

## **The transition from primary colorectal cancer to isolated peritoneal malignancy is associated with a hypermutant, hypermethylated state**

Sally Hallam<sup>1</sup>, Joanne Stockton<sup>1</sup>, Claire Bryer<sup>1</sup>, Celina Whalley<sup>1</sup>, Valerie Pestinger<sup>1</sup>, Haney Youssef<sup>1</sup>, Andrew D Beggs<sup>1</sup>

1 = Surgical Research Laboratory, Institute of Cancer & Genomic Science, University of Birmingham, B15 2TT.

### **Correspondence to:**

Andrew Beggs, [a.beggs@bham.ac.uk](mailto:a.beggs@bham.ac.uk)

**KEYWORDS:** Colorectal cancer, peritoneal metastasis

### **ABBREVIATIONS:**

Colorectal cancer (CRC), Colorectal peritoneal metastasis (CPM), Cytoreductive surgery and heated intraperitoneal chemotherapy (CRS & HIPEC), Disease free survival (DFS), Differentially methylated regions (DMR), Overall survival (OS), TableFormalin fixed paraffin embedded (FFPE), Hepatocellular carcinoma (HCC)

**ARTICLE CATEGORY:** Research article

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## **NOVELTY AND IMPACT:**

Colorectal peritoneal metastasis (CPM) are associated with limited and variable survival despite patient selection using known prognostic factors and optimal currently available treatments. There is a paucity of research concerning the biology of CPM. This study describes the biological landscape of CPM and the molecular features associated with CPM development, conferring poor prognosis and has identified that the majority of CPM develop a hypermutant phenotype that may be suitable for treatment with anti-PD1/CTLA4 immunotherapy.

## ABSTRACT

Colorectal Peritoneal metastases (CPM) develop in 15% of colorectal cancers. Cytoreductive surgery and heated intraperitoneal chemotherapy (CRS & HIPEC) is the current standard of care in selected patients with limited resectable CPM. Despite selection using known prognostic factors survival is varied and morbidity and mortality are relatively high. There is a need to improve patient selection and a paucity of research concerning the biology of isolated CPM. We aimed to determine the biology associated with transition from primary CRC to CPM and poor prognosis CPM, to identify those suitable for treatment with CRS & HIPEC and to identify targets for existing repurposed or novel treatment strategies. A cohort of patients with CPM treated with CRS & HIPEC was recruited and divided according to prognosis. Molecular profiling of the transcriptome, epigenome and genome of CPM and matched primary CRC was performed.

CPM were characterised by frequent Wnt/  $\beta$  catenin negative regulator mutations, mismatch repair mutations and resulting high tumour mutational burden and dysregulation of methylation suggested by frequent TET2 mutations and mutations suggesting an immune evasive phenotype. Several novel therapies could be targeted to these frequent mutations including porcupine inhibitors, immune checkpoint inhibitors and methylation inhibitors. Here we show the molecular features associated with CPM development and with poor prognosis. Potential applications include improving patient selection for treatment and in the development of novel and personalised treatments.

## INTRODUCTION

Little is known about the biology of isolated colorectal peritoneal metastasis (CPM), which although a relatively rare phenomenon is one with a high mortality rate (1). Understanding tumour biology may identify which patients with primary colorectal cancer (CRC) are at risk of developing CPM, and which are suitable for treatment with cytoreductive surgery and heated intra-peritoneal chemotherapy (CRS & HIPEC). CRS & HIPEC (usually using an agent such as mitomycin C) aims to achieve macroscopic tumour resection with multiple visceral and peritoneal resections and ablation of microscopic disease. Five-year survival however varies widely, and morbidity and mortality are relatively high (1). There is a need therefore to improve patient selection, allowing alternative existing or novel treatment strategies to be used for patients unlikely to respond.

Primary CRC research has identified biological drivers of metastasis and markers of response to specific treatments, for example KRAS mutation in selection for anti-EGFR mAb therapy (2). Gene expression signatures can be used to predict oncological outcomes and treatment response, for example the CMS & CRIS classifiers of colorectal cancer biology (3, 4). Gene expression profiling in primary CRC has identified signatures associated with the development of metastasis (5). One small study combining a small number of CPM with a larger cohort of appendix adenocarcinoma identified a signature predictive of reduced overall survival (OS) following CRS & HIPEC; these are however two biologically distinct tumours, appendix having significantly improved prognosis (2).

The dysregulation of methylation is a key step in tumorigenesis CpG island promoter methylation (CIMP) appears to be stable between matched primary CRC and hepatic

metastasis suggesting an epigenetic methylation programme is established prior to the development of metastasis (6, 7) (8). The hypermethylation of *KRAS*, Wnt modulators, tumour suppressor genes, *CIMP* and hypomethylation of the oncogenes are associated with an unfavourable response to chemotherapy and anti-EGFR antibodies as well as tumour recurrence and reduced OS in primary and metastatic CRC (9-14). Chromosomal instability is ubiquitous in cancer, increased copy number alteration, indicative of chromosomal instability is found in metastatic CRC (15, 16). Lopez-Garcia et al (17) demonstrated that the evolution of chromosomal instability is depending on cellular tolerance, either via dysregulation of TP53 or via alternate escape mechanisms such as dysfunction of BCL9L regulated caspase signalling.

CRC metastatic drivers are less clearly defined. Some studies found mutations exclusive to metastatic sites (18, 19), others found similar patterns of mutation between primary and metastasis (20). Studies have examined the somatic mutations in CPM and their prognostic implications. These studies are limited to individual or small panels of mutations routinely tested for in clinical practice with limited evidence to suggest which genes should be included in panel sequencing in CPM. Schneider et al examined the *KRAS* and *BRAF* mutation status of patients with CPM who underwent CRS & HIPEC (21). They found mutations of *RAS/RAF* were associated with reduced OS independent of the use of targeted anti-EGFR treatment (21). Sasaki et al examined the *KRAS*, *BRAF* and *PIKCA* mutation status of patients with metastatic CRC, with or without CPM (22). They found the incidence of *BRAF* mutation was significantly associated with the presence of CPM but not with prognosis (22).

The landscape of metastatic colorectal cancer was studied by the MSK-IMPACT (23) group which undertook panel based sequencing of 1,134 metastatic colorectal

cancers. Of these 19 patients were defined as “peritoneal” malignancy which defined a wide range of anatomical sites (liver, ovary, abdomen) rather than pure peritoneal disease. These tumours were also not studied with matched primary tumour of origin, nor did they undergo comprehensive genomic characterisation, as early versions of the MSK panel were insufficient in size to call tumour mutational burden.

There is a need to improve the outcomes for patients with CPM and significant variation in survival despite patient selection for treatment using known prognostic factors. There is a paucity of knowledge concerning CPM tumour biology. Understanding tumour biology will identify patients with primary CRC at risk of developing CPM, those suitable for treatment with CRS & HIPEC or alternative existing and novel treatment strategies. This study aims to determine the landscape of gene expression, methylation, and somatic mutation profile associated with the transition from primary CRC to isolated CPM and determine the association between these and prognosis following CRS & HIPEC in order to identify therapeutic targets.

## **MATERIALS AND METHODS**

### **Patient cohorts**

This study obtained ethical approval from the North West Haydock Research Ethics Committee, (15/NW/0079), project ID (17/283). Participants gave informed consent. Consecutive retrospective patients were recruited from an internally held database of all patients undergoing CRS & HIPEC at Good Hope hospital from 2011–2017. Patients with CPM (adenocarcinoma), no extra-abdominal metastasis, a CC0 resection and a PCI of < 12 were eligible for inclusion. Patients were divided into two prognostic groups. With palliative chemotherapy DFS is 11-13 months and therefore patients, post-treatment (CRS & HIPEC) with disease free survival (DFS) <12 months were defined as “poor prognosis” (24). Patients undergoing therapy with DFS > 12 months were defined as “good prognosis”. Demographic, tumour and treatment details were compared between the prognostic cohorts. For continuous variables, the students T-test was applied to normally distributed data and Mann Whitney-U to non-normally distributed data. Categorical variables were compared with the Chi-squared test or Fishers exact test. A p-value of < 0.5 was considered statistically significant. DFS survival between the good and poor prognosis cohorts was compared using the Kaplan Meier method. Statistical analysis was performed in IBM SPSS Statistics for Windows, Version 24.0 (25).

### **Nucleic acid extraction**

DNA and RNA were extracted from histologically confirmed Formalin fixed, paraffin embedded (FFPE) scrolls using the Covaris E220 evolution focused-ultrasonicator and the truTRAC® FFPE total NA Kit. Nucleic acid concentration was quantified using the Qubit™ 3.0 Fluorometer and Qubit™ RNA / DNA HS (high sensitivity) assay kit.

Nucleic acid quality was measured by electrophoresis using the Agilent 2200 TapeStation Nucleic Acid System, Agilent 2200 TapeStation Software A.01.05 and the Agilent High Sensitivity RNA / DNA ScreenTape and reagents.

### **RNA library preparation, sequencing and bioinformatics**

RNA library preparation was performed using the Lexogen Quant Seq 3' mRNA-Seq Library Prep kit. RNA libraries were denatured, diluted, loaded onto a 75-cycle High output flow cell and sequenced using the NextSeq500 at 2.5-5 million reads (26).

Quality control, trimming and alignment to the reference genome, (NCBI build 37, hg19) was performed with the Partek® Flow® genomics suite software package (Partek, St Louis, MI, USA). The gene expression profiles of primary and good and poor prognosis cohorts were using gene Specific Analysis (GSA) Modelling using the Partek® flow® with a false discovery rate (FDR) of <0.1.

