis Colon Rectum* **36**: 627–635.

Herrmann C, Van de Sande B, Potier D, Aerts S (2012) i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res* **40**: e114.

Hlubek F, Brabletz T, Budczies J, Pfeiffer S, Jung A, Kirchner T (2007) Heterogeneous expression of Wnt/beta-catenin target genes within colorectal cancer. *Int J Cancer* **121**: 1941–1948.

Horcic M, Koelzer VH, Karamitopoulou E, Terracciano L, Puppa G, Zlobec I, Lugli A (2013) Tumor budding score based on 10 high-power fields is a promising basis for a standardized prognostic scoring system in stage II colorectal cancer. *Hum Pathol* **44**: 697–705.

Horst D, Budczies J, Brabletz T, Kirchner T, Hlubek F (2009) Invasion associated up-regulation of nuclear factor kappaB target genes in colorectal cancer. *Cancer* **115**: 4946–4958.

Imrichová H, Hulselmans G, Atak ZK, Potier D, Aerts S (2015) i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res* **43**: W57–W64.

Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, Mellano A, Senetta R, Cassenti A, Sonetto C, Inghirami G, Trusolino L, Fekete Z, De Ridder M, Cassoni P, Storme G, Bertotti A, Medico E (2015) Stromal contribution to the colorectal cancer transcriptome. *Nat Genet* **47**: 312–319.

Jensen DH, Dabelsteen E, Specht L, Fiehn A, Therkildsen MH, Jønson L, Vikesaa J, Nielsen FC, Von Buchwald C (2015) Molecular profiling of tumour budding implicates TGFβ-mediated epithelial-mesenchymal transition as a therapeutic target in oral squamous cell carcinoma. *J Pathol* **236**: 505–516.

Karamitopoulou E, Zlobec I, Kölzer V, Kondi-Pafiti A, Patsouris ES, Gennatas K, Lugli A (2013) Proposal for a 10-high-power-fields scoring method for the assessment of tumor budding in colorectal cancer. *Mod Pathol* **26**: 295–301.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Lamouille S, Xu J, Derynck R (2014) Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**: 178–196.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lugli A, Karamitopoulou E, Zlobec I (2012) Tumour budding: a promising parameter in colorectal cancer. *Br J Cancer* **106**: 1713–1717.

Nakamura T, Mitomi H, Kanazawa H, Ohkura Y, Watanabe M (2008) Tumor budding as an index to identify high-risk patients with stage II colon cancer. *Dis Colon Rectum* **51**: 568–572.

Nassar D, Latil M, Boeckx B, Lambrechts D, Blanpain C (2015) Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma. *Nat Med* **21**: 946–954.

Ohtsuki K, Koyama F, Tamura T, Enomoto Y, Fujii H, Mukogawa T, Nakagawa T, Uchimoto K, Nakamura S, Nonomura A, Nakajima Y (2008) Prognostic value of immunohistochemical analysis of tumor budding in colorectal carcinoma. *Anticancer Res* **28**: 1831–1836.

Okuyama T, Nakamura T, Yamaguchi M (2003) Budding is useful to select high-risk patients in stage II well-differentiated or moderately differentiated colon adenocarcinoma. *Dis Colon Rectum* **46**: 1400–1406.

Okuyama T, Oya M, Ishikawa H (2002) Budding as a risk factor for lymph node metastasis in pT1 or pT2 well-differentiated colorectal adenocarcinoma. *Dis Colon Rectum* **45**: 628–634.

Peinado H, Olmeda D, Cano A (2007) Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* **7**: 415–428.

Risso D, Schwartz K, Sherlock G, Dudoit S (2011) GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**: 480.

Roh MS, Lee JI, Choi PJ (2004) Tumor budding as a useful prognostic marker in esophageal squamous cell carcinoma. *Dis Esophagus* **17**: 333–337.

Song Y, Washington MK, Crawford HC (2010) Loss of FOXA1/2 is essential for the epithelial-to-mesenchymal transition in pancreatic cancer. *Cancer Res* **70**: 2115–2125.

Talbot LJ, Bhattacharya SD, Kuo PC (2012) Epithelial-mesenchymal transition, the tumor microenvironment, and metastatic behavior of epithelial malignancies. *Int J Biochem Mol Biol* **3**: 117–136.

Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, Thiery JP (2014) Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med* **6**: 1279–1293.

Ueno H, Murphy J, Jass JR, Mochizuki H, Talbot IC (2002) Tumour 'budding' as an index to estimate the potential of aggressiveness in rectal cancer. *Histopathology* **40**: 127–132.

Weyemi U, Redon CE, Choudhuri R, Aziz T, Maeda D, Boufraqech M, Parekh PR, Sethi TK, Kasoji M, Abrams N, Merchant A, Rajapakse VN, Bonner WM (2016) The histone variant H2A.X is a regulator of the epithelial-mesenchymal transition. *Nat Commun* **7**: 10711.

Wöhlke M, Schiffmann L, Prall F (2011) Aggressive colorectal carcinoma phenotypes of invasion can be assessed reproducibly and effectively predict poor survival: interobserver study and multivariate survival analysis of a prospectively collected series of 299 patients after potentially curative resections with long-term follow-up. *Histopathology* **59**: 857–866.

Xu J, Lamouille S, Derynck R (2009) TGF-beta-induced epithelial to mesenchymal transition. *Cell Res* **19**: 156–172.

Zlobec I, Lugli A (2010) Epithelial mesenchymal transition and tumor budding in aggressive colorectal cancer: tumor budding as oncotarget. *Oncotarget* **1**: 651–661.

Zlobec I, Lugli A, Baker K, Roth S, Minoo P, Hayashi S, Terracciano L, Jass JR (2007) Role of APAF-1, E-cadherin and peritumoral lymphocytic infiltration in tumour budding in colorectal cancer. *J Pathol* **212**: 260–268.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)