Journal of Pathology
J Pathol 2010; **220**: 338–347
Published online 9 October 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/path.2640

Original Paper

Prognostic relevance of DNA copy number changes in colorectal cancer

George Poulogiannis,[1] Koichi Ichimura,[1] Rifat A Hamoudi,[1] Feijun Luo,[1] Suet Y Leung,[2] Siu T Yuen,[2] David J Harrison,[3] Andrew H Wyllie[1] and Mark J Arends[1]*

[1] *Department of Pathology, University of Cambridge, Cambridge, UK*
[2] *Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong*
[3] *Division of Pathology, Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, UK*

**Correspondence to: Mark J Arends, Department of Pathology, University of Cambridge, Cambridge, UK. E-mail: mja40@cam.ac.uk*

No conflicts of interest were declared.

Received: 6 April 2009
Revised: 19 August 2009
Accepted: 12 September 2009

Abstract

In a study of 109 colorectal cancers, DNA copy number aberrations were identified by comparative genomic hybridization using a DNA microarray covering the entire genome at an average interval of less than 1 Mbase. Four patterns were revealed by unsupervised clustering analysis, one of them associated with significantly better prognosis than the others. This group contained tumours with short, dispersed, and relatively few regions of copy number gain or loss. The good prognosis of this group was not attributable to the presence of tumours showing microsatellite instability (MSI-H). Supervised methods were employed to determine those genomic regions where copy number alterations correlate significantly with multiple indices of aggressive growth (lymphatic spread, recurrence, and early death). Multivariate analysis identified DNA copy number loss at 18q12.2, harbouring a single gene, *BRUNOL4* that encodes the Bruno-like 4 splicing factor, as an independent prognostic indicator. The data show that the different patterns of DNA copy number alterations in primary tumours reveal prognostic information and can aid identification of novel prognosis-associated genes.

Keywords: survival; colorectal cancer; microsatellite instability; array-comparative genomic hybridization; BRUNOL4; CELF4

Introduction

The molecular genetic abnormalities associated with sporadic colorectal carcinoma (CRC) reflect several events directly involved in carcinogenesis and more than one pathway of tumour evolution [1,2]. Less certain, however, is the extent to which patterns of genomic abnormality and gene expression are associated with tumour behaviour and hence have prognostic value. Consistent molecular biomarkers in primary CRC that predict the natural history of each tumour would be of immense value, both in planning individualized treatment regimes and in assessing the efficacy of new therapeutic agents. Hence many studies have sought prognostic associations with the transcriptional or mutational status of genes known or suspected to be associated with colorectal carcinogenesis, including *APC, KRAS, BRAF, TP53*, the TGFβ signalling pathway, members of the *BCL2* family, and several other single genes associated with proliferation, apoptosis, and stromal invasion [3–8]. Other approaches are based on identification of the pathway presumed responsible for the type of genetic instability displayed by CRCs, including defective DNA mismatch repair (leading to microsatellite instability) [9–11], inappropriate genome-wide patterns of promoter methylation [12–14], or multiple errors in chromosome number and structure [15–22]. Several such studies have supported the view, for example, that CRCs showing microsatellite instability tend to have a better prognosis than those with stable microsatellite DNA [10,11]; that tumours bearing *KRAS* mutations are associated with a poorer prognosis than those with wild-type *KRAS* [6,7]; and that allele losses within chromosome arms 8p, 17p, and 18q, together with gains within 8q and 20q, are associated with poor outcome and metastases [15,18–22]. There is, however, substantial controversy over the clinical applicability and even the reproducibility of many of these associations [21,23–25].

Pan-genomic array technology has permitted the exploration of prognostic indices that involve fewer initial assumptions about underlying mechanisms, whilst facilitating unbiased examination of the representation or expression of many thousands of genes

simultaneously. At the level of either transcript abundance or DNA copy number, these methods seek to identify clusters of properties that correlate with various types of tumour behaviour [26–28]. Unlike earlier cytogenetically-based methods, by which copy number abnormalities could be identified with confidence only if they extended over several megabases, array comparative genomic hybridization (array-CGH) permits DNA copy number gains or losses to be evaluated precisely at each clone site represented in the array, resolution being limited only by the density of the array itself.

Here, using array-CGH at around 1 Mbase resolution, on a set of more than 100 primary sporadic CRCs, we demonstrate the existence of several different patterns of DNA copy number abnormality (DCNA) and show that these provide prognostic information. Further analysis, of our own and others' data, reveals that this association of DCNA with prognosis relates to the probability with which specific genetic loci are involved. Fine mapping through one of these loci using a tiling path array provides proof of principle that this approach can lead to the identification of genes — previously unrecognized in colorectal carcinogenesis — whose copy number status is an independent prognostic variable. The data also provide an explanation for the uncertainty surrounding the prognostic value of classifying tumours on the basis of microsatellite instability.

Materials and methods

The laboratory work and most of the computation was performed in the Department of Pathology, University of Cambridge.

Colorectal tumour clinical samples and DNA extraction

Clinicopathological characteristics of the 109 CRC patients included in the current study are summarized in Table 1. Several sets of consecutive, presumptively curative, resection specimens were collected between 1990 and 1994 at the Royal Infirmary, Edinburgh (74 in all), and between 1991 and 1998 at Queen Mary Hospital, Hong Kong (HK) (35 in all). Additional criteria met by these specimens included availability of follow-up data, matched fresh-frozen and paraffin-embedded material, and good quality DNA as assessed by PCR amplification and adequate signal-to-noise ratios with low standard deviations of the array-CGH data. The HK series also included exact information on treatment conditions (adjuvant chemotherapy and/or radiotherapy or no therapy). The minimum clinical follow-up was 30 months, with a median of 67 months. Tumour and corresponding non-neoplastic colonic tissues were snap-frozen immediately on receipt from the surgical theatres and stored at −80 °C for subsequent DNA extraction by standard methods [29]. In cases where normal mucosa was not available, DNA was extracted from blood leukocytes. For most patients, DNA was also extracted from formalin-fixed, paraffin-embedded tumour and normal tissue for MSI analysis, following the QIAamp DNA mini-kit tissue protocol (Qiagen, Dorking, UK). Local Institutional Ethical Committees have formally