CMS and CRIS classifications were performed using 'CMScaller' (v0.99.1) in the R package, version 2.10.2. (27) (28) Fishers exact test was used to compare contingency between primary and CPM and good and poor prognosis CPM test in IBM SPSS Statistics for Windows, Version 24.0. (25) A p value of < 0.05 was considered significant.

### **Methylation array and bioinformatics**

DNA was treated with sodium bisulphite using the Zymo EZ-DNA methylation kit, according to manufacturer's instructions. Degraded FFPE DNA was restored prior to methylation array with the Infinium HD FFPE restore kit, according to manufacturer's instructions. Methylation array was performed according to the Infinium MethylationEPIC BeadChip Kit manufacturer's instructions. BeadChips were imaged

using the Illumina iScan system. Initial data quality was checked using GenomeStudio Methylation Module Software.

Raw data was loaded into the RStudio version 3.5.0 software using the minifi package. Bioinformatics analysis was performed using the Chip Analysis Methylation Pipeline (ChAMP) R package, version 2.10.2.(29, 30) Probes with signals from less than three functional beads, low confidence with a detection p-value  $>0.01$ , covering SNPs, non-CpG and those located on the X and Y chromosome where filtered. Beta-mixture quantile normalization (BMIQ) was applied and a singular value decomposition (SVD) performed to identify batch effects. The association between methylation and prognosis was determined using the Bioconductor R package limma and bumphunter functions. Copy number alteration calling was performed using the CHAMP CNA function with a significance threshold of, p-value  $< p < x10^{-10}$ .

### **Exome capture, high-throughput sequencing and bioinformatics**

DNA was sheared using the Covaris E220 evolution focused-ultrasonicator to produce a uniform 150bp fragment size. Libraries were prepared using the TruSeq® Exome Kit then denatured, diluted, loaded onto a 150-cycle High output flow cell and sequenced using the NextSeq500.

Sequencing reads were assessed using FastQC. Sequences with a Phred score of  $<30$  were removed giving a base call accuracy of 99.9%. Sequence reads were aligned to the human reference genome, (hg19) using the Burrows-Wheeler Aligner (BWA) package (31). SAMTools was used to generate chromosomal coordinate-sorted BAM files and Picard was used to remove PCR duplicates (32). Somatic

variants were called from matched tumour-normal samples using Strelka2 in tumour/normal mode (33). Somatic variants were viewed, filtered and annotated in genomics workbench (34). Mutations with a MAF of >1% in known variant databases, (dbSNP and 100,000 genomes) were filtered. Mutations were annotated with information from known variant databases, (dbSNP and 100,000 genomes), PhastCons score and functional consequences. The prognostic groups were compared using Fischer exact test to identify potential candidate driver mutations for poor prognosis CPM. Somatic mutations were entered into the IntOGen platform for further analysis (35). The IntOGen-mutation platform incorporates a number of pipelines to identify cancer driver mutations and activated pathways (35). The OncodriveFM pipeline identifies mutations with a high functional impact using three scoring methods (Sorting Intolerant From Tolerant, (SIFT) (36), PolyPhen2 (37), and Mutation Assessor scores) (35, 38). and assess the likelihood that such mutations are cancer drivers. The OncodriveCLUST pipeline assesses the clustering of mutations to identify relevant activated pathways (35). MSI assessment was carried out using MSI\_classifier\_v3 ([https://rpubs.com/sigven/msi\\_classification\\_v3](https://rpubs.com/sigven/msi_classification_v3)).

## RESULTS

### Patient cohort

From 2011-2017 a total of n=161 patients underwent CRS & HIPEC at University Hospitals Birmingham, n=88 patients for metachronous CPM.

Patients were excluded for the following reasons: rectal primary n=5, incomplete CC2 resection n=8, peritoneal carcinomatosis index (PCI) of  $\geq 12$  n=20, follow up period of  $\leq 12$  months n=27 leaving n=28 patients. Complete information regarding the primary

CRC pathology and treatment was available for n=26 patients who form the basis of this study. Each patient had matched normal, primary CRC and CPM samples.

Thirteen patients had a DFS of 24 months (15–72 range) following CRS & HIPEC and formed the ‘good prognosis’ cohort, thirteen patients had a DFS of 6 months (2-11 range) and formed the poor prognosis cohort. There were no significant differences between cohorts in demographics, primary CRC or CPM tumour, treatment or follow up (table 1).

Following nucleic acid extraction all patients had adequate CPM RNA for RNAseq (n=13 good, n=13 poor prognosis), n=25 had matched primary CRC samples. For methylation array n=24 patients (n=12 good, n=12 poor prognosis) had adequate DNA. As the Infinium methylation array comprises a 32-prep kit, n=4 good and n=4 poor prognosis primary tumours were matched to these. For exome sequencing n=24 patients (n=12 good, n=12 poor prognosis) had adequate DNA from both the primary and CPM samples, extraction of DNA from normal tissue resulted in n=21 samples (n=9 good, n=12 poor prognosis).

## **Exome sequencing**

### **Primary tumours:**

Across all six sequencing runs, we obtained a median of 59.6X coverage (42.3-166.4) with a median uniformity of 88.1% (70.67-89.32). In the primary CRC cohort, after normal subtraction, there was a total of n=112,420 somatic SNV's, (primary-normal

subtraction). Variants present at >1% in the dbSNP common or 1000-Genomes databases were filtered leaving n=38,366 variants.. Seven (7/24, 30%) samples had a high tumour mutational burden, TMB  $\geq$  10 mut/Mb (39). Mutations were present in 64 of 95 known CRC driver genes (supplementary table S1) (40).

Somatic mutations were entered into the IntOGen pipeline to identify candidate driver mutations. OncodriveFM identified n=46 candidate driver genes with high levels of functional mutation (Q-value<0.05) in the primary CRC cohort (supplementary table S2). The top 10 potential driver genes were *KRAS* (68% samples), *TP53* (62% samples), *CAD* (40% samples), *APC* (35% samples), *NME1-NME2* (30% samples), *NMD3* and *SDHA* (31% samples), *CDC42BPA* (23% samples) and *ARHGEF12* (23% samples).

### **CPM:**

In the matched CPM cohort, a total of n=244,531 somatic SNV's were identified (CPM-primary subtraction). Nine samples, (56%) had a high tumour mutational burden TMB  $\geq$  10 mut/Mb (39). Mutations were identified in n=69 of n=95 known CRC driver genes, n=51 were shared between the primary and CPM, n=13 were novel (supplementary table S1) (40). Of the somatic variants identified in CPM, n=58,958 (29%) were present in the primary CRC, n=205,552 variants occurred exclusively in the CPM suggesting a significant accumulation of mutations in the transition to CPM (figure 1). OncodriveFM identified n=265 potential driver genes with high levels of functional mutation (Q-value<0.05) in the CPM cohort: *FLNB*, *SPTB*, *PPL*, *TP53*, *PDE4DIP*, *RIOK2*, *CDC16*, *NUP98*, *CDC16* and *SVEP1* (supplementary table S2), however these results must be treated with caution due to the bias of the hypermutator

phenotype. KEGG pathway analysis of mutations demonstrated enrichment in pathways concerning the immune system, signalling, metabolism and cancer (supplementary table S3).

Clonality analysis with SuperFreq showed distinct differences between the good and poor prognosis groups, with a median of 2 clones in the good prognosis group of primary tumours (range 1-4) and 3 clones in the poor prognosis group (range 2-7). In the peritoneal metastases there were a median of 3 clones in both the good (range 1-4) and poor (range 2-5) groups. Of note, in the poor prognosis group during clonal expansion, the dominant clone in the peritoneal metastasis group arose de-novo rather than being a prior clone that existed in the primary tumour (Supplementary figure 1).

In the primary tumours, 4/21 (19.0%) were MSS and 17/21 MSI (81.0%) whereas in the isolated peritoneal metastasis group, 9/21 were MSI (47.4%) and 10/21 were MSS (52.6%) demonstrating that there was a significantly higher rate of MSI in the isolated peritoneal metastasis group ( $p < 0.05$ , Chi<sup>2</sup>).

Patients with poor prognosis CPM had a higher frequency of somatic mutations: 60% of all mutations in CPM cohort vs. 40%. Patients with poor prognosis more commonly had a high tumour mutational burden, TMB  $\geq 10$  mut/Mb (39), 56% vs. 44%. Of the somatic mutations identified in poor prognosis CPM, n=35,461 (30%) were present in patients with good prognosis CPM, n=145,089 variants occurred exclusively in patients with poor prognosis CPM, suggesting a high tumour mutational burden was associated with the development of poor prognosis CPM (figure 1).

Comparison of somatic mutations in patients with good and poor prognosis identified two potential candidate genes, *FAM13A* and *PIEZO2* (Fishers exact  $p < 0.05$ , FDR=0.53) (table 2).

## Gene expression

### Differences between primary CRC and matched CPM:

Primary CRC and matched CPM showed differential expression of  $n=65$  genes with an FDR  $< 0.1$ . (Figure 2) Sixteen genes showed significantly decreased expression in CPM compared with primary CRC (table 3). Forty-nine genes showed significantly increased expression in CPM compared with primary CRC (table 3). A KEGG pathway analysis was performed to identify the enriched biological functions among the differentially expressed genes (table 4). The expression of *FABP6*, an intercellular bile acid transporter, was decreased 34.30-fold in CPM. *OLFM4* is a target of the Wnt/ $\beta$ -catenin pathway, its expression was reduced 3.77-fold in CPM. *DCN* and *PTEN* are able to initiate a number of signalling pathways including *ERK* and *EGFR* leading to growth suppression, their expression was increased 3.3-fold and 3.25 fold in CPM, this was unexpected and in contrast to the literature . (41). *NF- $\kappa$ B* expression was increased 3.24-fold in CPM, its upregulation may reflect increased *NF- $\kappa$ B* activity in the development of CPM (42).

