



Published in final edited form as:

Cancer Res. 2008 March 1; 68(5): 1284–1295. doi:10.1158/0008-5472.CAN-07-2864.

## Chromosomal Breakpoints in Primary Colon Cancer Cluster at Sites of Structural Variants in the Genome

Jordi Camps<sup>1</sup>, Marian Grade<sup>1,3</sup>, Quang Tri Nguyen<sup>1</sup>, Patrick Hörmann<sup>1</sup>, Sandra Becker<sup>1</sup>, Amanda B. Hummon<sup>1</sup>, Virginia Rodriguez<sup>2</sup>, Settara Chandrasekharappa<sup>2</sup>, Yidong Chen<sup>1</sup>, Michael J. Difilippantonio<sup>1</sup>, Heinz Becker<sup>3</sup>, B. Michael Ghadimi<sup>3</sup>, and Thomas Ried<sup>1</sup>

<sup>1</sup>Genetics Branch, Center for Cancer Research, National Cancer Institute/NIH <sup>2</sup>Genome Technology Branch, National Human Genome Research Institute/NIH, Bethesda, Maryland

<sup>3</sup>Department of General and Visceral Surgery, University Medicine Göttingen, Göttingen, Germany

### Abstract

Genomic aberrations on chromosome 8 are common in colon cancer, and are associated with lymph node and distant metastases as well as with disease susceptibility. This prompted us to generate a high-resolution map of genomic imbalances of chromosome 8 in 51 primary colon carcinomas using a custom-designed genomic array consisting of a tiling path of BAC clones. This analysis confirmed the dominant role of this chromosome. Unexpectedly, the position of the breakpoints suggested colocalization with structural variants in the human genome. In order to map these sites with increased resolution and to extend the analysis to the entire genome, we analyzed a subset of these tumors ( $n = 32$ ) by comparative genomic hybridization on a 185K oligonucleotide array platform. Our comprehensive map of the colon cancer genome confirmed recurrent and specific low-level copy number changes of chromosomes 7, 8, 13, 18, and 20, and unveiled additional, novel sites of genomic imbalances including amplification of a histone gene cluster on chromosome 6p21.1-21.33 and deletions on chromosome 4q34-35. The systematic comparison of segments of copy number change with gene expression profiles showed that genomic imbalances directly affect average expression levels. Strikingly, we observed a significant association of chromosomal breakpoints with structural variants in the human genome: 41% of all copy number changes occurred at sites of such copy number variants ( $P < 2.2e^{-16}$ ). Such an association has not been previously described and reveals a yet underappreciated plasticity of the colon cancer genome; it also points to potential mechanisms for the induction of chromosomal breakage in cancer cells.

### Introduction

Colorectal cancer is the second leading cause of cancer death in Europe and in the United States, with ~300,000 new cases and 200,000 deaths each year (1). Cytogenetic and

**Requests for reprints:** Thomas Ried, Genetics Branch, Center for Cancer Research, National Cancer Institute/NIH, Building 50, Room 1408, 50 South Drive, Bethesda, MD 20892. Phone: 301-594-3118; Fax: 301-435-4428; riedt@mail.nih.gov..

Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

molecular cytogenetic studies clearly established that the colorectal cancer genome is defined by a specific distribution of genomic imbalances, most prominently, gains of chromosomes and chromosome arms 7, 8q, 13, and 20q as well as losses of chromosomes 4q, 8p, 17p, and 18q (2).

Within the last decade, microarray technology has been extensively applied to survey the cellular transcriptome of common solid tumors, including colorectal cancer, and for colon cancers, gene expression signatures were subsequently correlated with clinical outcome (for reviews, see refs. 3–5). However, high-resolution mapping of chromosomal copy number changes has only recently been achieved using BAC or cDNA clone-based arrays (6–10).

Chromosome 8q is one of the most frequently gained chromosomal arms in colorectal cancers (2), and it is conceivable that it contains more oncogenes than just the *MYC* oncogene, which maps to chromosome band 8q24.21. A potential role of chromosome 8q for the development of lymph node metastases has been previously reported (11), and overexpression of a gene, *PRL-3*, that maps to chromosome 8q24.3 has been implied in the development of liver metastases (12). Moreover, the 8q24 locus contains single nucleotide polymorphisms that are associated with an increased risk for the development of colon cancer (13–15).

