

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features

BMC Cancer 2010, **10**:227 doi:10.1186/1471-2407-10-227

Pei Woon Ang (angp02@student.uwa.edu.au)
Marie Loh (csimlcs@nus.edu.sg)
Natalia Liem (csiv13@nus.edu.sg)
Pei Li Lim (peili.85@gmail.com)
Fabienne Grieu (fabiennegrieu@yahoo.com.au)
Aparna Vaithilingam (csiav@nus.edu.sg)
Cameron Platell (cameron.platell@uwa.edu.au)
Wei Peng Yong (wei_peng_yong@nuhs.edu.sg)
Barry Iacopetta (barry.iacopetta@uwa.edu.au)
Richie Soong (csirs@nus.edu.sg)

ISSN 1471-2407

Article type Research article

Submission date 4 November 2009

Acceptance date 21 May 2010

Publication date 21 May 2010

Article URL <http://www.biomedcentral.com/1471-2407/10/227>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

© 2010 Ang *et al.*, licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features

Pei Woon Ang^{1,2}, Marie Loh^{1,2}, Natalia Liem³, Pei Li Lim³, Fabienne Grieu¹, Aparna Vaithilingam², Cameron Platell^{1,4}, Wei Peng Yong³, Barry Iacopetta¹, Richie Soong^{2*}

¹ School of Surgery, University of Western Australia, Perth, Australia

² Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore

³ Department of Haematology and Oncology, National University Hospital, Singapore, Singapore

⁴ Department of Surgery, St John of God Hospital, Subiaco, Australia

*Corresponding author:

Richie Soong (A/Prof)

Cancer Science Institute of Singapore,

Centre for Life Science #02-15, 28 Medical Drive,

Singapore 117456, Republic of Singapore.

Phone: +65 65168055; Fax: +65 68739664

e-mail: csirs@nus.edu.sg

Abstract

Background: Most previous studies of the CpG island methylator phenotype (CIMP) in colorectal cancer (CRC) have been conducted on a relatively small numbers of CpG sites. In the present study we performed comprehensive DNA methylation profiling of CRC with the aim of characterizing CIMP subgroups.

Methods: DNA methylation at 1,505 CpG sites in 807 cancer-related genes was evaluated using the Illumina GoldenGate® methylation array in 28 normal colonic mucosa and 91 consecutive CRC samples. Methylation data was analyzed using unsupervised hierarchical clustering. CIMP subgroups were compared for various clinicopathological and molecular features including patient age, tumor site, microsatellite instability (MSI), methylation at a consensus panel of CpG islands and mutations in *BRAF* and *KRAS*.

Results: A total of 202 CpG sites were differentially methylated between tumor and normal tissue. Unsupervised hierarchical clustering of methylation data from these sites revealed the existence of three CRC subgroups referred to as CIMP-low (CIMP-L, 21% of cases), CIMP-mid (CIMP-M, 14%) and CIMP-high (CIMP-H, 65%). In comparison to CIMP-L tumors, CIMP-H tumors were more often located in the proximal colon and showed more frequent mutation of *KRAS* and *BRAF* ($P<0.001$).

Conclusions: Comprehensive DNA methylation profiling identified three CRC subgroups with distinctive clinicopathological and molecular features. This study suggests that both *KRAS* and *BRAF* mutations are involved with the CIMP-H pathway of CRC rather than with distinct CIMP subgroups.

Background

DNA hypermethylation-induced gene silencing is a common event in many malignancies and serves as an alternative mechanism to genetic mutation for the loss of tumor suppressor functions [1, 2]. Although the mechanisms that underlie aberrant DNA methylation in cancer cells remain to be elucidated, current evidence suggests that it may be an early and possibly even an initiating event in the development of colorectal cancer (CRC).

A subset of CRC has been shown to exhibit frequent and concurrent hypermethylation at specific gene promoters and is referred to as the CpG island methylator phenotype (CIMP+) [3]. CIMP+ CRC is associated with distinct clinicopathological and molecular features including proximal tumor location, preponderance in elderly females, poorly differentiated and mucinous tumor histology, microsatellite instability (MSI) and frequent *BRAF* V600E mutation [3-10]. CIMP+ CRC often lack the hallmark genetic alterations in *APC*, *p53* and 18q that characterize the classic adenoma-carcinoma sequence. Instead, CIMP+ tumors are thought to develop along an alternate serrated adenoma pathway in which hypermethylation rather than mutation is used to inactivate tumor suppressor genes [11].

In an effort to establish CIMP+ CRC as a distinct subgroup of CRC, Laird and colleagues analysed the methylation of 195 individual gene promoter regions in 295 CRC using the quantitative MethyLight assay [10]. From their results, they proposed a panel of 5 CpG island methylation markers to standardize the classification of CIMP+ CRC. However, different groups have continued to use a variety of methylation markers to define CIMP+ CRC [7, 12-14]. The lack of consensus markers has led to reports of several CIMP subgroups according to

the frequency of CpG island methylation [13-15]. The investigators who originally proposed CIMP recently described two subgroups of CIMP+, termed CIMP-1 and CIMP-2, that displayed increased frequencies of *BRAF* and *KRAS* mutations, respectively [14]. Similarly, Nagasaka *et al* described two distinct patterns of gene methylation in CRC that also segregated with *BRAF* and *KRAS* mutations [13, 16]. Using a panel of 8 methylation markers, Ogino *et al* identified a CRC subgroup which they termed CIMP-low that was associated with frequent *KRAS* mutation, *MGMT* methylation and occurrence in males [17].