Table 1. Clinicopathological and genetic features of the sporadic CRCs

Characteristic	Number	Percentage (%)
Tumours included in the study	109	
Males/females	54/55	49.5/50.5
Age (years)		
≤55/>55	28/80	25.9/74.1
Median/mean/range	67/64.1/30–90	
Dukes' staging		
A	8	7.5
B	48	44.9
C	42	39.2
D	9	8.5
Primary site		
Caecum	6	5.6
Ascending colon	5	4.6
Hepatic flexure	5	4.6
Transverse	4	3.7
Descending colon	3	2.8
Sigmoid	26	24.1
Recto-sigmoid	2	1.8
Rectum	57	52.8
Tumour side		
Right/left	20/88	18.5/81.5
Histological type		
Adenocarcinoma	92	84.4
Mucinous	10	9.2
Mucinous and adenocarcinoma	5	4.6
Mucinous and signet ring	2	1.8
Signet ring	0	0
Degree of differentiation		
Well	13	11.9
Moderate	87	79.8
Poor	9	8.3
Overall survival data (months)		
Median/mean/range	28.7/41.4/0.3–147.2	
Dead/alive	57/50	53.3/46.7
Recurrence status		
Recurrence/no recurrence	40/64	38.5/61.5
MSI status		
MSS	93	85.3
MSI-L	3	2.8
MSI-H	13	11.9
Immunohistochemistry for MMR proteins		
MLH1-negative	10	9.2
MSH2-negative	1	0.9
MSH6-negative	2	1.8

MSI = microsatellite instability; MSS = microsatellite-stable; MSI-L = low-frequency MSI; MSI-H = high-frequency MSI.

accepted these studies on the basis of consent and anonymization protocols in force at the time of sample collection.

Analysis of microsatellite instability (MSI)

MSI was assessed by comparison of paired tumour and normal mucosa samples using ten microsatellite loci including the five recommended by the National Cancer Institute Bethesda workshop on MSI [30]. Seven dinucleotide repeats (ACTC, D5S107, D5S406, D13S153, D2S123, D5S346, and D17S250) and three mononucleotide tracts (BAT25, BAT26 and BAT40) were amplified by polymerase chain reaction, with one fluorescently labelled primer (MWG Biotech AG) as previously described [31,32]. Tumours were classified as showing microsatellite instability at high frequency (MSI-H) if three or more microsatellite markers showed variation in length between tumour and paired control DNA. Tumours with length variation detected at only one or two loci were scored as MSI-L (indicating microsatellite instability at low frequency) and were grouped for analysis with microsatellite-stable (MSS) tumours, defined as having no length variation across all ten markers [31].

Immunohistochemistry

Formalin-fixed, paraffin-embedded tumours were analysed by immunohistochemistry using mouse monoclonal antibodies against hMLH1 (PharMingen, MO52537), hMSH2 (Calbiochem, NA 27–100), and hMSH6 (Abcam, ab14204) as previously described [31,33,34].

Array comparative genomic hybridization

The 0.97 Mb resolution array platform was constructed as previously described [35] using clones obtained from the Wellcome Trust Sanger Institute (Hinxton, UK). For each hybridization, 500 ng of tumour and reference genomic DNA were labelled with either Cy5-dCTP or Cy3-dCTP (Amersham Biosciences, Little Chalfont, UK) using a BioPrime kit (Invitrogen, Paisley, UK). The reference DNA consisted of a pool of mixed female or male normal peripheral blood leukocyte DNA from 20 unrelated individuals. Labelled samples were hybridized; the arrays were scanned using GenePix Pro 5.1 on a GenePix 4100A personal scanner (Axon Instruments, Union City, CA, USA); and normalization and data analysis were performed as previously described [35,36].

Statistical analysis

Categorical variables (patient age class, gender and recurrence status, and tumour stage, location, histological type, extent of mucin production, degree of differentiation, and MSI status) were compared using chi-squared contingency tests. Yates' or Bonferroni corrections were not applied, as all identified correlations have previously been confirmed in large series studies [9,37,38]. Kaplan–Meier curves were used to estimate the time-related probabilities of survival among patients with CRCs of different genetic, clinical, and histopathological characteristics, and the log-rank test was used to estimate the significance of differences between survival distributions.

Copy number differences between categorical variables were compared using the aCGH package in R 2.8.1 and a two-sample-based t-statistic incorporating permutation-adjusted p values for step-down multiple testing procedures [39,40]. This analysis was performed on those autosomal clone loci that were gained or lost in at least 10% of the samples. The distributions of DNA copy number alterations among different clusters were compared with one-way ANOVA. Unsupervised hierarchical clustering was carried out on the log2 Cy5/Cy3 fluorescence ratio data using the Euclidean distance metric and Ward's linkage algorithm.

Array-CGH microarray data

The pan-genomic 1 Mb and the chromosome 18 tiling path array-CGH data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through accession number GSE12520.

Results

Clinicopathological features of sporadic CRCs

In terms of location within the colon and rectum, histopathological features, grade, stage (including the presence of lymph node metastasis), gender association, and incidence of high frequency microsatellite instability (MSI-H), this set of 109 tumours closely resembles most other reported series of sporadic CRC (Table 1). The great majority (~84%) were staged as Dukes' B or C, with the expected correlation with survival (Supporting information, Supplementary Figure 1A). Thirteen (12%) exhibited MSI-H on DNA analysis. Of these, ten lacked immunohistochemical expression of hMLH1 in tumour nuclei, whilst one lacked hMSH2 and two hMSH6. The apparent trend towards better survival associated with MSI-H did not reach statistical significance ($p = 0.071$; Supporting information, Supplementary Figure 1C), but the frequency of recurrence in these patients was less than in the MSI-L/MSS group ($p = 0.013$), observations similar to those reported in larger series of CRC [27]. Over the complete set of CRCs, as expected, there was a strong inverse relationship between survival and clinical history of recurrence ($p < 0.0001$) (Supporting information, Supplementary Figure 1B). The relative weakness of commonly used prognostic indices is also obvious, however. Thus, of patients with Dukes' B tumours, 13/43 (~30%) died within 5 years of diagnosis, whilst at least 17% of patients with Dukes'