## Differential gene expression profile Poor vs. Good prognosis CPM

One hundred and fifty-nine genes showed increased expression in poor prognosis CPM (table 5, figure 3). Five genes showed decreased expression in in poor prognosis CPM, however none had a fold change  $\geq 1.5$  suggesting minimal difference in expression between the good and poor prognosis cohorts (table 5). KEGG pathway analysis demonstrated enrichment in endocytosis, metabolism, phagocytosis, cell movement and architecture, bacterial and viral cell infection, transcription and the expression of genes controlling apoptosis, cell cycle, oxidative stress resistance and longevity (table 3). The expression of *CEACAM1*, a member of the carcinoembryonic antigen (*CEA*) immunoglobulin family, was increased 8.27-fold in poor prognosis CPM (43).

*AXIN1* encodes a cytoplasmic protein which forms the  $\beta$ -Catenin destruction complex, a negative regulator of the WNT signalling pathway (44). *AXIN1* expression was increased 5.42-fold in poor prognosis CPM which suggests increased activity of the Wnt signalling pathway (45) and may suggest sensitivity to porcupine inhibition.

Amongst the n=51 primary CRC and CPM samples n=35 were representative of each CMS subtype, the remaining n=16 samples did not have a consistent pattern (figure 4). Comparison of the CMS subtypes in primary and CPM and prognostic groups revealed an apparent transition from primary CRC to CPM. No primary CRC samples were classified as CMS4 (mesenchymal subtype characterized by prominent transforming growth factor activation, stromal invasion and angiogenesis) compared to 31% of CPM ( $p=0.085$ ). Secondly, patients with poor prognosis CPM were more commonly CMS4, 46% vs. 15% ( $p=0.005$ , table 6). Amongst the n=51 primary CRC

and CPM samples n=42 were representative of each CRIS subtype, the remaining nine samples did not have a consistent pattern (figure 5). There was no significant difference in CRIS classification between primary and CPM, p-value 0.365, or between good and poor prognosis CPM, p-value 0.148, (table 6). The non-concordance with CMS subtypes may reflect the increased stromal mesenchymal tissue present in patients with CPM and particularly those with CPM. As CRIS classification is based only on tumour cells such changes in intratumoural heterogeneity are not accounted for.

## **Methylation**

### **Methylation Microarray Analysis, primary CRC vs. matched metachronous**

#### **CPM**

Thirty-two samples in total were hybridised successfully to the Illumina HumanMethylation EPIC microarrays. DMPs were called between the primary CRC and CPM. The top ranked differentially methylated probe was cg04146982, BF 34.5, adjusted p-value  $5.67 \times 10^{-16}$  (chr8:144,943,810-144,943,810, hg19 coordinates), which tags a CpG dinucleotide 3651 bp upstream of the transcription start site of gene Epiplakin 1, (EPPK1) (46). EPPK1 is part of the Plakin family an important component of the cell cytoskeleton (47). The other DMP was cg12209861, BF 7.1, adjusted p-value 0.059 (chr4:37,459,078-37,459,078, hg19 coordinates), 3526 bp upstream of the transcription start site of gene Chromosome 4 Open Reading Frame 19, (*C4orf19*). DMRs were called between primary CRC and CPM via the dmrLasso function of the CHAMP pipeline (table 7). The top 10 most significant DMRs were in the region of *IGF2*, *ZNF461*, *RASGFR1*, *CELF4*, *ZSCAN18*, *EDNRB*, *ZBED9*, *VTRNA2-1*, *ZNF256*

and *EGFLAM*. KEGG pathway analysis did not reveal any significantly enriched pathways.

Comparison of CNA between primary and CPM via methylation arrays did not identify and significant differences in CNA between primary and CPM at a stringent p-value of  $< x10^{-10}$  however a number of CNA were identified at a lower significance threshold,  $p=2.78 \times 10^{-07}$  (table 8). Genes showing CNA gains of known significance in patients with CPM included; *TRIM3*, 5, 6, 21 and 22, *MT1A*, 2A, 3, 4 encode proteins of the metallothionein family.

### **Methylation Microarray Analysis, Poor vs. Good prognosis CPM**

The top ranked differentially methylated probe was cg07951355, BF=6, (chr1:40123717) which tags an intergenic region 1076 bp before gene *NT5C1A*. Cg25909064, BF 4 adjusted p-value 0.47 (chr11:120081487-120082345) which tags an intron of gene *OAF* and cg12977942, BF 4 adjusted p-value 0.47 (chr5:92839309-92839309) which tags an intron of gene *NR2F1-AS1* (46). Six significant DMRs (table 7) were identified in the regions of *NKX6-2*, *CHFR*, *GATA3*, *IRX5* and *HCK* and *BC019904*. KEGG pathway analysis did not reveal any significantly enriched pathways.

Comparison of CNA between the CPM prognostic groups identified recurrent gene losses at chromosomes 3, 4, 14, 15, 17 and 19 (table 8). CNA losses clustered in the RAS-MAPK-ERK signalling pathway suggesting dysregulation in poor prognosis CPM.

Comparison of CNA between the CPM prognostic groups identified n=19 gene gains at chromosomes 9, 10 and 11. Genes showing CNA gains in patients with poor prognosis CPM included: *SIT1*, *RNF38*, *MELK*, *PAX5*, *SHB*, *ZEB1*, *DEAF1*, *ANTXR*, *EPS8L2* and *PIDD1*

## DISCUSSION

This study determined the gene expression, CNA, methylation and somatic mutation profile of primary CRC and matched isolated CPM to determine whether there were a change associated with the development of CPM or predicting prognosis for patients with CPM. To our knowledge, this is the first such analysis in a cohort of patients with isolated CPM suitable for treatment with CRS & HIPEC. The MSKCC cohort of metastatic cancer (23) had a diverse range of metastatic cancer, none of whom overlapped with the type we have studied, which is isolated colorectal peritoneal metastasis, with matched primary samples, suitable for cytoreduction.

Comparison of patients with primary CRC and metachronous CPM identified biological changes associated with the transition from primary CRC to CPM. Hypermethylation, CNA and hypermutation resulted in the inactivation of tumour suppressors and oncogene activation in CPM, (TP53, VTRA2-1, TRIM proteins). These changes suggest a rapid rate of tumour growth unchecked by tumour suppressor or apoptotic mechanisms.

Increased MAPK and Wnt/ $\beta$ -catenin pathway activation was noted in CPM. Gene expression of negative regulators of the Wnt pathway was reduced, (OLFM4, DEFA6), negative Wnt regulators contained somatic mutations, (APC, RNF43, FAM123B and TSC1), and the MAPK marker, RASFGFR1 was hypermethylated

suggesting persistent activation of MAPK and Wnt pathways. Multiple mutations of negative Wnt signalling regulators make this an attractive therapeutic target. Porcupine inhibitors mediate the palmitoylation of Wnt ligands, blocking Wnt signalling. The porcupine inhibitor LGK974 inhibits the upstream negative Wnt regulator mutant RNF43 and is a potential therapeutic target in CPM (48).

CPM contained a high proportion of MSH6 somatic mutations suggesting deficiency in the mismatch repair pathway and MSI. MSH6 mutations are commonly found in isolated peritoneal metastasis (49). As expected for tumours with mismatch repair deficiency both the primary CRC and CPM cohort had a high tumour mutational burden, crucially this suggests they may have a good response to treatment with immune checkpoint inhibitors such as pembrolizumab (50), a new therapeutic avenue for these difficult to treat patients. The frequency of hypermutation seen in our study (48%) was considerably higher than that observed for both the MSKCC metastatic disease cohort (5%) and the TCGA Colorectal (51) cohort (10%). The expression of genes regulating innate immunity however was downregulated, (DEFA6, DMBT1, MUC2) or altered via somatic mutations, (HLA-A antigen) suggesting immune evasion in the transition to CPM which may reduce the likelihood of successful PD-1 therapy.

The expression of genes suppressing invasion, migration and EMT was downregulated or hypermethylated, (MUC2, MMP26, ILK, FLNB, SPTB, PPL, and SVEP1) and those triggering these processes upregulated, (CYR61, CXCL12, CTGF, CSTB). These changes suggest a mechanism by which CPM cells metastasise from the primary CRC. In keeping with changes in EMT regulators there appeared to be a transition in CMS subtypes towards CMS4 from primary CRC to CPM. The CMS4 subtype is an interesting therapeutic target, *TGF $\beta$*  signalling inhibitors and targeted immunotherapies have been trialled with success in pre-clinical models to block cross

talk between the tumour microenvironment and halt disease progression of stromal rich CMS4 CRC (52, 53).

Methylation appeared to be dysregulated in CPM with a bias towards a hypermethylator phenotype caused by somatic mutation of the TET2 tumour suppressor and CDH7 chromatin regulator. Active DNA demethylation by TET enzymes is an important tumour suppressor mechanism in a variety of cancers (54-56). Downregulation of CES2, a gene known to activate the prodrug irinotecan, a commonly used chemotherapy in the adjuvant treatment of primary CRC and CPM was seen in this cohort. Resistance to the treatment of primary CRC may in part explain the development of CPM.