Most previous studies of CIMP+ CRC have investigated a relatively small number of CpG island markers for methylation. The GoldenGate Methylation BeadArray (Illumina, Inc.) technology provides the opportunity for high-throughput methylation analysis of a large number of CpG sites. In the present study the GoldenGate Methylation Cancer Panel I containing 1,505 CpG loci within 807 cancer-related genes was used to study methylation patterns in 91 unselected CRC. These genes were selected based on their involvement in cell growth control, differentiation, migration, apoptosis, DNA damage repair and oxidative metabolism. The GoldenGate technology allowed us to identify three distinct CRC subgroups according to their methylation pattern which showed distinctive clinicopathological and molecular characteristics and differed in their frequencies of *BRAF* and *KRAS* mutation.

Methods

Tissue samples

Unselected cases of CRC and adjacent normal colonic mucosa were obtained from 91 patients undergoing surgical resection at St John of God Hospital, Subiaco, Western Australia. All samples were snap-frozen in liquid nitrogen at the time of surgery and stored at -80°C until use. This set of tumors contains well-annotated clinicopathological information including age, gender, tumor location, staging, presence of lymphocytic infiltration and careful pathological assessment of perineural (PNI), lymphovascular (LVI) and extramural invasion (EMVI). Informed consent was obtained from all patients and the project was approved by the Human Research Ethics Committee of St John of God Hospital.

BRAF mutation, KRAS mutation and microsatellite instability

DNA was extracted from approximately 25mg of tissue using standard phenol-chloroform extraction. Hotspot mutations in *BRAF* (V600E) and *KRAS* (codons 12 and 13) were identified using fluorescent single strand conformation polymorphism (F-SSCP) as described previously [18, 19]. Deletions in the BAT-26 mononucleotide repeat were detected using F-SSCP and this was used to establish MSI+ status [20].

MethyLight determination of CIMP^W status

Sodium bisulfite modification was performed using the EZ DNA methylation kit according to the manufacturer's instructions (Zymo Research, Orange, CA) and eluted into 20 µl of 10 mmol/L Tris-HCl (pH 8). The required amount of genomic DNA to ensure reliable evaluation of DNA methylation following bisulfite modification was determined as described previously

[21]. DNA methylation levels for the panel of markers (*RUNX3*, *CACNA1G*, *IGF2*, *NEUROG1*, *SOCS1*) described by Weisenberger *et al* [10] were measured using MethyLight as described by the authors. The percentage of methylated reference (PMR) was calculated and normalised against β -actin to account for variability in the amount of input bisulfite-treated DNA. SssI methylase-treated DNA was used as the methylated standard. A threshold PMR value of >4 was used to classify loci as methylated or non-methylated. In the present study, CIMP^W refers to the classification of CIMP using the panel of markers described by Weisenberger *et al.*, whereby CIMP^W-high is defined as 3 or more methylated loci, CIMP^W-low as 1 or 2 methylated loci and CIMP^W-negative as no methylated loci.

DNA methylation profiling using Illumina GoldenGate® methylation bead array

Comprehensive DNA methylation profiling using the Illumina Goldengate Methylation Arrays® (Illumina, San Diego, CA) was carried out as described by Bibikova *et al* [22] on 91 CRC and 28 randomly selected, matched normal colonic mucosa samples. Briefly, DNA was quantified by real-time PCR and treated with bisulfite as for the MethyLight assay. Human sperm DNA and Universal methylated DNA (Chemicon, Temcula, CA) were included in each run as unmethylated and methylated controls, respectively. The bisulfite-converted DNA was probed at 1,505 individual CpG loci contained within 807 genes in the GoldenGate Methylation Cancer Panel I according to the manufacturer's instructions (Illumina). Hybridised arrays were scanned using the BeadArray Reader (Illumina). Extraction and normalization of intensity data was performed using the Beadscan software. To ensure adequate sample quality, only samples having >75% loci with a detection *P*-value of <0.05 were included for analysis.

Statistical analysis

The methylation level at each CpG site, or β -value, was defined as the ratio of methylated allele to the sum of methylated and unmethylated alleles and ranged from 0 (completely unmethylated) to 1 (completely methylated). Normalisation of background intensity was estimated from a set of built-in negative controls and subtracted from each methylation data point as performed in other studies [23, 24]. All statistical analyses were carried out using β -value as a continuous variable unless specified otherwise. To compare the number of methylated genes between different tumor subgroups, β -values were binarized using a methylated threshold of 0.297. Using this threshold, methylated controls in the array were classified as unmethylated at a 5% false discovery rate. A total of 84 CpG sites contained within 39 X-chromosome genes were excluded from the analysis in order to eliminate gender-specific bias.

Unsupervised and supervised hierarchical clustering analyses were performed with the *heatmap.2* function in the *gplots* library. Unsupervised clustering was used to characterize methylation patterns in an unbiased fashion, as performed in other studies using methylation arrays [14, 25-27]. Supervised clustering analysis was used to further investigate methylation differences observed in unsupervised clustering. The optimal number of clusters was determined using the Hubert & Levine internal cluster quality index [28]. The robustness of this number was evaluated by bootstrap resampling analysis (n=1000). Additional evidence to support the delineation of clusters was obtained through unsupervised principal component analysis. The frequency and level of CpG methylation across different clusters was compared using a two-sample proportion test based on both binarised and continuous β -values. The

association of clinicopathological and molecular variables with each cluster was analysed using continuous β -values and the two-sample proportion t-test. All statistical analyses were performed in R version 2.7.1 (The R Foundation for Statistical Computing) at 5% significance level unless otherwise stated. Where applicable, Bonferroni correction was applied to adjust for multiple testing.