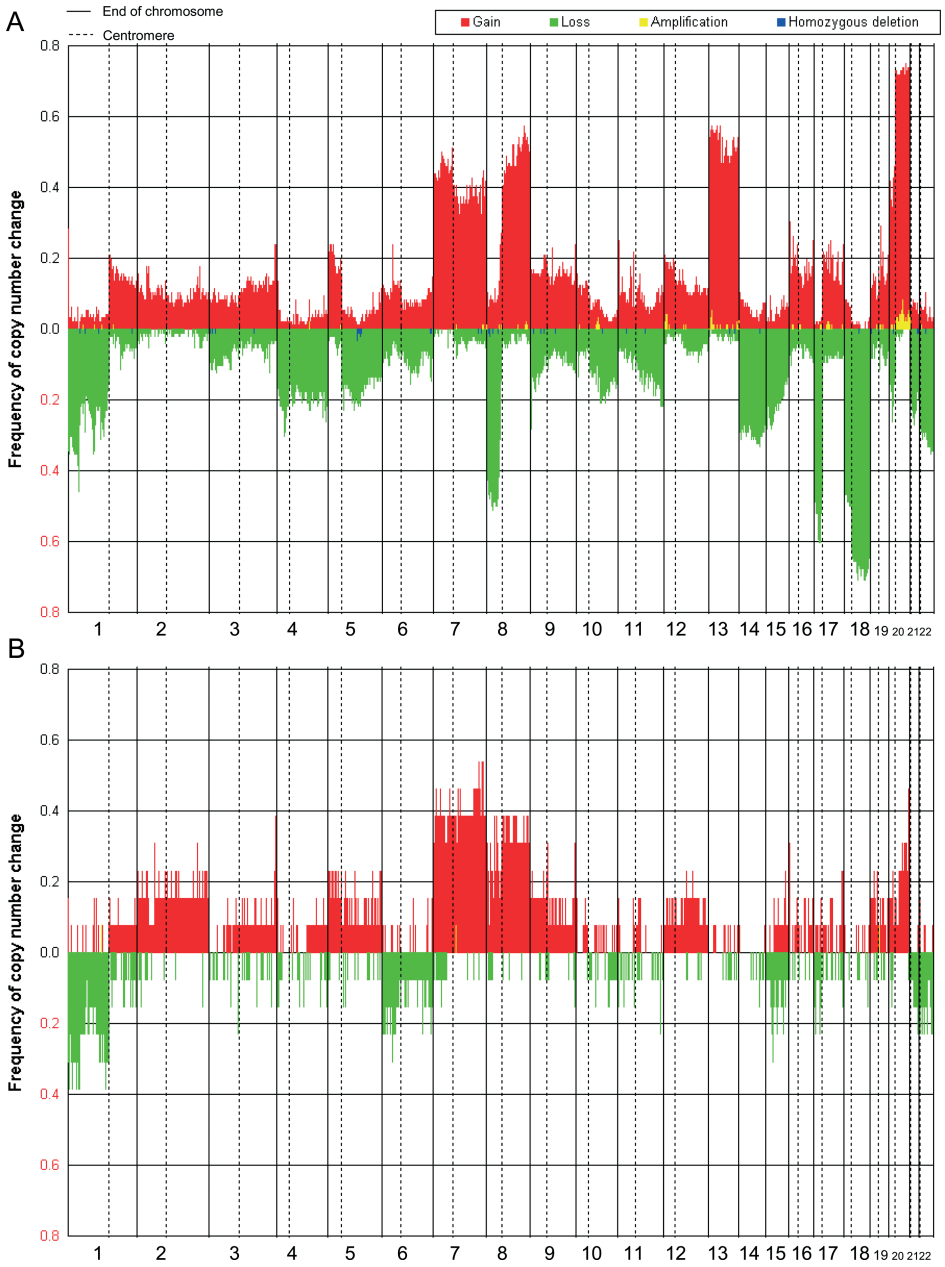

Figure 1. Comparison of the overall frequencies of DNA copy number alterations (*y*-axis) as assessed by 1 Mb array-CGH between (A) microsatellite-stable (MSS) ($n = 96$) (top) and microsatellite-unstable (MSI-H) ($n = 13$) sporadic CRCs (B). Losses are shown in green; gains in red. Data are presented ordered by chromosomal map position of the clones (*x*-axis)

C tumours appeared to be symptom-free at the last recorded follow-up. Similarly, 6/12 (50%) patients with MSI-H tumours died within 5 years of diagnosis.

DNA copy number alterations differ in frequency and site between MSI-H and MSS CRCs

DCNAs were less frequent in MSI-H tumours than in MSS/MSI-L tumours, as assessed by 1 Mb array-CGH. Thus, the mean number of clone sites in the array that revealed tumour DCNAs was 167 for MSI-H and 383 for MSS tumours ($p = 0.002$, one-way ANOVA). The most notable difference was in copy number loss, which affected a mean of 59 clone sites for MSI-H and 184 for MSS/MSI-L tumours ($p = 0.003$). None of the MSI-H tumours showed homozygous deletions, whilst only one showed genomic amplification. However, no MSI-H tumour was completely free from copy number alterations.

There were conspicuous differences between MSI-H and MSS tumours in the frequency with which copy number alterations extended over large portions of chromosome arms (Figures 1 and 2), notably in gains within 13q (nil MSI-H, 50% MSS) and 20q (20% MSI-H, 70% MSS), and losses within 8p (nil MSI-H, 50% MSS), 14q (nil MSI-H, 30% MSS), 17p (10% MSI-H, 50% MSS), 18p (nil MSI-H, 50% MSS), and 18q (nil MSI-H, 70% MSS). The losses within 18q provided the largest difference between MSI-H and MSS tumours ($p < 0.05$ even after adjustment for multiple comparisons). In contrast, however, both MSI-H and MSS tumours exhibited