*CEACAM1* expression correlates with metastasis and reduced survival in CRC and was upregulated in this cohort of patients (57). Novel therapies in the form of CEA TCB IgG-based T-cell bispecific antibodies (Cibisatamab) may therefore be of benefit (58). Additionally there was a downregulation of gene expression of negative regulators of the Wnt pathway, (AXIN1) and somatic mutations of key Wnt regulators, (FAM13A) and hypermethylation of MAPK and TGF- $\beta$  pathway markers, (RAB8A, RAB34, FGF5 and BMP3) suggesting persistent activation of MAPK, TGF- $\beta$  and Wnt in poor prognosis CPM.

A relative weakness of this study is the small cohort of patients, the biological changes identified here form a starting point in identifying the tumour biology associated with the development of CPM and predicting poor prognosis disease. However, we have identified multiple potential targets for therapy, along with the important finding that CPM appears to be a hypermutated, hypermethylated, immune evasive cancer which allows it to be potentially targeted by emerging novel therapeutics.

Patients with colorectal peritoneal metastasis (CPM) secondary to colorectal cancer have limited survival with the best available treatments. Despite selection for treatment using known prognostic factors survival varies widely and can be difficult to predict. There is a paucity of knowledge concerning the biology of CPM, it is likely that there are additional biological markers of response to currently available as well as novel or re-purposed alternative treatments. Here we have comprehensively profiled a cohort of patients with isolated CPM and identified a number of therapeutically targetable alterations including mutations in Wnt/ $\beta$  catenin regulators, the mismatch repair pathway and methylation regulators.

#### **DATA AVAILABILITY:**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Table 1: Comparison of good and poor prognosis patient cohorts**

		<b>Good Prognosis</b>	<b>Poor prognosis</b>	<b>P value</b>
<b>Age, mean +/- SD</b>		58 +/-13	58 +/-9	0.97
<b>Gender, male</b>		n=7 (54)	n=7 (54)	0.68
<b>Tumour location</b>	Right	n=9 (69)	n=6 (46)	0.33
	Transverse	n=1 (8)	n=0 (0)	
	Left	n=3 (23)	n=7 (54)	
<b>T stage primary</b>	3	n=3 (23)	n=3 (23)	0.66
	4a	n=5 (38.5)	n=7 (54)	
	4b	n=5 (38.5)	n=3 (23)	
<b>N stage primary</b>	0	n=4 (31)	n=1 (8)	0.86
	1	n=7 (54)	n=5 (38)	
	2	n=2 (15)	n=7 (54)	
<b>DFI months</b>		25 +/-9	24 +/-12	0.83
<b>PCI score, median (range)</b>		5 (3-12)	8 (2-12)	0.019
<b>CC score</b>	CC0	n=13 (100)	n=13 (100)	1
	CC1	n=0 (0)	n=0 (0)	

	CC2	n=0 (0)	n=0 (0)	
<b>Follow up, months, median (range)</b>		29 (19-72)	16 (5-55)	0.11
<b>Adjuvant treatment</b>	Yes	n=11 (85)	n=12 (92)	0.38
	No	n=2 (15)	n=1 (8)	
<b>DFS, median (range)</b>		24 (15-72)	6 (2-11)	<0.0001
<b>OS, median (range)</b>		29 (19-72)	16 (5-55)	0.12

(N= number value in parenthesis, percentage, DFI, disease free interval, PCI, peritoneal carcinomatosis index, CC score, completeness of cytoreduction, DFS, disease free survival, OS, overall survival)

Log rank  $p < 0.0001$ .

**Table 2 Potential candidate variants, poor prognosis CPM**

Chr	Position	Reference	Allele	P-value	FDR	Sample frequency (case)	Sample frequency (control)	Gene ID
4	93084410	C	G	0.007	0.53	62.5	0	FAM13A
18	11552313	G	C	0.023	0.53	50	0	PIEZO2

CPM identified through Fisher exact test, genomics workbench (Chr, chromosome, FDR, false discovery rate)

**Table 3 The top 10 genes with significantly altered expression (FDR<0.1) in CPM samples compared with primary CRC samples**

Rank	Gene name	Function	Fold change	FDR p value
<b>Reduced expression CPM samples vs. primary CRC</b>				
1	FABP6	Intracellular bile acid transporter	-34.30	1.74 x 10-06
2	DEFA6	Cytotoxic peptide involved in host intestine defence	-8.15	8.55 x 10-06
3	DMBT1	Tumour suppressor	-6.06	2.43 x 10-04
4	TTC38	Protein coding gene	-4.56	5.80 x 10-05
5	OLFM4	Wnt/ $\beta$ -catenin pathway target	-3.77	1.01 x 10-04
6	IGHA1	Immune receptor	-3.66	4.23 x 10-05
7	CES2	Intestinal enzyme controlling drug clearance	-3.20	6.84 x 10-05
8	NDUFS6	Enzyme in electron transport chain of mitochondria	-2.70	7.74 x 10-05
9	P2RY11	G-protein coupled receptor	-2.53	6.37 x 10-04
10	MUC2	Encodes a mucinous intestinal coating	-2.34	7.22 x 10-04

Increased expression CPM samples vs. primary CRC				
1	CD53	Tetraspanin	7.29	5.87 x 10 <sup>-05</sup>
2	CYR61	Extracellular signalling protein	4.24	3.12 x 10 <sup>-04</sup>
3	CXCL12	G-protein coupled receptor	3.64	9.25 x 10 <sup>-04</sup>
4	NR2F1	Nuclear hormone receptor and transcriptional regulator	3.53	7.09 x 10 <sup>-04</sup>
5	CTGF	Connective tissue growth factor	3.49	1.55 x 10 <sup>-04</sup>
6	CSTB	Cystatin	3.41	6.13 x 10 <sup>-04</sup>
7	TSC22D3	Anti-inflammatory protein glucocorticoid (GC)-induced leucine zipper	3.36	3.94 x 10 <sup>-04</sup>
8	DCN	Tumour suppressor gene	3.30	6.19 x 10 <sup>-05</sup>
9	PTEN	Tumour suppressor gene	3.25	9.28 x 10 <sup>-04</sup>
10	NF-κBIA	Inhibits the <i>NF-κB</i> transcription factor	3.24	1.06 x 10 <sup>-04</sup>

**Table 4 CMS and CRIS classification good vs. poor prognosis CPM**

	<b>Non-consensus</b>	<b>CMS1</b>	<b>CMS2</b>	<b>CMS3</b>	<b>CMS4</b>		<b>Total</b>
<b>Good prognosis CPM</b>	10 (77)	0 (0)	0 (0)	1 (8)	2 (15)		13
<b>Poor prognosis CPM</b>	2 (15)	1 (8)	3 (23)	1 (8)	6 (46)		13
	<b>Non-consensus</b>	<b>CRISA</b>	<b>CRISB</b>	<b>CRISC</b>	<b>CRISD</b>	<b>CRISE</b>	
<b>Good prognosis CPM</b>	5 (39)	2 (15)	2 (15)	3 (23)	0 (0)	1 (8)	13

<b>Poor prognosis CPM</b>	1 (8)	2 (15)	5 (39)	2 (15)	3 (23)	0 (0)	13
---------------------------	-------	--------	--------	--------	--------	-------	----

*CMS Fishers exact p-value 0.005, CRIS Fischer's exact p-value 0.148, values in parenthesis percentages*

## REFERENCES:

1. Koumpa FS, Xylas D, Konopka M, Galea D, Veselkov K, Antoniou A, et al. Colorectal Peritoneal Metastases: A Systematic Review of Current and Emerging Trends in Clinical and Translational Research. *Gastroenterol Res Pract*. 2019;2019:5180895.
2. Levine EA, Blazer DG, 3rd, Kim MK, Shen P, Stewart JHt, Guy C, et al. Gene expression profiling of peritoneal metastases from appendiceal and colon cancer demonstrates unique biologic signatures and predicts patient outcomes. *J Am Coll Surg*. 2012;214(4):599-606; discussion -7.
3. Isella C, Brundu F, Bellomo SE, Galimi F, Zanella E, Porporato R, et al. Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat Commun*. 2017;8:15107.
4. Lal N, White BS, Goussous G, Pickles O, Mason MJ, Beggs AD, et al. KRAS Mutation and Consensus Molecular Subtypes 2 and 3 Are Independently Associated with Reduced Immune Infiltration and Reactivity in Colorectal Cancer. *Clin Cancer Res*. 2018;24(1):224-33.
5. Friederichs J, Rosenberg R, Mages J, Janssen KP, Maeckl C, Nekarda H, et al. Gene expression profiles of different clinical stages of colorectal carcinoma: toward a molecular genetic understanding of tumor progression. *Int J Colorectal Dis*. 2005;20(5):391-402.
6. Ju HX, An B, Okamoto Y, Shinjo K, Kanemitsu Y, Komori K, et al. Distinct profiles of epigenetic evolution between colorectal cancers with and without metastasis. *Am J Pathol*. 2011;178(4):1835-46.