Results

DNA methylation patterns in normal and tumor tissue

Unsupervised hierarchical clustering of DNA methylation data from 1,505 CpG sites in 28 samples of normal colonic mucosa revealed no distinct clusters [Additional file 1]. As expected, the methylation status of 84 CpG sites in 39 genes located in the X-chromosome was perfectly correlated with gender [Additional file 1]. These genes were excluded from subsequent analyses. For the 91 tumor samples, three clusters were observed when methylation data from all 1,505 loci were included in the analysis [Additional file 2].

A total of 202 CpG sites, corresponding to 132 genes (90 hypermethylated and 42 hypomethylated), were differentially methylated between tumor and normal colonic mucosa ($P < 0.001$, FDR 5%) [Additional file 3]. Unsupervised hierarchical clustering of methylation data from these 202 tumor-specific markers identified three major tumor groups (Fig. 1), referred to here as CIMP-high (CIMP-H; 59/91, 65%), CIMP-mid (CIMP-M; 13/91, 14%) and CIMP-low (CIMP-L; 19/91, 21%). The mean methylation level (β -value) of the 202 CpG sites for these groups was 0.617, 0.506 and 0.370, respectively ($P < 0.001$). Binarization of the methylation readings using a β -value cut-off of 0.297 revealed a decreasing number of methylated CpG sites for the three groups (167, 136 and 105 respectively; $P < 0.001$).

Although branching of the dendrogram suggested the existence of two subgroups within CIMP-H (Fig. 1), the mean methylation level and the frequency of methylation between these groups were not significantly different ($P = 0.37$ and $P = 0.90$ respectively). Additional evidence for the

validity of tumor segregation was obtained through unsupervised principal component analysis (PCA) [Additional file 4]. CIMP-H could be clearly segregated from CIMP-L. CIMP-H and CIMP-M could also be discriminated from each other, although less distinctly. This is presumably because of a greater similarity between these two groups [Additional file 4]. In further support, 3 was the most frequent optimal number of clusters in bootstrap resampling analysis.

CIMP subgroups show distinctive clinicopathological and molecular features

The distribution of clinicopathological and molecular features for 91 CRC in relation to the methylation pattern obtained from analysis of all 202 differentially methylated CpG sites is shown in Fig. 1. Calculation of associations between these features and the three CIMP subgroups are shown in Table 1. Similar to previous reports on CIMP+, the CIMP-H tumors in this study were significantly associated with older age, proximal tumor location and *BRAF* mutation relative to CIMP-M and CIMP-L tumors. CIMP-H was also significantly associated with MSI+ when compared to CIMP-M, but not CIMP-L tumors. Two of the 15 MSI+ tumors were observed in the CIMP-L group and 13 in the CIMP-H group. Interestingly, the two patients with CIMP-L MSI+ tumors were aged 44 and 60 years, suggesting the underlying cause of the MSI+ phenotype was germline or somatic mutation of the mismatch repair genes rather than *hMLH1* methylation. Indeed, no *hMLH1* methylation was detected in both these cases in the GoldenGate Methylation Array data.

All 16 tumors classified as CIMP^W-high by MethyLight analysis using the panel of markers proposed by Weisenberger *et al* (>3/5 sites methylated) were contained within the CIMP-H group, while all 18 tumors classified as CIMP^W-low (1/5 or 2/5 sites methylated) segregated

into the CIMP-H or CIMP-M groups. Unfortunately, CpG sites for only 2 (*RUNX3*, *IGF2*) of the 5 genes in the CIMP^W panel were included in the Golden Gate arrays, thus preventing comparison of CIMP status by array and MethyLight analysis. Nevertheless, there was good correlation between the array and MethyLight methods for methylation levels of *RUNX3* and *IGF2* (all $p < 0.05$) using Pearson correlation.

All 15 tumors with *BRAF* mutation were CIMP-H. A significantly higher frequency of *KRAS* mutation was observed in CIMP-H compared to CIMP-L or CIMP-M tumors. None of the 13 CIMP-M tumors contained a *KRAS* mutation. The presence of extramural vascular invasion (EMVI) was more frequent in CIMP-M compared to CIMP-H or CIMP-L tumors. The presence of a tumor-infiltrating lymphocytic response (TILS) was not associated with any of the CIMP subgroups.

Differentially methylated genes in CIMP subgroups

Five clusters of CpG loci, termed A to E, were apparent following unsupervised hierarchical clustering of methylation data for the 202 CpG loci that showed tumor-specific methylation (Fig. 1). CpG sites in clusters A and C were more highly methylated in CIMP-M and CIMP-H tumors compared to CIMP-L tumors, while the converse was true for the CpG sites in cluster D. CpG sites in cluster B and cluster E showed uniformly high and low methylation, respectively, in each of the 3 CIMP subgroups.

Using published data from studies on human stem cells [29], 50% (39/98) of the genes within clusters A and C were found to be targets for binding by Polycomb repressive complex 2 (PRC2) components and/or H3K27 trimethylation. In contrast, only 12% (5/41) of the genes

within clusters B, D and E were targets ($P < 0.001$). These observations support previous reports that hypermethylated genes in cancer are frequent targets of PRC2-mediated H3K27 trimethylation [30].