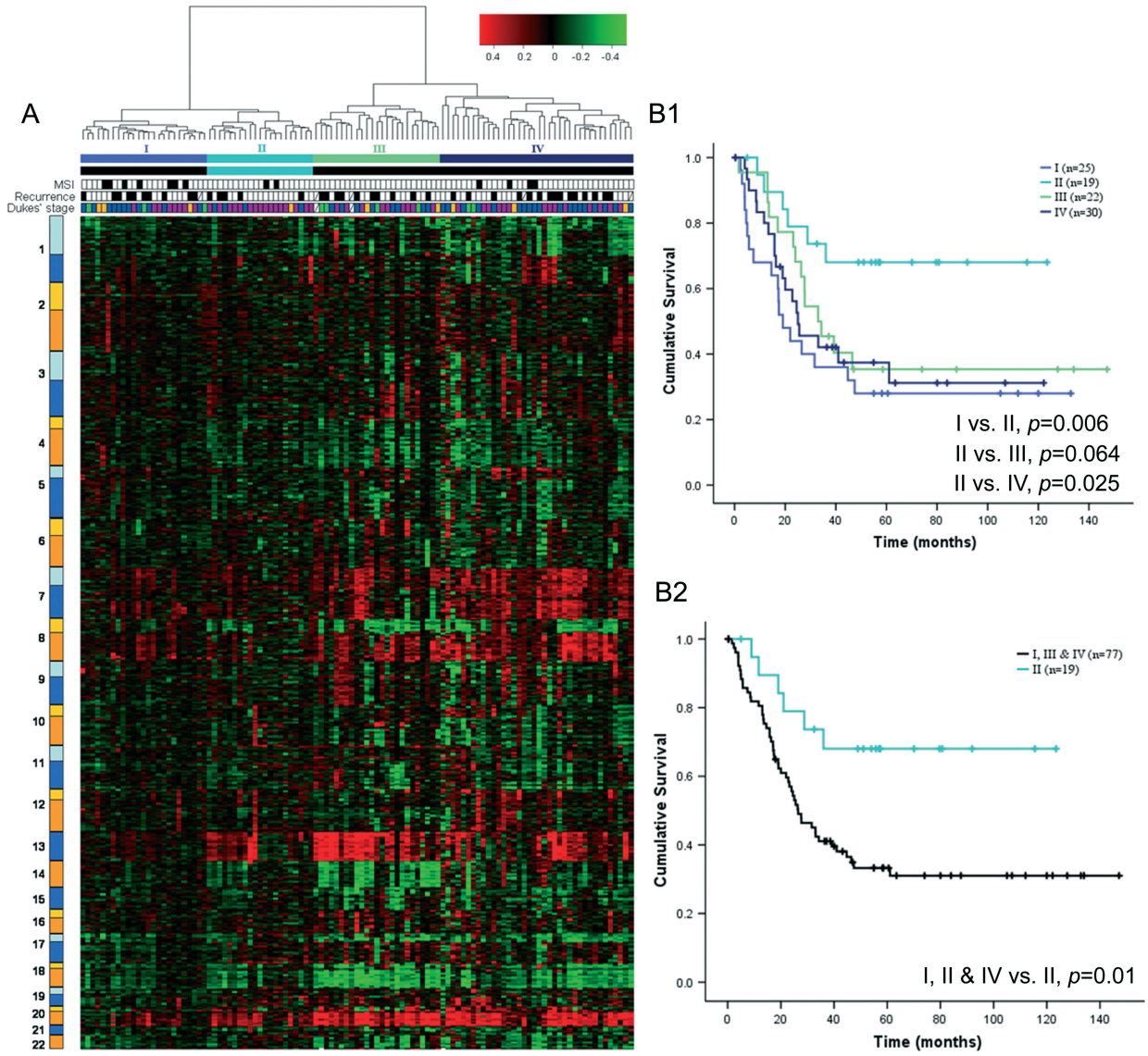

Figure 2. (A) Unsupervised hierarchical clustering into four groups (I–IV; blue, turquoise, light green, dark purple) of genomic profiles (rows) on 109 sporadic CRCs (columns). Red and green cytobands indicate DNA copy number gain and loss, respectively. Boxes below the dendrogram indicate MSI status: MSS (white), (MSI-H) (black); recurrence status: no recurrence (white), recurrence (black); Dukes' stage: A (yellow), B (purple), C (blue), D (green). Diagonal slash: data not available. (B1) Kaplan–Meier survival curves of the four groups, I–IV (in blue, turquoise, light green, and dark purple, respectively). (B2) Survival curves of group II (turquoise) against combined groups I, III, and IV (black). Log-rank statistical analysis *p* values are provided

similar frequencies of losses in 1p and gains in 7p, 7q, and 8q.

Unsupervised clustering analysis reveals sporadic CRCs with different DNA copy number profiles and survival

Unsupervised hierarchical clustering based on the log2 tumour/normal ratio for each of the 2904 clones of the array was applied to the entire set of 109 tumours. The data, ordered by genome position, are shown in Figure 2A. Clustering analysis identified four subsets, referred to here as groups I, II, III, and IV. These showed both different patterns and significantly different overall frequencies of DCNA (Table 2).

Thus, groups III and IV showed many copy number alterations that extended over most or all of chromosome arms 1p, 8p, 17p, and 18q, and gains in 7, 8q, 13, and 20q. In contrast, groups I and II showed DCNAs at fewer, and usually smaller, sites. The proportions of tumours in groups I, II, III, and IV were 23%, 19%, 23%, and 35%, respectively. Groups I and II differed

Table 2. Mean DNA copy number alterations among the four groups obtained from unsupervised clustering analysis

Cluster group	Loss	Gain	Breakpoint	Total
Group I	78	104	178	182
Group II	103	122	201	225
Group III	256	232	292	488
Group IV	207	253	328	460
p value	<0.0001	<0.0001	<0.0001	<0.0001

Figures are mean copy number alterations per genome, in each group, as read from the array hybridization pattern, and compared by one-way ANOVA. Breakpoints: as defined in the text. Total = losses + gains.

from each other in that group II had slightly more DCNAs (both gains and losses) and breakpoints per genome than group I (where a 'breakpoint' is defined as the point at which the DNA copy number undergoes change between adjacent clone sites). MSI-H tumours were found in groups with both high and low frequency of DCNAs. Thus, seven of the 13 MSI tumours appeared in group I (where they represented 28% of the total in that group), two in group II (9%), and four in group IV (11%). No statistically significant relationship was found between the histopathological features of the tumours and any of the four cluster groups.

Although unsupervised clustering failed to stratify tumours according to their microsatellite stability, it clearly identified group II as having significant over-representation of lymph node-negative tumours (ie Dukes' stage B) relative to the other groups ($p < 0.005$, <0.02, and <0.003 relative to groups I, III, and IV, respectively; chi-squared test) and better survival than the other groups ($p = 0.01$, log-rank test; Figure 2B). Removal of the 13 MSI-H tumours from comparisons between the four groups increased the significance of the difference in prognosis between group II and the others. Thus, the better prognosis associated with group II tumours cannot be attributed to its content of MSI-H cancers.

To establish the generality of this relationship between DCNA and prognosis, the same unsupervised hierarchical clustering analysis was applied to data from a published study from Korea, in which 59 CRCs had been studied by array-CGH at a very similar resolution to our own [24]. Clustering analysis also divided these tumours into similar subgroups, with a statistically significant relationship to survival ($p = 0.04$). In the light of this evidence that groups of CRCs with differing prognosis can be distinguished on the basis of their patterns of DCNA, a statistical analysis was implemented to test the association of DCNA with categorical, clinical outcomes across our entire series of 109 tumours, with the aim of identifying key alterations that might detect those patients most at risk from aggressively growing tumours.