7. Konishi K, Watanabe Y, Shen L, Guo Y, Castoro RJ, Kondo K, et al. DNA methylation profiles of primary colorectal carcinoma and matched liver metastasis. *PLoS One*. 2011;6(11):e27889.
8. Murata A, Baba Y, Watanabe M, Shigaki H, Miyake K, Ishimoto T, et al. Methylation levels of LINE-1 in primary lesion and matched metastatic lesions of colorectal cancer. *Br J Cancer*. 2013;109(2):408-15.
9. Miranda E, Destro A, Malesci A, Balladore E, Bianchi P, Baryshnikova E, et al. Genetic and epigenetic changes in primary metastatic and nonmetastatic colorectal cancer. *Br J Cancer*. 2006;95(8):1101-7.
10. Sharrard RM, Royds JA, Rogers S, Shorthouse AJ. Patterns of methylation of the c-myc gene in human colorectal cancer progression. *Br J Cancer*. 1992;65(5):667-72.
11. Siemens H, Neumann J, Jackstadt R, Mansmann U, Horst D, Kirchner T, et al. Detection of miR-34a promoter methylation in combination with elevated expression of c-Met and  $\beta$ -catenin predicts distant metastasis of colon cancer. *Clin Cancer Res*. 2013;19(3):710-20.
12. Tang D, Liu J, Wang DR, Yu HF, Li YK, Zhang JQ. Diagnostic and prognostic value of the methylation status of secreted frizzled-related protein 2 in colorectal cancer. *Clin Invest Med*. 2011;34(2):E88-95.
13. Zhang X, Shimodaira H, Soeda H, Komine K, Takahashi H, Ouchi K, et al. CpG island methylator phenotype is associated with the efficacy of sequential oxaliplatin- and irinotecan-based chemotherapy and EGFR-related gene mutation in Japanese patients with metastatic colorectal cancer. *Int J Clin Oncol*. 2016;21(6):1091-101.

14. Kim SH, Park KH, Shin SJ, Lee KY, Kim TI, Kim NK, et al. CpG Island Methylator Phenotype and Methylation of Wnt Pathway Genes Together Predict Survival in Patients with Colorectal Cancer. *Yonsei Med J.* 2018;59(5):588-94.
15. Lee SY, Haq F, Kim D, Jun C, Jo HJ, Ahn SM, et al. Comparative genomic analysis of primary and synchronous metastatic colorectal cancers. *PLoS One.* 2014;9(3):e90459.
16. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell.* 2018;173(3):595-610 e11.
17. Lopez-Garcia C, Sansregret L, Domingo E, McGranahan N, Hobor S, Birkbak NJ, et al. BCL9L Dysfunction Impairs Caspase-2 Expression Permitting Aneuploidy Tolerance in Colorectal Cancer. *Cancer Cell.* 2017;31(1):79-93.
18. Goryca K, Kulecka M, Paziewska A, Dabrowska M, Grzelak M, Skrzypczak M, et al. Exome scale map of genetic alterations promoting metastasis in colorectal cancer. *BMC Genet.* 2018;19(1):85.
19. Huang D, Sun W, Zhou Y, Li P, Chen F, Chen H, et al. Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer Metastasis Rev.* 2018;37(1):173-87.
20. Mlecnik B, Bindea G, Kirilovsky A, Angell HK, Obenauf AC, Tosolini M, et al. The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. *Sci Transl Med.* 2016;8(327):327ra26.
21. Schneider MA, Eden J, Pache B, Laminger F, Lopez-Lopez V, Steffen T, et al. Mutations of RAS/RAF Proto-oncogenes Impair Survival After Cytoreductive Surgery and HIPEC for Peritoneal Metastasis of Colorectal Origin. *Ann Surg.* 2018;268(5):845-53.

22. Sasaki Y, Hamaguchi T, Yamada Y, Takahashi N, Shoji H, Honma Y, et al. Value of KRAS, BRAF, and PIK3CA Mutations and Survival Benefit from Systemic Chemotherapy in Colorectal Peritoneal Carcinomatosis. *Asian Pac J Cancer Prev*. 2016;17(2):539-43.
23. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, et al. Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell*. 2018;33(1):125-36 e3.
24. Verwaal VJ. Long-Term Results of Cytoreduction and HIPEC Followed by Systemic Chemotherapy. *Cancer Journal*. 2009;15(3):212-5.
25. Corp I. IBM SPSS Statistics for Windows, Version 24.0 Armonk, NY IBM Corp; Released 2016 [
26. Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*. 2014;30(3):301-4.
27. Rinaudo P, Boudah S, Junot C, Thévenot EA. biosigner: A New Method for the Discovery of Significant Molecular Signatures from Omics Data. *Front Mol Biosci*. 2016;3:26.
28. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep*. 2017;7(1):16618.
29. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics*. 2014;30(3):428-30.
30. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-9.

31. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-95.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
33. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-7.
34. Qiagen. **CLC Genomics Workbench 11.0** [Available from: <https://www.qiagenbioinformatics.com>].
35. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, et al. IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods*. 2013;10(11):1081-2.
36. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-4.
37. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
38. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011;39(17):e118.
39. Büttner R, Longshore JW, López-Ríos F, Merkelbach-Bruse S, Normanno N, Rouleau E, et al. Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open*. 2019;4(1):e000442.
40. IntOGen. IntOGen Colorectal cancer driver genes [Available from: <https://www.intogen.org/search>].

41. Copija A, Waniczek D, Walkiewicz k, Głogowski L, Augustyniak H, Nowakowska-Zajdel E. PTEN – clinical significance in colorectal cancer. *OncoReview*. 2016;6(2):A86-90.
42. Perkins N. The diverse and complex roles of NF- $\kappa$ B subunits in cancer. *Nature Reviews Cancer*. 2012;12(2):121-32.
43. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. Database (Oxford). 2010;2010:baq020.
44. Morin P, Sparks A, Korinek V, Barker N, Clevers H, Vogelstein B, et al. Activation of  $\beta$ -Catenin-Tcf Signaling in Colon Cancer by Mutations in  $\beta$ -Catenin or APC. *Science*. 1997;275(5307):1787-90.
45. Zhang L, Guo L, Peng Y, Chen B. Expression of T-STAR gene is associated with regulation of telomerase activity in human colon cancer cell line HCT-116. *World J Gastroenterol*. 2006;12(25):4056-60.
46. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
47. Hu L, Huang Z, Wu Z, Ali A, Qian A. Mammalian Plakins, Giant Cytolinkers: Versatile Biological Functions and Roles in Cancer. *Int J Mol Sci*. 2018;19(4).
48. Bahrami A, Amerizadeh F, ShahidSales S, Khazaei M, Ghayour-Mobarhan M, Sadeghnia HR, et al. Therapeutic Potential of Targeting Wnt/ $\beta$ -Catenin Pathway in Treatment of Colorectal Cancer: Rational and Progress. *J Cell Biochem*. 2017;118(8):1979-83.
49. Kim CG, Ahn JB, Jung M, Beom SH, Kim C, Kim JH, et al. Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers. *Br J Cancer*. 2016;115(1):25-33.

50. Schrock AB, Ouyang C, Sandhu J, Sokol E, Jin D, Ross JS, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol*. 2019.
51. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330-7.
52. Becht E, de Reyniès A, Giraldo NA, Pilati C, Buttard B, Lacroix L, et al. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clin Cancer Res*. 2016;22(16):4057-66.
53. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet*. 2015;47(4):320-9.
54. Tahara T, Yamamoto E, Madireddi P, Suzuki H, Maruyama R, Chung W, et al. Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology*. 2014;146(2):530-38.e5.
55. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3(6):415-28.
56. Carter CA, Oronsky BT, Roswarski J, Oronsky AL, Oronsky N, Scicinski J, et al. No patient left behind: The promise of immune priming with epigenetic agents. *Oncoimmunology*. 2017;6(10):e1315486.
57. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev*. 2013;32(3-4):643-71.

58. Bacac M, Fauti T, Sam J, Colombetti S, Weinzierl T, Ouaret D, et al. A Novel Carcinoembryonic Antigen T-Cell Bispecific Antibody (CEA TCB) for the Treatment of Solid Tumors. Clin Cancer Res. 2016;22(13):3286-97.

## FIGURES

Figure 1: Mutations exclusive to and shared between primary CRC and matched CPM and good and poor prognosis CPM

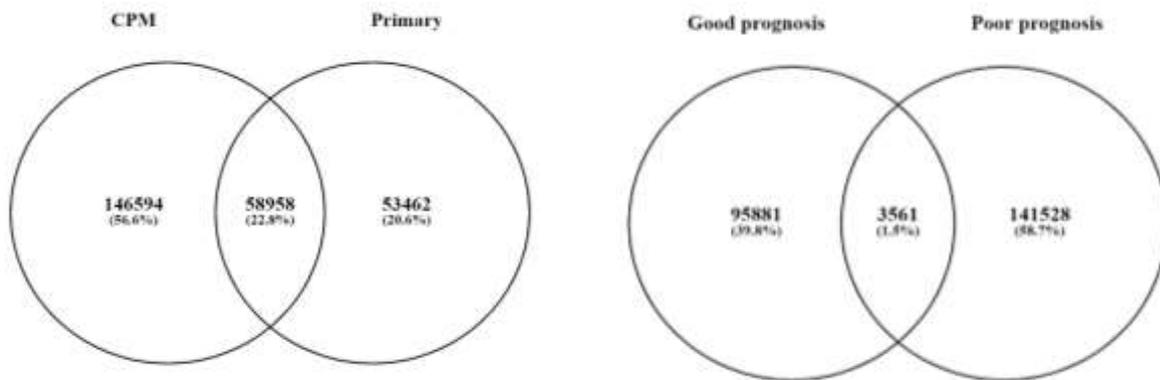


Figure 2: Heatmap of differential gene expression between CPM vs. primary samples. Sample type is indicated at the upper border of the heatmap. Blue CRS = CPM; P red=primary