Discussion

The current study is the first to use array-based technology to enable comprehensive methylation profiling of CRC. A total of 1,505 CpG sites contained within 807 genes were assessed in 91 consecutive cases of CRC. The GoldenGate[®] arrays employed here were recently used to profile methylation in head and neck cancer [31], renal cancer [32], glioblastoma [33] and hematological neoplasms [24, 27]. The validity of these arrays for the quantitative assessment of methylation was shown in several previous studies by comparison with other quantitative methods [23, 26, 34]. The finding that methylation of CpG sites in X-linked genes correlated with gender provided further validation [Additional file 1]. Many of the genes found to be hypermethylated in this study were previously reported to be methylated in CRC [Additional file 5]. Finally, in agreement with earlier work on cancer [30], many of the genes showing *de novo* hypermethylation in this study of CRC (cluster A and C genes, Fig. 1) are known targets for PRC2 [29].

Similar to earlier studies in CRC that evaluated a limited number of methylation markers [3-10], comprehensive methylation profiling in the present study revealed the existence of distinct tumor subgroups (Fig. 1). The three major subgroups identified by unsupervised hierarchical clustering were classified as CIMP-H, CIMP-M and CIMP-L according to the level and frequency of methylation. In agreement with previous studies, CIMP-H tumors were associated with older patient age, proximal site and *BRAF* mutation (Table 1). All 16 tumors identified as CIMP^W-high using a proposed consensus panel of 5 markers were contained within the CIMP-H group, as well as all 15 tumors containing a *BRAF* mutation. Using small numbers of methylation markers in unselected CRC, the original studies by Toyota *et al*

reported CIMP+ frequencies of 62% [3] and 51% [35] whereas subsequent studies reported lower frequencies of 15-32% [5, 7-10, 14]. In contrast, by investigating a large number of methylation sites and using unsupervised hierarchical clustering to analyze the results, we observed a relatively high proportion (65%) of CIMP-H tumors in the present study.

Previous studies have reported inconsistent results for the association between CpG island methylation and *KRAS* mutation [8, 13, 17, 35, 36], probably because of the different methylation markers used in each study. Analysis of a large number of CpG sites in the present study revealed that CIMP-H tumors showed a significantly higher *KRAS* mutation frequency compared to both CIMP-M and CIMP-L tumors (Table 1). This result agrees with some studies [4, 5, 8, 35] but not others that found an inverse association between *KRAS* mutation and CIMP+ [7, 9, 10].

Since *BRAF* mutations are strongly associated with CIMP and mutually exclusive to *KRAS* mutations ([10]; Fig. 1), a point of interest is whether methylation patterns differ between tumors with *BRAF* and *KRAS* mutations. Supervised clustering analysis with Bonferroni correction revealed that only 1 of the 202 tumor-specific CpG sites was differentially methylated between these tumor groups (HTR1B_P222_F, upregulated in *BRAF* mutant tumors, $p=8.1 \times 10^{-6}$). HTR1B (5-hydroxytryptamine (serotonin) receptor 1B) is a G protein-coupled multi-pass membrane protein involved in regulation of the serotonin system [37]. The gene is hypermethylated in lung cancer and its chromosome locus (16q14.1) is frequently deleted in a number of cancer types [38]. However, no links with *BRAF* or *RAS* mutations or signaling have been reported.

A novel finding of this array-based analysis was the existence of an apparent CIMP-M group (Fig. 1). These tumors displayed a higher frequency of EMVI compared to both CIMP-L and CIMP-H, and a significantly higher stage compared to CIMP-H (Table 1). CIMP-M tumors were almost exclusively located in the distal colon or rectum (12/13, 92%). MSI and *KRAS* and *BRAF* mutations were notably absent in these tumors, although this may be due to reportedly lower frequencies of these alterations in distal tumors [39]. Taken together, these results suggest CIMP-M tumors could be a distinct clinical and molecular entity, although confirmation in larger, independent tumor series is required.

After adjustment for multiple testing, 170 CpG sites were hypermethylated in CIMP-H compared to CIMP-L. The 112 genes containing these CpG sites are ranked according to significance in Additional file 5. Of these, 54 were previously reported as methylated in cancer, 38 as methylated in gastrointestinal cancers and 30 in CRC [Additional file 5]. Of the top 10 genes that were hypermethylated in CIMP-H compared to CIMP-L tumors, 5 have previously been implicated in the pathogenesis of gastrointestinal tumors (*NTRK3*, *HS3ST2*, *TWIST1*, *CD40* and *EYA4*). Somatic mutation of *NTRK3* has been reported in human colon cancer [40], while methylation of *EYA4* has been documented previously in ulcerative colitis-associated dysplasia [41] and CRC [42].

CIMP-M tumors were found to have a relatively high incidence of EMVI (38%) compared to CIMP-H and CIMP-L tumors (Table 1). Supervised analysis revealed that *HS3ST2*, also known as *3-OST-2*, was the only gene to be differentially methylated between tumors showing presence or absence of EMVI. Methylation-associated silencing of *HS3ST2* expression has been demonstrated in breast, lung, pancreatic and colon cancers [43]. This gene encodes an

enzyme that modifies heparin sulfate proteoglycans [44] involved in cell adhesion and migration [45], thus suggesting a possible mechanistic link between *HS3ST2* methylation and EMVI.