DNA copy number alterations at discrete chromosomal locations are associated with recurrence, survival, and lymph node status

The total tumour set was divided into two roughly equal subsets for each of three clinicopathological prognostic criteria: documented tumour recurrence in the observation period ($n = 40$) versus non-recurrence ($n = 64$); Dukes' C stage ($n = 42$) versus Dukes' B ($n = 48$); and patients deceased within 5 years of diagnosis ($n = 43$) versus those censored within the first 5 years or dying only after 5 years ($n = 56$). The frequencies of significant DNA copy number gains and losses among these groups are plotted in Figures 3A–3C against their chromosomal position. Also shown, for each chromosomal position, is the magnitude of the t-statistic, which reflects the significance of the difference between these frequencies. The chromosomal positions at which there are statistically significant differences between each set of comparisons are depicted in Figures 3A–3C and summarized in Figure 3D. The most significant associations with recurrence are copy number losses within 14q and 18q. Risk of death within 5 years of diagnosis is also associated significantly with copy number loss at these sites, together with loss within 6p and gains in 13q. Lymph node involvement (Dukes' C versus Dukes' B), in addition to these, also correlates positively with copy number gains in 2p, 4p, 7q, and 16q, and with losses in 1p, 10q, 11q, 12q, 17q, 18q, 19q, and 22. A list was then compiled of the clone sites at which significant differences were recorded in DNA copy number between the supervised groups in all three indices of aggressive growth (lymph node-positive, and recurrence and death within 5 years), in an attempt to identify associations with multiple components of poor prognosis. This list comprises just four clone sites (all losses) in 6p22.1 (RP1-265C24), 14q24.3 (RP3-449M8), 18q12.2 (RP11-19F9), and 18q21.1 (RP11-748M14) (Table 3).

Multivariate analysis using a Cox-proportional model with time-dependent covariates was performed to assess the independent prognostic significance of the clone abnormalities shown in Table 3. This analysis showed that, of these clone sites, only the DNA copy number loss of clone RP11-19F9, located on chromosome 18q12.2, is significantly associated with poorer survival independently of other prognostically relevant cofactors such as the age of the patient or the stage of the tumour (Supporting information, Supplementary Table 2). Kaplan–Meier analysis also confirmed that when the patients are compared on the basis of the DNA copy number status at the RP11-19F9 clone site alone, there is a statistically significant difference in survival ($p = 0.007$, log-rank test), those with copy number loss showing a poorer prognosis (Supporting information, Supplementary Figure 2F).

A tiling-path array of chromosome 18 was used to fine-map the prognostically relevant locus within 18q12.2 (RP11-19F9). Thirty samples were screened (26 with known copy number losses in RP11-19F9 and four without as controls). The commonest region of deletion was shown to span clones RP11-94N4 and RP11-234D3, both of which lie within a single gene, *BRUNOL4*. In total, nine (30%) of the 30 screened tumours showed focal deletions that included this gene, while the remainder showed monosomy of chromosome 18 or loss of 18q (Supporting information, Supplementary Figure 2).

Discussion

DNA copy number alterations in primary CRC predict prognosis

Our study is based on the hypothesis that appropriate analysis of primary CRCs is capable of providing