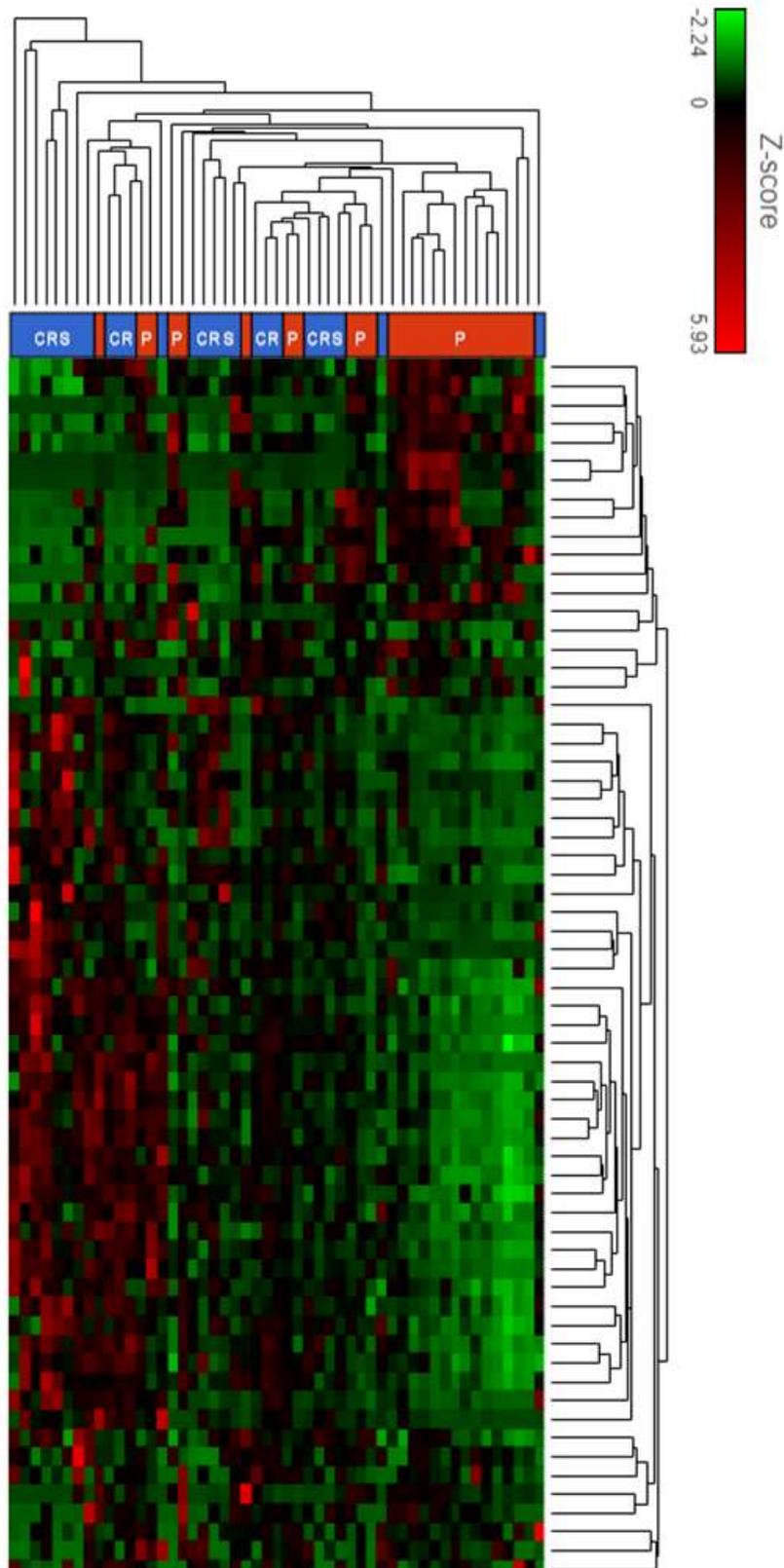


Figure 3: Heatmap differential gene expression good vs. poor CPM. Sample type is indicated at top of heatmap. Poor = poor prognosis; Good = good prognosis.

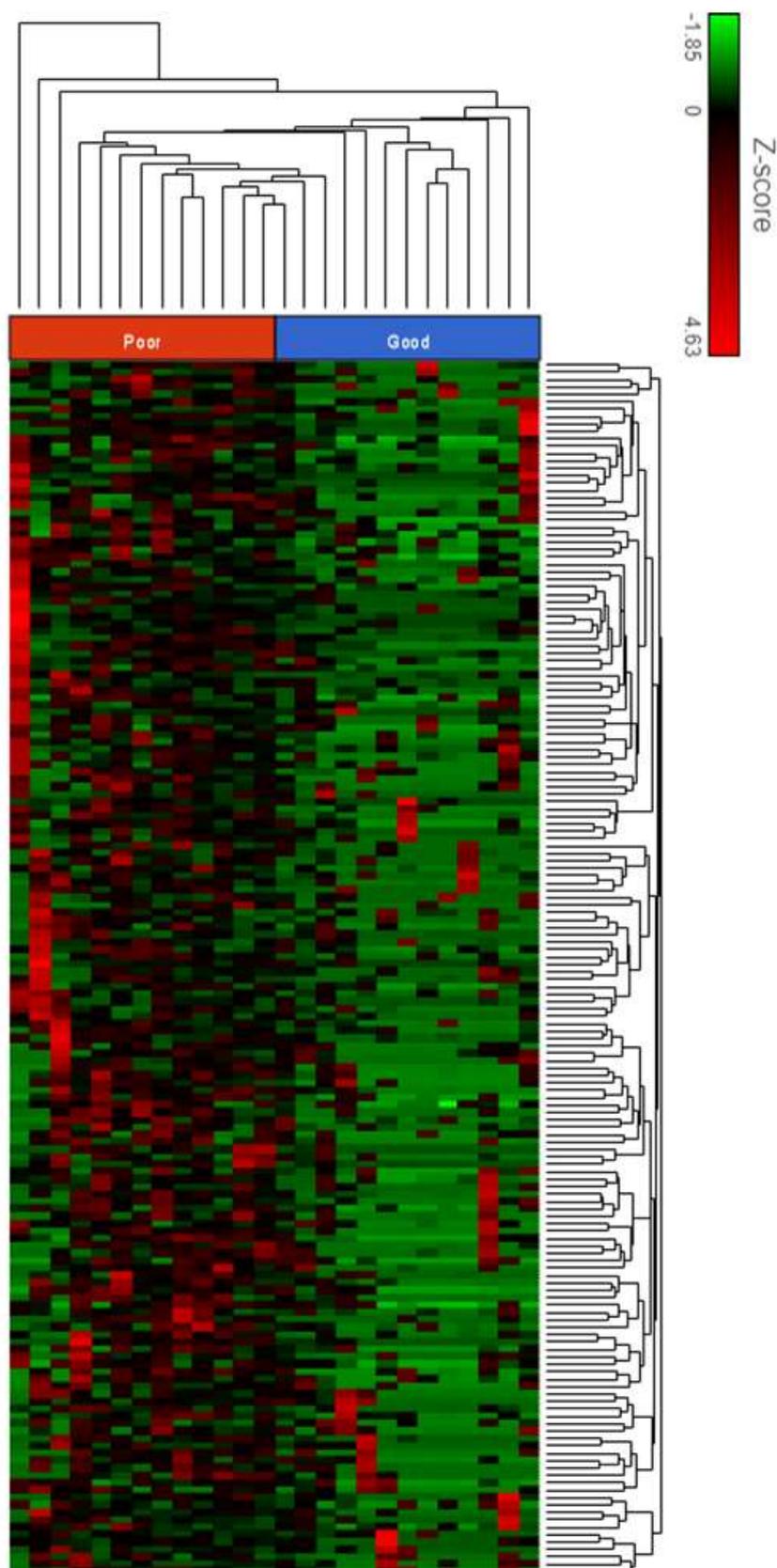
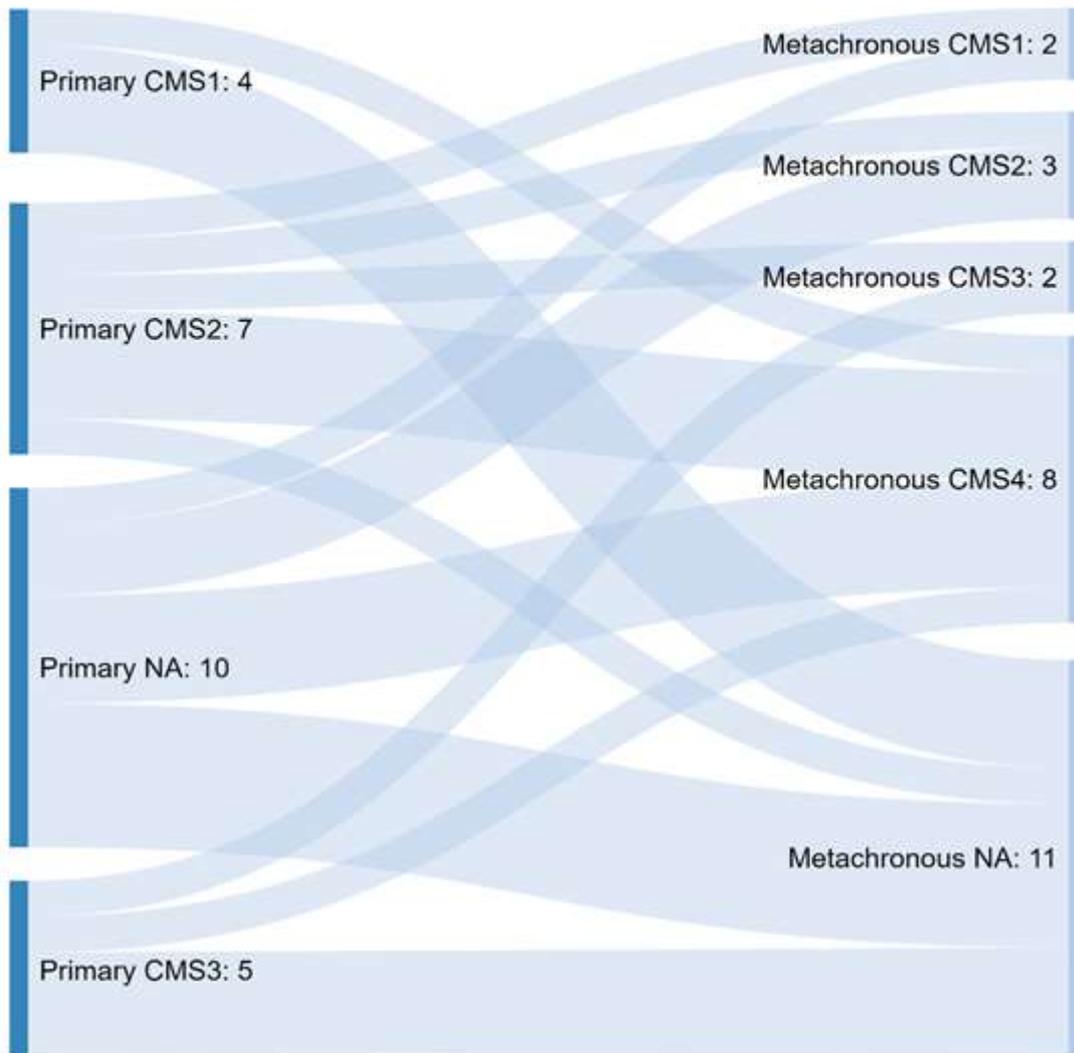