The use of Illumina GoldenGate® Beadarray technology in this study allowed a large number of CpG sites to be evaluated for methylation in an unbiased fashion. However, there are several limitations with this approach for the characterization of CIMP subgroups in CRC. Firstly, only a small fraction of all genes were investigated for methylation and in 70% of these just one CpG site per gene was evaluated. Secondly, it is unclear whether the methylation level at these sites relates to expression of the genes. Thirdly, some of the probes used in this assay contain single nucleotide polymorphisms (SNPs) or repetitive elements that could influence methylation analysis [25]. The cost effectiveness of using arrays to characterize CIMP-H is questionable, given the strong concordance between CIMP-H from this study and CIMP^w. Further studies should clarify if the additional information provided by methylation arrays is worth the complexity and expense.

Conclusions

Methylation profiling of 807 cancer-related genes revealed the presence of three CRC subgroups with distinct clinicopathological and molecular features. Similar to earlier studies that investigated fewer methylation markers, CIMP-H CRC were associated with older patient age, proximal location and mutations in *BRAF* and *KRAS*. Further investigations in large and independent population-based series are required to validate these findings and to assess the clinical utility of CIMP subgroups.

Competing interests

The authors declare that they have no competing interests

Authors' contributions

PWA, FG, NL, PLL, AV carried out the experimental work. ML performed the statistical analysis. CP and WPY contributed gave critical clinical perspective to the results, and BI and RS co-ordinated the study and compiled the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

This work is supported by the Singapore Cancer Syndicate (SCS#BU51), and the Singapore National Research Foundation and the Ministry of Education under the Research Center of Excellence Programme. We thank Patrick Tan and his laboratory staff for assistance with data generation and technical support. PWA is supported by the International Postgraduate Research Scholarship.

References

1. Esteller M: **Epigenetics in cancer**. *N Engl J Med* 2008, **358**(11):1148-1159.
2. Issa JP: **CpG island methylator phenotype in cancer**. *Nat Rev Cancer* 2004, **4**(12):988-993.
3. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP: **CpG island methylator phenotype in colorectal cancer**. *Proc Natl Acad Sci U S A* 1999, **96**(15):8681-8686.
4. Barault L, Charon-Barra C, Jooste V, de la Vega MF, Martin L, Roignot P, Rat P, Bouvier AM, Laurent-Puig P, Faivre J *et al*: **Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases**. *Cancer Res* 2008, **68**(20):8541-8546.
5. Hawkins N, Norrie M, Cheong K, Mokany E, Ku SL, Meagher A, O'Connor T, Ward R: **CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability**. *Gastroenterology* 2002, **122**(5):1376-1387.
6. Ogino S, Cantor M, Kawasaki T, Brahmandam M, Kirkner GJ, Weisenberger DJ, Campan M, Laird PW, Loda M, Fuchs CS: **CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies**. *Gut* 2006, **55**(7):1000-1006.
7. Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS: **Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample**. *J Mol Diagn* 2007, **9**(3):305-314.
8. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML: **Evaluation of a large, population-based sample supports a CpG**

- island methylator phenotype in colon cancer.** *Gastroenterology* 2005, **129**(3):837-845.
9. van Rijnsoever M, Grieu F, Elsaleh H, Joseph D, Iacopetta B: **Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands.** *Gut* 2002, **51**(6):797-802.
 10. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D *et al*: **CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.** *Nat Genet* 2006, **38**(7):787-793.
 11. Jass JR: **Classification of colorectal cancer based on correlation of clinical, morphological and molecular features.** *Histopathology* 2007, **50**(1):113-130.
 12. Ferracin M, Gafa R, Miotto E, Veronese A, Pultrone C, Sabbioni S, Lanza G, Negrini M: **The methylator phenotype in microsatellite stable colorectal cancers is characterized by a distinct gene expression profile.** *J Pathol* 2008, **214**(5):594-602.
 13. Nagasaka T, Koi M, Kloor M, Gebert J, Vilkin A, Nishida N, Shin SK, Sasamoto H, Tanaka N, Matsubara N *et al*: **Mutations in both KRAS and BRAF may contribute to the methylator phenotype in colon cancer.** *Gastroenterology* 2008, **134**(7):1950-1960, 1960 e1951.
 14. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K *et al*: **Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer.** *Proc Natl Acad Sci U S A* 2007, **104**(47):18654-18659.
 15. Kawasaki T, Ohnishi M, Nosho K, Suemoto Y, Kirkner GJ, Meyerhardt JA, Fuchs CS, Ogino S: **CpG island methylator phenotype-low (CIMP-low) colorectal cancer**

- shows not only few methylated CIMP-high-specific CpG islands, but also low-level methylation at individual loci. *Mod Pathol* 2008, **21**(3):245-255.
16. Nagasaka T, Sasamoto H, Notohara K, Cullings HM, Takeda M, Kimura K, Kambara T, MacPhee DG, Young J, Leggett BA *et al*: **Colorectal cancer with mutation in BRAF, KRAS, and wild-type with respect to both oncogenes showing different patterns of DNA methylation.** *J Clin Oncol* 2004, **22**(22):4584-4594.
 17. Ogino S, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS: **CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations.** *J Mol Diagn* 2006, **8**(5):582-588.
 18. Li WQ, Kawakami K, Ruszkiewicz A, Bennett G, Moore J, Iacopetta B: **BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status.** *Mol Cancer* 2006, **5**:2.
 19. Wang C, van Rijnsoever M, Grieu F, Bydder S, Elsaleh H, Joseph D, Harvey J, Iacopetta B: **Prognostic significance of microsatellite instability and Ki-ras mutation type in stage II colorectal cancer.** *Oncology* 2003, **64**(3):259-265.
 20. Iacopetta B, Grieu F: **Routine detection of the replication error phenotype in clinical tumor specimens using fluorescence-SSCP.** *Biotechniques* 2000, **28**(3):566-568, 570.
 21. Ang PW, Toh HB, Iacopetta B, Soong R: **An improved quality control for bisulfite-PCR-based DNA methylation analysis: cycle threshold value.** *Clin Chem Lab Med* 2008, **46**(8):1117-1121.
 22. Bibikova M, Fan JB: **GoldenGate assay for DNA methylation profiling.** *Methods Mol Biol* 2009, **507**:149-163.

23. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E *et al*: **High-throughput DNA methylation profiling using universal bead arrays.** *Genome Res* 2006, **16**(3):383-393.
24. Martin-Subero JI, Ammerpohl O, Bibikova M, Wickham-Garcia E, Agirre X, Alvarez S, Bruggemann M, Bug S, Calasanz MJ, Deckert M *et al*: **A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms.** *PLoS One* 2009, **4**(9):e6986.
25. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, Yang AS: **Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns.** *Hum Mol Genet* 2009, **18**(24):4808-4817.
26. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R *et al*: **Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context.** *PLoS Genet* 2009, **5**(8):e1000602.
27. O'Riain C, O'Shea DM, Yang Y, Le Dieu R, Gribben JG, Summers K, Yeboah-Afari J, Bhaw-Rosun L, Fleischmann C, Mein CA *et al*: **Array-based DNA methylation profiling in follicular lymphoma.** *Leukemia* 2009, **23**(10):1858-1866.
28. Hubert LJ, Levin JR: **A general statistical framework for assessing categorical clustering in free recall.** *Psychological Bulletin* 1976, **83**:1072-1080.
29. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K *et al*: **Control of developmental regulators by Polycomb in human embryonic stem cells.** *Cell* 2006, **125**(2):301-313.

30. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I *et al*: **Epigenetic stem cell signature in cancer**. *Nat Genet* 2007, **39**(2):157-158.
31. Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, Nelson HH, Wiemels JL, Zheng S, Posner MR *et al*: **Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma**. *Carcinogenesis* 2009, **30**(3):416-422.
32. McDonald FE, Morris MR, Gentle D, Winchester L, Baban D, Ragoussis J, Clarke NW, Brown MD, Kishida T, Yao M *et al*: **CpG methylation profiling in VHL related and VHL unrelated renal cell carcinoma**. *Mol Cancer* 2009, **8**:31.
33. Martinez R, Martin-Subero JJ, Rohde V, Kirsch M, Alaminos M, Fernandez AF, Ropero S, Schackert G, Esteller M: **A microarray-based DNA methylation study of glioblastoma multiforme**. *Epigenetics* 2009, **4**(4):255-264.
34. Ladd-Acosta C, Pevsner J, Sabuncuyan S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP: **DNA methylation signatures within the human brain**. *Am J Hum Genet* 2007, **81**(6):1304-1315.
35. Toyota M, Ohe-Toyota M, Ahuja N, Issa JP: **Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype**. *Proc Natl Acad Sci U S A* 2000, **97**(2):710-715.
36. Nosh K, Irahara N, Shima K, Kure S, Kirkner GJ, Schernhammer ES, Hazra A, Hunter DJ, Quackenbush J, Spiegelman D *et al*: **Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample**. *PLoS One* 2008, **3**(11):e3698.

37. Drago A, Alboni S, Nicoletta B, De Ronchi D, Serretti A: **HTR1B as a risk profile maker in psychiatric disorders: a review through motivation and memory.** *Eur J Clin Pharmacol* 2010, **66**(1):5-27.
38. Takai D, Yagi Y, Wakazono K, Ohishi N, Morita Y, Sugimura T, Ushijima T: **Silencing of HTR1B and reduced expression of EDN1 in human lung cancers, revealed by methylation-sensitive representational difference analysis.** *Oncogene* 2001, **20**(51):7505-7513.
39. Slattery ML, Curtin K, Wolff RK, Boucher KM, Sweeney C, Edwards S, Caan BJ, Samowitz W: **A comparison of colon and rectal somatic DNA alterations.** *Dis Colon Rectum* 2009, **52**(7):1304-1311.
40. Wood LD, Calhoun ES, Silliman N, Ptak J, Szabo S, Powell SM, Riggins GJ, Wang TL, Yan H, Gazdar A *et al*: **Somatic mutations of GUCY2F, EPHA3, and NTRK3 in human cancers.** *Hum Mutat* 2006, **27**(10):1060-1061.
41. Osborn NK, Zou H, Molina JR, Lesche R, Lewin J, Lofton-Day C, Klatt KK, Harrington JJ, Burgart LJ, Ahlquist DA: **Aberrant methylation of the eyes absent 4 gene in ulcerative colitis-associated dysplasia.** *Clin Gastroenterol Hepatol* 2006, **4**(2):212-218.
42. Schatz P, Distler J, Berlin K, Schuster M: **Novel method for high throughput DNA methylation marker evaluation using PNA-probe library hybridization and MALDI-TOF detection.** *Nucleic Acids Res* 2006, **34**(8):e59.
43. Miyamoto K, Asada K, Fukutomi T, Okochi E, Yagi Y, Hasegawa T, Asahara T, Sugimura T, Ushijima T: **Methylation-associated silencing of heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2 (3-OST-2) in human breast, colon, lung and pancreatic cancers.** *Oncogene* 2003, **22**(2):274-280.