Recently, a new class of genetic variation among humans has become recognized as a major source of genetic diversity. Termed structural variations, these polymorphisms can present themselves as copy number variants (CNV) and segmental duplications, which could be CNVs, but are not necessarily so (16–19). These polymorphisms could induce chromosomal rearrangements (20). One of our previous analyses of chromosomal aberrations in cell lines established from different carcinomas indicated that genomic copy number changes could be triggered by jumping trans-locations, many of which originated in the pericentromeric heterochromatin of several chromosomes (21). These regions frequently contain segmental duplications and other structural variants of the genome (22). Taken together, these data enticed us to systematically explore the genomic aberration profile and the potential involvement of structural variants of the human genome in the genesis of chromosomal aberrations in this common cancer. We therefore established a high-resolution map of genomic copy number changes in 51 primary colon carcinomas using comparative genomic hybridization (CGH) on both a BAC-based genomic tiling array for chromosome 8 and, for a subset of those, using a 185K oligonucleotide platform for whole genome coverage.

## Materials and Methods

### Patients and Sample Collection

The 51 patients included in this study were diagnosed with primary adenocarcinomas of the colon, and treated at the Department of General Surgery, University Medicine Göttingen, Göttingen, Germany. All patients received standardized surgery and histopathologic workup, and tumor staging was based on WHO criteria (23). Twenty-five tumors were associated with lymph node metastases [International Union Against Cancer (UICC)-III], whereas 26 tumors were not (UICC-II). Tumor samples were obtained immediately after surgery and stored on ice for inspection by an experienced pathologist. Consistent with

standard procedures, only samples with a tumor cell content of at least 70% were included in this study. Biopsies of normal adjacent mucosa were collected from some patients when possible. Table 1 summarizes the clinical data and experimental setup.

### DNA and RNA Isolation

Bioprotic material was in the range of 24 to 370 mg, and nucleic acids were extracted using TRIZOL (Invitrogen) following standard procedures.<sup>4</sup> On average, we obtained 200 Ag each of RNA and DNA. Nucleic acid quantification was determined using the Nanodrop ND-1000 UV-VIS spectrophotometer (Nanodrop). The quality of the nucleic acids after preparation was assessed using a 2100 Bioanalyzer (Agilent Technologies).

### Array CGH

**BAC array CGH platform**—The 1,463 BAC clones and DNA used to construct the chromosome 8 Human-BAC microarray were a subset of the Human “32K” BAC Re-Array library from the BACPAC Resources (Children's Hospital Oakland Research Institute, Oakland, CA).<sup>5</sup> The platform and details of the procedure are described in ref. 24.

Genomic DNA was digested using *RsaI* and *AluI* (Roche Applied Science), and the appropriate fragment size was confirmed on an agarose gel. After protein removal using a phenol-chloroform extraction, 600 ng of digested DNA were labeled using the Bioprime Labeling Kit (Invitrogen) to incorporate Cy5-dCTP or Cy3-dCTP (Amersham). Sex-matched tumor and reference DNA were combined and hybridized to the custom chromosome 8 BAC array in specifically designed hybridization cassettes (TeleChem International). After overnight hybridization, slides were washed and scanned on an Axon scanner using GenePixPro (3.0) software (Axon Instruments).

**Oligo array CGH platform**—Oligonucleotide array CGH (aCGH) was performed according to the protocol provided by the manufacturer (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis, protocol version 4.0, June 2006; Agilent Technologies), with minor modifications. Commercially available pooled control DNA (Promega) was used as sex-matched reference DNA in all hybridizations. Briefly, 3 µg of genomic DNA was digested for 2 h with *AluI* and *RsaI* (Promega). QIAprep Spin Miniprep Kit (Qiagen) was used for purifying the digested DNA. Tumor and reference DNA was labeled with Cy3-dUTP and Cy5-dUTP (Promega), respectively, in a random priming reaction using Bioprime Array CGH Genomic Labeling Module (Invitrogen). After 2 h of reaction, unincorporated nucleotides were removed using Microcon YM-30 columns (Millipore). Cy3 and Cy5-labeled samples were combined in equal amounts according to the incorporation of labeled nucleotides as measured using Nanodrop. Hybridization and washes were performed according to the manufacturer's protocol. Slides were scanned using a scanner (G2565BA; Agilent Technologies), and Agilent Feature Extraction software (version 9.1; Agilent Technologies) was applied for image analysis. To visualize the aCGH

<sup>4</sup><http://www.riedlab.nci.nih.gov/protocols.asp>

<sup>5</sup><http://bacpac.chori.org/>

data, we used Agilent CGH Analytics 3.4 software (Agilent Technologies). The quality of the slides was assessed using metrics provided by CGH Analytics.

### Gene Expression Profiling

Gene expression profiles for all 51 primary colon tumors and 21 associated mucosa samples were established as previously reported (25).