Figure 4: Sankey diagram of the CMS transition from primary to CPM

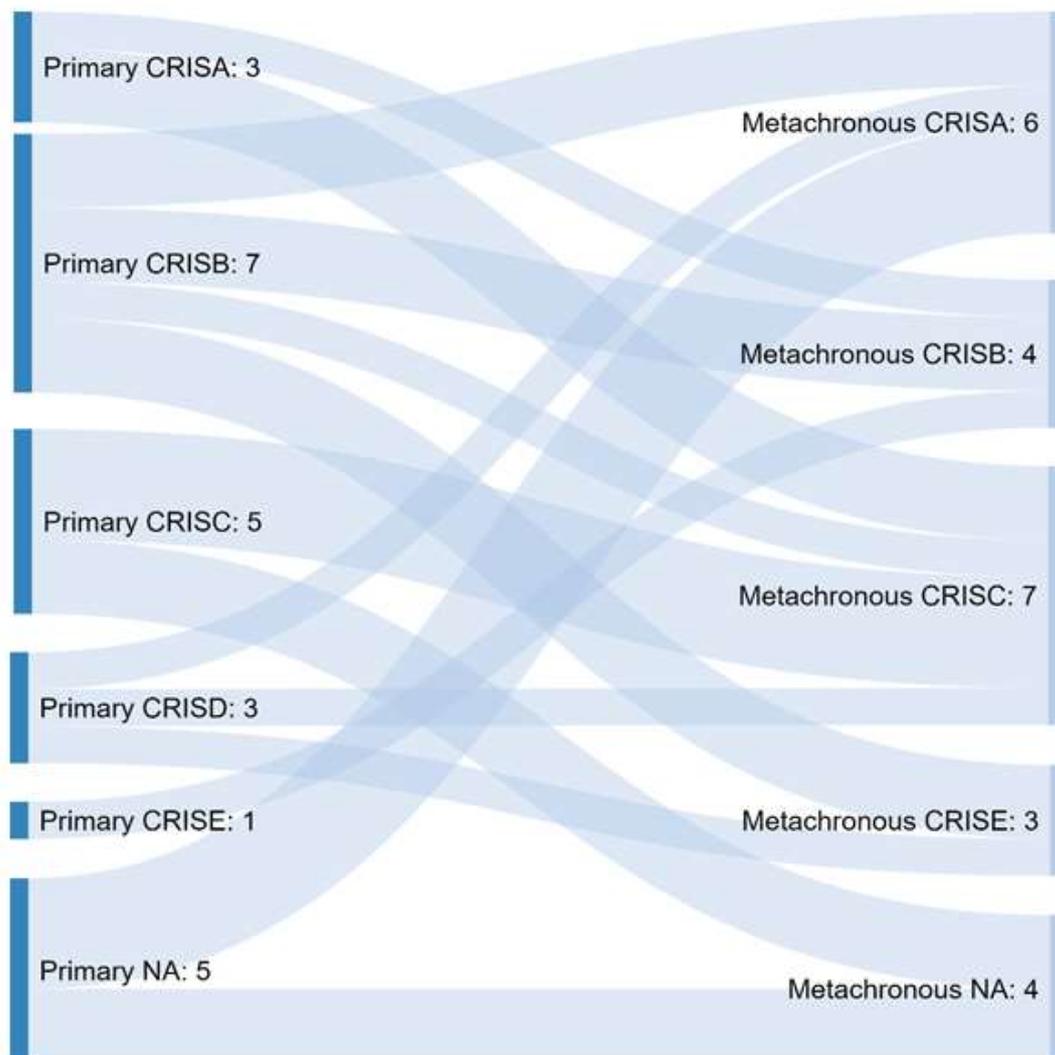


	Non-consensus	CMS1	CMS2	CMS3	CMS4	Total
Primary	10 (40)	4 (16)	6 (24)	5 (20)	0 (0)	25
CPM	12 (46)	1 (4)	3 (12)	2 (7)	8 (31)	26

**Figure 4 Sankey diagram of the CMS transition from primary to CPM and CMS classification**

*Fischers exact p-value 0.085, values in parenthesis percentages*

Figure 5: Sankey diagram of the CRIS transition from primary to CPM



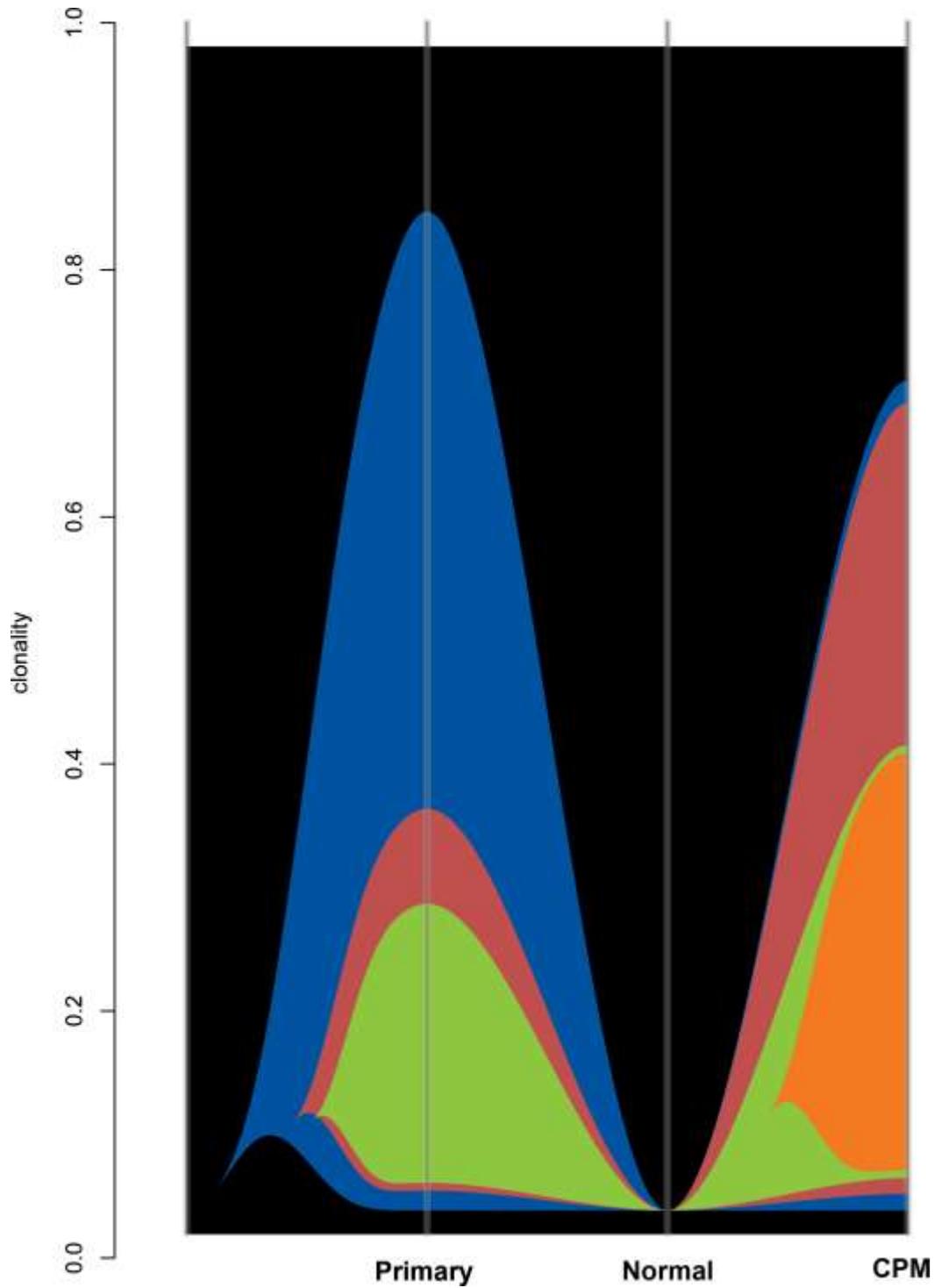
	Non-consensus	CRISA	CRISB	CRISC	CRISD	CRISE	Total
Primary	4 (16)	6 (24)	4 (16)	8 (32)	0 (0)	3 (12)	25
CPM	5 (19)	4 (16)	7 (30)	5 (19)	3 (12)	1 (4)	26

**Figure 5 Sankey diagram of the CRIS transition from primary to CPM**

*Fishers exact p-value 0.365, values in parenthesis percentages*

## Supplementary tables & figures

Supplementary figure 1: Clonality river plot of primary tumour, normal germline and colorectal peritoneal metastasis (CPM). Clonality on Y axis.