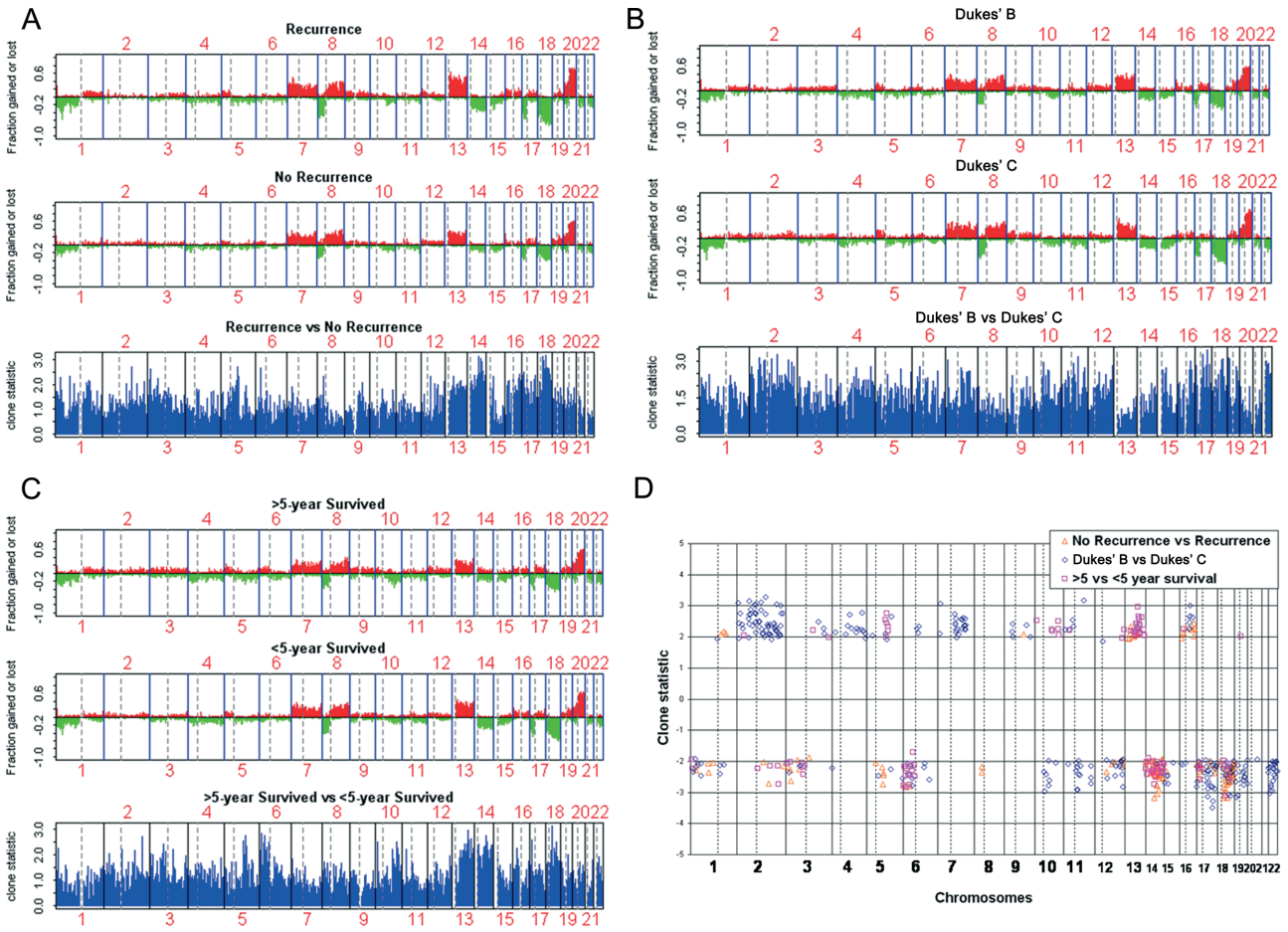

Figure 3. Frequency plots of the array-CGH data displaying on the *y*-axis DNA copy number changes (gains in red; losses in green) and the height of the statistic (in blue) indicating the strength of significance of the difference in frequency between the two sets of comparison for each clone (by chromosome position on the *x*-axis). The categorical variables which are compared are in the order of (A) recurrence versus no recurrence, (B) Dukes' B versus Dukes' C, and (C) >5 years versus <5 years survival. (D) Scatter plot indicating a summary of the *t*-statistic height (*y*-axis) of the clones showing significant differences ($p < 0.05$) from all the different sets of comparisons

Table 3. Genomic copy number changes associated with all indices of poor prognosis

Clone name	Band	Change	Genes
RP1-265C24	6p22.1	Loss	Histone genes, *OR2B2*, *OR2B6*, *ZNF165*, *ZNF435*, *ZNF192*, *ZNF193*, *ZKSCAN4*, *NKAPC*, *PGBG1*, *ZNF323*, *ZKSCAN3*, *ZSCAN12*, *ZSCAN23*, *GPX6*, *GPX5*, *ZNF452*
RP3-449M8	14q24.3	Loss	*ENTPD5*, *C14orf45*, *ALDH6A1*, *LIN52*, *FAM161B*, *VSX2*, *ABCD4*, *C14orf115*, *TMEM90A*, *NPC2*, *ISCA2*, *LTBP2*, *KIAA0317*, *AC007956.5*, *FCF1*, *YLPM1*, *PROX2*, *RPS6KL1*, *PGF*
RP11-19F9	18q12.2	Loss	*KIAA1328*, *BRUNOL4*
RP11-748M14	18q21.1	Loss	*SMAD2*, *ZBTB7C*

tumour-specific prognostic information. The need for such analysis is particularly acute in the majority group of Dukes' B and C carcinomas, where conventional clinicopathological indices are frequently misleading. Several genome-wide analyses have already demonstrated an association between the risk of metastases or recurrence and the patterns in the primary tumour of either mRNA transcripts or miRNA [40–46] or DCNA [47–50]. Some data indicate that patterns of DCNA may even predict the probability of metastases to local and regional lymphatics, as opposed to blood-borne metastases to the liver and lung [49,51]. Only a minority of these studies describe convincingly large numbers of cases, however, and until recently, most published genome-wide studies of DCNA have been limited by the low resolution of chromosome CGH. Here, in a pan-genomic analysis of over 100 CRCs, initially using an unsupervised clustering algorithm, we have demonstrated that patterns of DCNA do indeed distinguish tumours with differing outcomes. In a series of tumours that appears concordant in many criteria with CRCs reported from centres across the world, we identified a good prognosis group (here called group II) that accounts for around 20% of all CRCs. Most of the tumours in this group were microsatellite-stable but showed dispersed, relatively sparse, short-length regions of DNA copy number alterations. Many, but not all, were node-negative.

Consistent DNA copy number alterations reveal both familiar and novel cancer-related sites

In supervised analyses, we identified a significant association between DNA copy number loss at four genomic regions (6p22.1, 14q24.3, 18q12.2, 18q21.1) and aggressive growth behaviour (Figure 3 and Table 3). Reassuringly, similar copy number loss in three of these chromosome regions has already been incriminated in the initiation or progression of malignant CRCs in a meta-analysis of (relatively low-resolution) chromosome CGH data from 373 tumours [49]. Furthermore, the most significant differences in DCNAs recorded in the present study, in comparisons between patient survival of less than or more than 5 years, coincide precisely with seven regions in chromosomes 18q, 14q (both losses), and 13q (gains) identified as associated with poor prognosis in the recently published study of 59 Korean CRCs [24] (Supporting information, Supplementary Table 1). Amongst these is copy number loss involving *SMAD2*, a gene in 18q encoding an element in the TGFβ signalling pathway, for which the most profound down-regulation of several candidates in 18q [47] and an association with poor prognosis [6] have already been reported in CRC. However, DCNAs that map within the *SMAD2* gene (at clone RP11-748M14), as well as most of the prognostically relevant loci shown in Table 3, prove not to be independent prognostic factors.

In contrast, multivariate analysis identified the copy number loss of RP11-19F9 as the only independent prognostic indicator in 18q represented in our 1 Mb array. Fine-mapping of the prognostically relevant locus using tiling-path array-CGH of chromosome 18 pointed to a single gene in 18q12.2, *BRUNOL4* (also known as *CELF4*), as the target of deletion in all cases. Bruno-like 4 is a factor that controls alternative splicing of the insulin receptor, cardiac troponin, and the microtubule-associated protein TAU [52]. Its status has not been associated with CRC before, to our knowledge, but its expression has been reported to be associated with paclitaxel sensitivity in breast cancer [53].

All CRCs show abnormal DNA copy number at multiple loci

The data described here emphasize, in primary colorectal cancer, the widespread incidence of DNA copy number abnormalities. Even MSI-H tumours, which often appear near-diploid on the basis of karyotypic or chromosomal CGH evidence, invariably show losses and gains at several loci. In particular, amongst those MSI-H tumours with the highest numbers of DNA copy number abnormalities, several show preferential losses of the same chromosome arms as those similarly affected in MSS tumours (1p−, 7+, 8q+), suggesting that copy number changes affecting these chromosome arms may be part of a pattern associated with the initiation of colorectal cancer in general, rather than the result of a specific carcinogenic pathway. Equally, CRCs with short, dispersed regions of copy number abnormality include both MSI-H and MSS tumours, but only a subset of these falls within the good prognosis group. This may provide an explanation for the long-running uncertainty over the prognostic significance of MSI-H versus MSS status.

In summary, this large study of the frequency and location of DNA copy number gains and losses in CRC shows that the patterns of these alterations relate to prognosis. The reasons for this association appear complex. The patterns themselves presumably reflect the probability that certain types of repeated genomic error are tolerated by the evolving tumour clones. Such errors may well affect genes indiscriminately, but the frequency with which specific, critical targets are affected is likely to determine the various properties responsible for aggressive growth. The data presented here provide proof of principle that fine-resolution study of DCNAs can lead to the identification of such targets, including genes whose connection with cancer was previously unsuspected.