44. Shworak NW, Liu J, Petros LM, Zhang L, Kobayashi M, Copeland NG, Jenkins NA, Rosenberg RD: **Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase. Isolation, characterization, and expression of human cdnas and identification of distinct genomic loci.** *J Biol Chem* 1999, **274**(8):5170-5184.
45. Perrimon N, Bernfield M: **Specificities of heparan sulphate proteoglycans in developmental processes.** *Nature* 2000, **404**(6779):725-728.

Figure legends

Fig. 1 Unsupervised hierarchical clustering of 202 tumor-specific probes (rows) in 91 CRC (columns).

The 3 tumor clusters generated by this analysis were termed CIMP-High (CIMP-H), CIMP-Mid (CIMP-M) and CIMP-Low (CIMP-L). Clinicopathological and molecular features are shown above the heatmap. White rectangles are cases with missing data. Gender: female (red), male (blue); Age: ≥ 67 years (red), < 67 (blue); Tumor location: proximal (red), distal (blue); Tumor stage (ACPS): A or B (blue), C or D (red); Lymphovascular invasion (LVI): present (red), absent (blue); Extramural vascular invasion (EMVI): present (red), absent (blue); Perineural invasion (PNI): present (red), absent (blue); Tumor infiltrating lymphocytes (TILS): present (red), absent (blue); CIMP^W: CIMP^W-high (red), CIMP^W-low (yellow), CIMP^W-negative (blue); *BRAF*: mutant (red), wildtype (blue); *KRAS*: mutant (red), wildtype (blue); Microsatellite instability (MSI): positive (red), negative (blue). Five CpG clusters (A-E) were apparent from the analysis and showed differential methylation amongst the 3 CIMP subgroups.

Table 1: Clinicopathological and molecular characteristics of CIMP subgroups

	CIMP subgroup (n, %)			<i>P</i>		
	L 19 (21)	M 13 (14)	H 59 (65)	L vs M	M vs H	L vs H
Female	6 (32)	4 (31)	30 (51)			
Male	13 (68)	9 (69)	29 (49)	0.952	0.211	0.175
Age \geq 67 years	6 (32)	5 (38)	37 (63)			
Age < 67 years	13 (68)	8 (62)	22 (37)	0.570	0.003	0.012
Proximal tumor site ¹	5 (26)	1 (8)	29 (49)			
Distal tumor site ²	14 (74)	12 (92)	29 (49)	0.152	<0.001	0.001
ACPS Stage A or B	8 (42)	4 (31)	36 (61)			
ACPS Stage C or D	11 (58)	9 (69)	23 (39)	0.520	0.025	0.105
LVI Negative	15 (79)	6 (46)	39 (66)			
LVI Positive	4 (21)	7 (54)	20 (34)	0.049	0.188	0.126
EMVI Negative	19 (100)	8 (62)	52 (88)			
EMVI Positive	0 (0)	5 (38)	7 (12)	0.005	0.031	0.024
PNI Negative	17 (89)	11 (85)	52 (88)			
PNI Positive	2 (11)	2 (15)	7 (12)	0.744	0.506	0.708
TILS Negative ³	9 (47)	5 (38)	26 (44)			
TILS Positive	7 (37)	8 (62)	33 (56)	0.326	0.547	0.521
CIMP ^W – negative ⁴	19 (100)	10 (77)	28 (47)			
CIMP ^W – low	0 (0)	3 (23)	15 (25)			
CIMP ^W – high	0 (0)	0 (0)	16 (27)	1.000	<0.001	<0.001
MSI+	2 (11)	0 (0)	13 (18)			
MSI-	17 (89)	13 (100)	46 (78)	0.125	<0.001	0.221
<i>BRAF</i> mutant	0 (0)	0 (0)	15 (25)			
<i>BRAF</i> wildtype	19 (100)	13 (100)	44 (75)	1.000	<0.001	<0.001
<i>KRAS</i> mutant	3 (16)	0 (0)	26 (44)			
<i>KRAS</i> wildtype	16 (84)	13 (100)	33 (56)	0.057	<0.001	0.014

L, CIMP-low; M, CIMP-mid; H, CIMP-high; LVI, lymphovascular invasion; EMVI, extramural vascular invasion; PNI, perineural invasion; TILS, tumour infiltrating lymphocytes; CIMP^W, classification of CIMP using the Weisenberger *et al* panel, whereby CIMP^W-high is defined as 3 or more methylated loci, CIMP^W-low as 1 or 2 methylated loci and CIMP^W-negative as no methylated loci. MSI, microsatellite instability; ¹Tumor location was unknown for 1 patient in CIMP-H, ²Number of distal CRC that were located at rectal site were 13, 1 and 11 in CIMP-L, CIMP-M and CIMP-H respectively, ³TILS data unknown for 3 patients in CIMP-L, ⁴*P* value for CIMP^W was generated from comparison between CIMP^W-high and CIMP^W-low or CIMP^W-negative.

Additional files

Additional file 1

File format: PowerPoint

Title: Unsupervised hierarchical clustering of 1505 probes (rows) in 28 normal colonic tissues (columns).

Description of data: Methylation of X-chromosome genes (enclosed within yellow rectangle) showed 100% correlation to gender as indicated by the red (female) and blue (male) bar above the heatmap.

Additional file 2

File format: PowerPoint

Title: Unsupervised hierarchical clustering of 1505 probes (rows) in 91 colorectal tumors (columns).

Description of data: Three tumor subgroups were revealed when methylation data from all 1,505 loci were analysed.