**Supplementary table 1: KEGG pathway enrichment analysis, primary CRC vs. matched metachronous CPM and good vs. poor prognosis CPM (p-value < 0.05)**

<b>Gene set</b>	<b>Class</b>	<b>KEGG pathway</b>	<b>Enrichment score</b>	<b>P-value</b>
<b>Primary CRC vs. matched CPM</b>				
hsa04810	Cellular Processes; Cell motility	Regulation of actin cytoskeleton	4.19	0.02
hsa05206	Human Diseases; Cancers: Overview	MicroRNAs in cancer	3.44	0.03
hsa05222	'Human Diseases; Cancers: Specific types	Small cell lung cancer	3.43	0.03
hsa04672	Organismal Systems; Immune system	Intestinal immune network for IgA production	3.30	0.04
hsa04670	Organismal Systems; Immune system	Leukocyte transendothelial migration	3.22	0.04

<b>Good vs. Poor prognosis CPM</b>				
hsa04144	Cellular Processes; Transport and catabolism	Endocytosis	11.67	8.56 x10-6
hsa00130	Metabolism; Metabolism of cofactors and vitamins	Ubiquinone and other terpenoid- quinone biosynthesis	5.07	6.28 x10-3
hsa05100	Human Diseases; Infectious diseases: Bacterial	Bacterial invasion of epithelial cells	4.76	8.56 x10-3
hsa05110	Human Diseases; Infectious diseases: Bacterial	Vibrio cholerae infection	4.14	0.02
hsa04145	Cellular Processes; Transport and catabolism	Phagosome	3.92	0.02
hsa05165	Human Diseases; Infectious diseases: Viral	Human papillomavirus infection	3.85	0.02
hsa00561	Metabolism; Lipid metabolism	Glycerolipid metabolism	3.64	0.03

hsa04966	Organismal Systems; Excretory system	Collecting duct acid secretion	3.63	0.03
hsa00230	Metabolism; Nucleotide metabolism	Purine metabolism	3.37	0.03
hsa05120	Human Diseases; Infectious diseases: Bacterial	Epithelial cell signalling in Helicobacter pylori infection	3.31	0.04
hsa00020	Metabolism; Carbohydrate metabolism	Citrate cycle (TCA cycle)	3.20	0.04
hsa01230	Biosynthesis of amino acids	Biosynthesis of amino acids	3.20	0.04
hsa04520	Cellular Processes; Cellular community – eukaryotes	Adherens junction	3.13	0.04
hsa03020	Genetic Information Processing; Transcription	RNA polymerase	11.67	8.56 x10-6
hsa04068	Environmental Information Processing; Signal transduction	FoxO signalling pathway	5.07	6.28 x10-3

**Supplementary table 2: The top 10 genes with significantly altered expression (FDR<0.1) in poor prognosis CPM**

Rank	Gene name	Description	Fold change	FDR p value
<b>Genes with reduced expression in poor prognosis vs. good prognosis CPM samples</b>				
1	KIAA0319L	Protein coding gene	1.30	2.89 x10-03
2	CORO7	F-actin regulator	1.16	7.36 x10-04
3	DIP2C	Disco interacting factor	1.13	1.90 x10-04
4	STK38	Serine/threonine kinase	1.05	6.38 x10-05
5	CLINT1	Clathrin Interacting factor	1.03	2.73 x10-03
<b>Genes with increased expression in poor prognosis vs. good prognosis CPM samples</b>				
1	VEZF1	Endothelial cell specific transcription factor	-10.40	3.76 x10-05
2	CEACAM1	Carcinoembryonic antigen immunoglobulin	-8.27	1.83 x10-03
3	BCYRN1	Small non-messenger RNA	-7.96	7.50 x10-05

4	RP11-192H23.5	Long non-coding RNA	-6.22	2.38 x10 <sup>-04</sup>
5	RELB	Proto-Oncogene	-6.17	9.50 x10 <sup>-04</sup>
6	DCAF12	Apoptosis regulator	-6.17	1.36 x10 <sup>-05</sup>
7	POLR3GL	RNA polymerase	-6.05	1.48 x10 <sup>-03</sup>
8	FEM1C	Protein coding gene	-5.69	1.87 x10 <sup>-03</sup>
9	AXIN1	Negative Wnt pathway regulator	-5.42	1.67 x10 <sup>-04</sup>
10	KHDRBS3	RNA-binding protein regulates alternative splicing	-5.01	1.23 x10 <sup>-03</sup>

**Supplementary table 3: The top 10 differentially methylated regions, (DMRs),  
CPM vs. primary CRC and good vs. poor prognosis CPM**

Chr	Start (bp)	End (bp)	Length (bp)	DMR p-value	Located Gene	UCSC C pG island
<b>CPM vs. primary CRC</b>						
chr11	2160904	2161586	682	0.002	IGF2	302
chr19	37157318	37157945	627	0.003	ZNF461	39
chr15	79383167	79383980	813	0.003	RASGRF1	195
chr18	35145983	35147090	1107	0.003	CELF4	196
chr19	58609269	58609987	718	0.003	ZSCAN18	61
chr13	78493229	78493712	483	0.003	EDNRB	70
chr6	28602543	28603027	484	0.004	ZBED9	32
chr5	135415693	135416613	920	0.005	VTRNA2-1	24
chr19	58458572	58459358	786	0.005	ZNF256	52
chr5	38258007	38259243	1236	0.008	EGFLAM	108
<b>Good vs. poor prognosis CPM</b>						
chr10	134600357	134600919	562	0.007	NKX6-2	569
chr12	133464496	133464737	473	0.003	CHFR	91
chr10	8096311	8096991	680	0.006	NA	509
chr20	30639816	30640256	440	0.009	HCK	71

chr16	54972078	54973579	1501	0.015	NA	179
chr11	2397201	2397655	454	0.016	BC019904	NA

**Supplementary table 4 :The top 10 CNAs, CPM vs. primary CRC and good vs. poor prognosis CPM**

Chr	Start	End	Type	Count	Score	GW_P	Gene
<b>CPM vs. primary CRC</b>							
chr11	Gain	3817692	3818253	22	0.91	2.78 X10- 07	NUP98
chr11	Gain	3825449	4629195	22	0.91	2.78 X10- 07	RRM1, TRIM21
chr11	Gain	4629529	5704417	22	0.91	2.78 X10- 07	MMP26, UBQLN3, TRIM5, 6
chr11	Gain	5705995	5705995	22	0.91	2.78 X10- 07	TRIM5

chr11	Gain	5712109	6518263	22	0.91	2.78 X10- 07	TRIM3, 5, 22, FAM160A2, CCKBR, SMPD1, APBB1, ARFIP2
chr11	Gain	6518907	6650480	22	0.91	2.78 X10- 07	DNHD1, ILK, TPP1, TAF10, RRP8
chr16	Gain	56635680	56645731	22	0.91	2.78 X10- 07	MT1A, 2A, 3, 4
chr16	Gain	57333722	57335017	22	0.91	2.78 X10- 07	TRNA <sup>Leu</sup>
chr16	Gain	57495650	57503371	22	0.91	2.78 X10- 07	POLR2C
chr16	Gain	71842906	71843647	22	0.91	2.78 X10- 07	AP1G1

Good vs. poor prognosis CPM							
chr3	6902689	6905031	Loss	12	1	2.22 X10- 16	GRM7
chr3	16554466	16554910	Loss	12	1	2.22 X10- 16	RFTN1
chr3	36985697	36986697	Loss	12	1	2.22 X10- 16	TRANK1
chr3	43020703	43022351	Loss	12	1	2.22 X10- 16	FAM198A
chr3	44596684	44596846	Loss	12	1	2.22 X10- 16	ZKSCAN7
chr3	46939978	46940439	Loss	12	1	2.22 X10- 16	PTH1R
chr4	25657365	25657628	Loss	12	1	2.22 X10- 16	SLC34A2

chr4	57975793	57976944	Loss	12	1	2.22 X10- 16	IGFBP7
chr4	81187125	81188051	Loss	12	1	2.22 X10- 16	FGF5
chr4	81951953	81952024	Loss	12	1	2.22 X10- 16	BMP3

*Chr, chromosomes, Type CNA, Start, start position of the CNA, End, end position of the CNA, count, number of samples carrying this type of event at this location, score, proportion of samples carrying this type of event at this location, GW\_P, poisson-binomial test for overlap of event across individuals at this location given proportion chromosome affected, Gene – gene encoded by this portion of the genome.*