Acknowledgements

This work was supported by the Addenbrooke's Charitable Trust, the John Lucas Walker Fund, and a Research Award from the Pathological Society of Great Britain and Ireland. We are very grateful to Professor Yeun-Jun Chung, College of Medicine, Catholic University of Korea, Seoul, for providing access to his group's original data, cited in this paper

References

1. Fearon ER, Gruber SB. Molecular abnormalities in colon and rectal cancer. In *The Molecular Basis of Cancer*, (2nd edn), Mendelsohn J, Howley PM, Israel MA, Liotta LA (eds), W.B. Saunders, Philadelphia, 2001; 289–312.
2. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, *et al*. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;**318:**1108–1113.
3. Rohde F, Rimkus C, Friederichs J, Rosenberg R, Marthen C, Doll D, *et al*. Expression of osteopontin, a target gene of deregulated Wnt signaling, predicts survival in colon cancer. *Int J Cancer* 2007;**121:**1717–1723.
4. Parsons DW, Wang TL, Samuels Y, Bardelli A, Cummins JM, DeLong L, *et al*. Colorectal cancer: mutations in a signalling pathway. *Nature* 2005;**436:**792.
5. Conlin A, Smith G, Carey FA, Wolf CR, Steele RJ. The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. *Gut* 2005;**54:**1283–1286.
6. Xie W, Rimm DL, Lin Y, Shih WJ, Reiss M. Loss of Smad signaling in human colorectal cancer is associated with advanced disease and poor prognosis. *Cancer J* 2003;**9:**302–312.
7. Esteller M, Gonzalez S, Risques RA, Marcuello E, Mangues R, Germa JR, *et al*. K-ras and p16 aberrations confer poor prognosis in human colorectal cancer. *J Clin Oncol* 2001;**19:**299–304.
8. Sinicrope FA, Rego RL, Okumura K, Foster NR, O'Connell MJ, Sargent DJ, *et al*. Prognostic impact of bim, puma, and noxa expression in human colon carcinomas. *Clin Cancer Res* 2008;**14:**5810–5818.
9. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. *Science* 1993;**260:**816–819.
10. Samowitz WS, Curtin K, Ma KN, Schaffer D, Coleman LW, Leppert M, *et al*. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol Biomarkers Prev* 2001;**10:**917–923.