Additional file 3

File format: Excel spreadsheet

Title: 202 CpG sites differentially methylated between normal and tumour tissues.

Description of data: Methylation status at 202 CpG loci differentially methylated between normal and tumour tissues and presence of repetitive element or single nucleotide polymorphisms within probes.

Additional file 4

File format: PowerPoint

Title: Principal component analysis of 202 CpG loci that were differentially methylated between tumor and normal colonic tissue.

Description of data: This identified principal component 2 as the top ranking dimension and which explained 20% of the variability in the dataset. CIMP-H tumors are denoted in green, CIMP-M in black and CIMP-L in red.

Additional file 5

File format: Excel spreadsheet

Title: 112 genes differentially methylated between CIMP-H and CIMP-L.

Description of data: Known functions of 112 genes differentially methylated between CIMP-H and CIMP-L and their reported methylation in cancer and putative roles in gastrointestinal cancer.

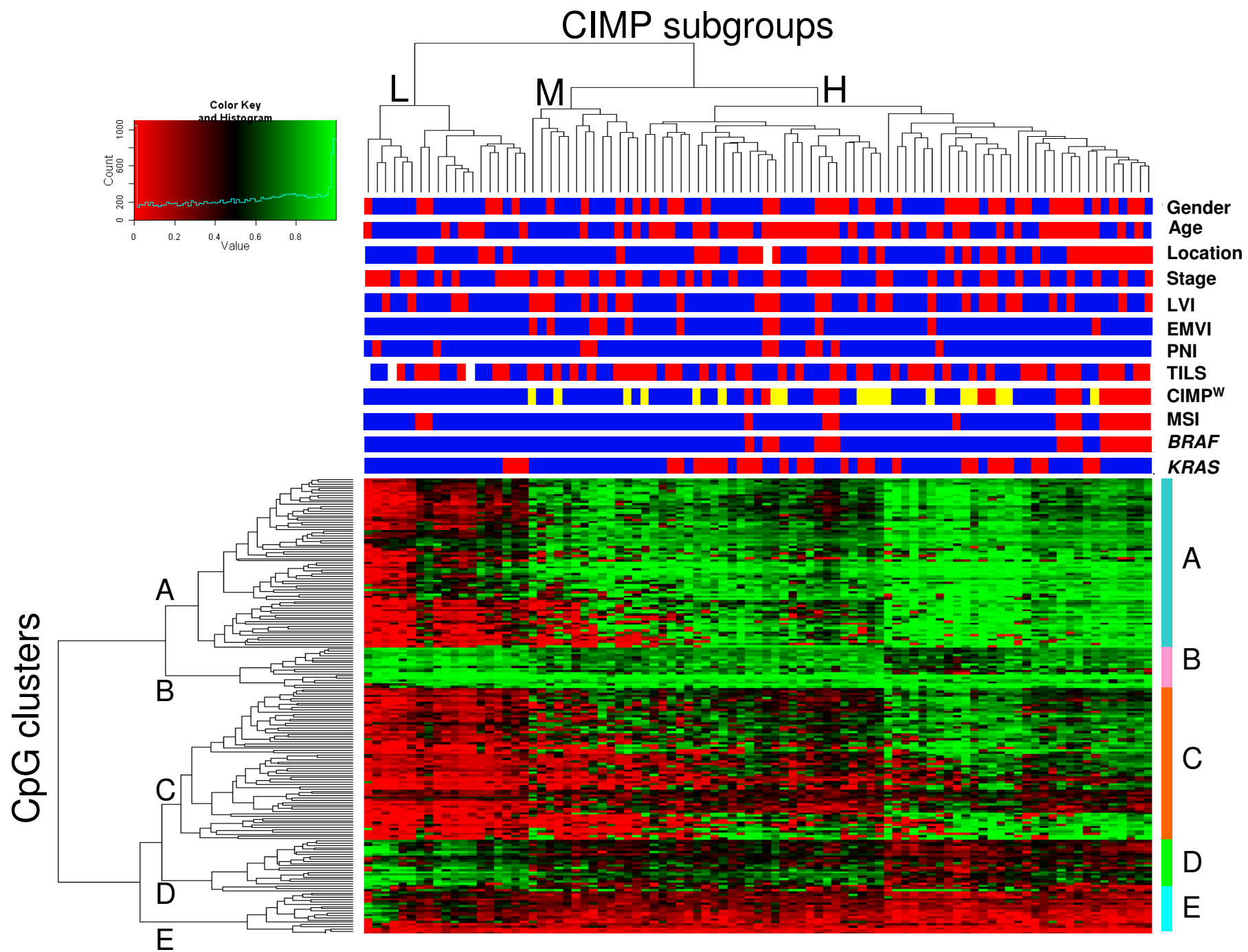


Figure 1

Additional files provided with this submission:

Additional file 1: Additional file 1_revised.ppt, 203K

<http://www.biomedcentral.com/imedia/1761232953393141/supp1.ppt>

Additional file 2: Additional file 2.ppt, 276K

<http://www.biomedcentral.com/imedia/1487394800321311/supp2.ppt>

Additional file 3: Additional file 3.xls, 64K

<http://www.biomedcentral.com/imedia/1162434971321311/supp3.xls>

Additional file 4: Additional file 4.ppt, 59K

<http://www.biomedcentral.com/imedia/7413021723213112/supp4.ppt>

Additional file 5: Additional file 5.xls, 132K

<http://www.biomedcentral.com/imedia/1930410284321311/supp5.xls>