11. Malesci A, Laghi L, Bianchi P, Delconte G, Randolph A, Torri V, *et al*. Reduced likelihood of metastases in patients with microsatellite-unstable colorectal cancer. *Clin Cancer Res* 2007;**13:**3831–3839.
12. van Rijnsoever M, Grieu F, Elsaleh H, Joseph D, Iacopetta B. Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. *Gut* 2002;**51:**797–802.
13. Hawkins N, Norrie M, Cheong K, Mokany E, Ku SL, Meagher A, *et al*. CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability. *Gastroenterology* 2002;**122:**1376–1387.
14. Whitehall VL, Walsh MD, Young J, Leggett BA, Jass JR. Methylation of O-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability. *Cancer Res* 2001;**61:**827–830.
15. Macartney-Coxson DP, Hood KA, Shi HJ, Ward T, Wiles A, O'Connor R, *et al*. Metastatic susceptibility locus, an 8p hot-spot for tumour progression disrupted in colorectal liver metastases: 13 candidate genes examined at the DNA, mRNA and protein level. *BMC Cancer* 2008;**8:**187.
16. Wang Z, Cummins JM, Shen D, Cahill DP, Jallepalli PV, Wang TL, *et al*. Three classes of genes mutated in colorectal cancers with chromosomal instability. *Cancer Res* 2004;**64:**2998–3001.
17. Diep CB, Thorstensen L, Meling GI, Skovlund E, Rognum TO, Lothe RA. Genetic tumor markers with prognostic impact in Dukes' stages B and C colorectal cancer patients. *J Clin Oncol* 2003;**21:**820–829.
18. Chang SC, Lin JK, Yang SH, Wang HS, Li AF, Chi CW. Relationship between genetic alterations and prognosis in sporadic colorectal cancer. *Int J Cancer* 2006;**118:**1721–1727.
19. Hidaka S, Yasutake T, Takeshita H, Kondo M, Tsuji T, Nanashima A, *et al*. Differences in 20q13.2 copy number between colorectal cancers with and without liver metastasis. *Clin Cancer Res* 2000;**6:**2712–2717.
20. Fujita S, Baba H, Yamamoto S, Akasu T, Moriya Y, Sugano K. Allelic status of chromosomes 17p, 18q, 22q [corrected] 3p and their clinical usefulness in colorectal cancer. *Anticancer Res* 2006;**26:**2833–2840.
21. Popat S, Zhao D, Chen Z, Pan H, Shao Y, Chandler I, *et al*. Relationship between chromosome 18q status and colorectal cancer prognosis: a prospective, blinded analysis of 280 patients. *Anticancer Res* 2007;**27:**627–633.
22. Postma C, Terwischa S, Hermsen MA, van der Sijp JR, Meijer GA. Gain of chromosome 20q is an indicator of poor prognosis in colorectal cancer. *Cell Oncol* 2007;**29:**73–75.
23. Arango D, Laiho P, Kokko A, Alhopuro P, Sammalkorpi H, Salovaara R, *et al*. Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. *Gastroenterology* 2005;**129:**874–884.
24. Kim MY, Yim SH, Kwon MS, Kim TM, Shin SH, Kang HM, *et al*. Recurrent genomic alterations with impact on survival in colorectal cancer identified by genome-wide array comparative genomic hybridization. *Gastroenterology* 2006;**131:**1913–1924.
25. Anwar S, Frayling IM, Scott NA, Carlson GL. Systematic review of genetic influences on the prognosis of colorectal cancer. *Br J Surg* 2004;**91:**1275–1291.
26. Eschrich S, Yang I, Bloom G, Kwong KY, Boulware D, Cantor A, *et al*. Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol* 2005;**23:**3526–3535.
27. Kim GP, Colangelo LH, Wieand HS, Paik S, Kirsch IR, Wolmark N, *et al*. Prognostic and predictive roles of high-degree microsatellite instability in colon cancer: a National Cancer Institute–National Surgical Adjuvant Breast and Bowel Project Collaborative Study. *J Clin Oncol* 2007;**25:**767–772.
28. Goel A, Arnold CN, Niedzwiecki D, Chang DK, Ricciardiello L, Carethers JM, *et al*. Characterization of sporadic colon cancer by patterns of genomic instability. *Cancer Res* 2003;**63:**1608–1614.
29. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, *et al*. *Short Protocols in Molecular Biology*. John Wiley: New York, 1995; Ch. 2.4.
30. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, *et al*. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 1998;**58:**5248–5257.
31. Frayling IM, Happerfield L, Mattocks C, Oakhill K, Arends MJ. Application of Molecular Diagnostics to Hereditary Nonpolyposis Colorectal Cancer. In *Molecular Diagnostics, for the Clinical Laboratorian* (2nd edn), Coleman WB, Tsongalis GJ (eds). Humana Press, New Jersey, 2005; 375–392.
32. Liu B, Nicolaides NC, Markowitz S, Willson JK, Parsons RE, Jen J, *et al*. Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nature Genet* 1995;**9:**48–55.
33. Frayling IM. Methods of molecular analysis: mutation detection in solid tumours. *Mol Pathol* 2002;**55:**73–79.
34. Garrity MM, Burgart LJ, Mahoney MR, Windschitl HE, Salim M, Wiesenfeld M, *et al*. Prognostic value of proliferation, apoptosis, defective DNA mismatch repair, and p53 overexpression in patients with resected Dukes' B2 or C colon cancer: a North Central Cancer Treatment Group Study. *J Clin Oncol* 2004;**22:**1572–1582.
35. McCabe MG, Ichimura K, Liu L, Plant K, Backlund LM, Pearson DM, *et al*. High-resolution array-based comparative genomic hybridization of medulloblastomas and supratentorial primitive neuroectodermal tumors. *J Neuropathol Exp Neurol* 2006;**65:**549–561.
36. Jones DT, Ichimura K, Liu L, Pearson DM, Plant K, Collins VP. Genomic analysis of pilocytic astrocytomas at 0.97 Mb resolution shows an increasing tendency toward chromosomal copy number change with age. *J Neuropathol Exp Neurol* 2006;**65:**1049–1058.
37. Gryfe R, Kim H, Hsieh ET, Aronson MD, Holowaty EJ, Bull SB, *et al*. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N Engl J Med* 2000;**342:**69–77.
38. Lothe RA, Peltomaki P, Meling GI, Aaltonen LA, Nystrom-Lahti M, Pylkkanen L, *et al*. Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. *Cancer Res* 1993;**53:**5849–5852.
39. Benatti P, Gafa R, Barana D, Marino M, Scarselli A, Pedroni M, *et al*. Microsatellite instability and colorectal cancer prognosis. *Clin Cancer Res* 2005;**11:**8332–8340.
40. Lassmann S, Weis R, Makowiec F, Roth J, Danciu M, Hopt U, *et al*. Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal- and microsatellite-unstable sporadic colorectal carcinomas. *J Mol Med* 2007;**85:**293–304.
41. Camps J, Armengol G, del Rey J, Lozano JJ, Vauhkonen H, Prat E, *et al*. Genome-wide differences between microsatellite stable and unstable colorectal tumors. *Carcinogenesis* 2006;**27:**419–428.
42. Jones AM, Douglas EJ, Halford SE, Fiegler H, Gorman PA, Roylance RR, *et al*. Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene* 2005;**24:**118–129.
43. Yamasaki M, Takemasa I, Komori T, Watanabe S, Sekimoto M, Doki Y, *et al*. The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer. *Int J Oncol* 2007;**30:**129–138.
44. Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, *et al*. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *J Am Med Assoc* 2008;**299:**425–436.
45. Watanabe T, Wu TT, Catalano PJ, Ueki T, Satriano R, Haller DG, *et al*. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med* 2001;**344:**1196–1206.
46. Schepeler T, Reinert JT, Ostenfeld MS, Christensen LL, Silahtaroglu AN, Dyrskjot L, *et al*. Diagnostic and prognostic microRNAs in stage II colon cancer. *Cancer Res* 2008;**68:**6416–6424.
47. Lips EH, van Eijk R, de Graaf EJ, Doornebosch PG, de Miranda NF, Oosting J, *et al*. Progression and tumor heterogeneity analysis in early rectal cancer. *Clin Cancer Res* 2008;**14:**772–781.

48. Clark-Langone KM, Wu JY, Sangli C, Chen A, Snable JL, Nguyen A, *et al*. Biomarker discovery for colon cancer using a 761 gene RT-PCR assay. *BMC Genomics* 2007;**8:**279.
49. Diep CB, Kleivi K, Ribeiro FR, Teixeira MR, Lindgjaerde OC, Lothe RA. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes Chromosomes Cancer* 2006;**45:**31–41.
50. Kurashina K, Yamashita Y, Ueno T, Koinuma K, Ohashi J, Horie H, *et al*. Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. *Cancer Sci* 2008;**99:**1835–1840.
51. Knosel T, Schluns K, Dietel M, Petersen I. Chromosomal alterations in lung metastases of colorectal carcinomas: associations with tissue specific tumor dissemination. *Clin Exp Metastasis* 2005;**22:**533–538.
52. Leroy O, Dhaenens CM, Schraen-Maschke S, Belarbi K, Delacourte A, Andreadis A, *et al*. ETR-3 represses Tau exons 2/3 inclusion, a splicing event abnormally enhanced in myotonic dystrophy type I. *J Neurosci Res* 2006;**84:**852–859.
53. Rouzier R, Rajan R, Wagner P, Hess KR, Gold DL, Stec J, *et al*. Microtubule-associated protein tau: a marker of paclitaxel sensitivity in breast cancer. *Proc Natl Acad Sci U S A* 2005;**102:**8315–8320.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article.

Table S1. Overlapping genomic copy number changes associated with poorer survival as recorded from this study and Kim *et al* [24].

Table S2. Univariate and multivariate analysis of the colorectal dataset indicating the prognostic significance of RP11-19F9 DNA copy number status.

Figure S1. Kaplan–Meier survival analysis for patients with sporadic CRC assessed for (A) Dukes' staging; (B) recurrence status; and (C) MSI status.

Figure S2. (A–D) Array-CGH profiles of four CRCs analysed with chromosome 18 tiling-path arrays showing DNA copy number loss of clone RP11-19F9 at 18q12.2; (E) pie-chart summarizing the type of DNA copy number change obtained from the array-CGH analysis of 30 tumour DNA samples by tiling-path array of chromosome 18; (F) Kaplan–Meier survival analysis for patients with tumours showing the DNA copy number loss of 18q12.2 (clone RP11-19F9).