### Data Analysis

**BAC aCGH platform**—In order to compensate for scanner distortion between the Cy3 and Cy5 channel readings, we applied a 90th inter-percentile range (90IPR) normalization procedure to equalize the spread of Cy3 measurement to the spread of Cy5 measurements per array (in natural scale):

$$cCy3 = Cy3 \times (90IPR.Cy5 / 90IPR.Cy3),$$

where  $cCy3$  is the corrected Cy3 measurement, and  $90IPR.Cy5$  and  $90IPR.Cy3$  are the 95th percentile minus the 5th percentile measurements in the Cy5 and Cy3 channels, respectively.  $cCy3$  and Cy5 measurements are then  $\log_2$ -transformed, and their  $\log_2$  (ratio) are median-centralized by array using the following formula:

$$\log_2(R) = \log_2(cCy3) - \log_2(Cy5) + MD.\log_2(Cy5) - MD.\log_2(cCy3),$$

where  $MD.\log_2(Cy5)$  and  $MD.\log_2(cCy3)$  are the medians of  $\log_2(Cy5)$  and  $\log_2(cCy3)$  measurements, respectively. An aCGH segmentation algorithm developed under MATLAB was applied to all normalized arrays to extract segmented regions. Consensus gain or loss regions were obtained as described previously (24).

**Oligo aCGH platform**—The analysis of the aCGH experiments was performed with in-house developed software based on R version 2.4.1<sup>6</sup> and the DNA copy package from Bioconductor.<sup>7</sup> One array that did not pass the quality control criteria (derivative log ratio spread or DLRS<sub>spread</sub> > 0.3) was discarded. We also discarded features with no precise chromosomal location. The final data set was comprised of 29 arrays and 181,984 features. The data were smoothed using “smooth.CNA” function (with arguments `smooth.region = 1`, and `smooth.SD.scale = 3`), followed by the generation of chromosome segments using circular binary segmentation (CBS; ref. 26). We centralized DNA segments to the most common ploidy per array through an algorithm similar to the one offered in Agilent CGH Analytics 3.4 software. The cumulative frequency of loss score for each feature is the percentage of samples for which the segment value is below the threshold  $\log_2(5/6)$  corresponding to a loss of one DNA copy in 30% of diploid cells. Cumulative frequency is scaled to 100% = 4 (e.g., 25% = 1) in order to take advantage of the maximum range of the representation in genome, chromosome, and gene views in Agilent CGH Analytics 3.4.

<sup>6</sup><http://www.R-project.org>

<sup>7</sup><http://www.bioconductor.org>

Likewise, the cumulative frequency of gain score for each feature is the percentage of samples for which the segment value is above the threshold  $\log_2(7/6)$ .

The significance of association of chromosomal breakpoints within CNV loci was calculated as follows: the statistics for breakpoints in CNV loci is the  $\chi^2$  goodness of fit between the observed fraction of breakpoint in CNV loci (count of observed breakpoint in CNV loci/total observed breakpoints), and the fraction of expected breakpoints in CNV loci (total base pair of CNV areas in array/total base pair covered in array). The significance threshold for this statistical test was  $P < \alpha = 0.05$  (two-sided).

The correlation between average CGH copy number and average gene expression was performed using Pearson's correlation for each CBS segment with (*a*) ratio average values (CBS segment mean from this article), as the X-axis versus (*b*) average of gene expression [ $\log_2(\text{ratio})$ ; from ref. 25], as the Y-axis. We excluded gene expression arrays with >30% missing data points, and to prevent distortion caused by outliers, we excluded segments containing less than six features for either gene expression or CGH prior to calculating the correlation, i.e., 10 samples and 314 of 369 segments were retained. The significance threshold for this statistical test was  $P < \alpha = 0.05$  (two-sided).

## Results

### CGH using chromosome 8–specific tiling BAC arrays

Chromosome arm 8q is one of the most common targets of genomic amplification in colon cancer. It is also associated with the development of both lymph node and distant metastases, and contains single nucleotide polymorphisms that predispose to the development of this malignancy (2, 11, 12, 14, 15). We therefore aimed to generate a high-resolution map of genomic copy number changes by analyzing 51 primary colon tumors by CGH using a BAC clone-based genomic tiling array. Twenty-five of these tumors were associated with lymph node metastases at the time of surgery (UICC-III), whereas the remaining patients were free of lymph node metastases (UICC-II,  $n = 26$ ). The clinical information is presented in Table 1.

Confirming previous results, 50% of the cases showed aberrations on chromosome 8; 37% had gains on the long arm, and 45% had losses on 8p. Two regions with the highest copy number increases mapped to genome locations 105 to 120 Mbp and 127 to 142 Mbp. This includes chromosome band 8q24.21, the genomic location of the *MYC* oncogene. Interestingly, in striking difference from the results suggested by conventional CGH, the short arm of chromosome 8 was not subject to loss in its entirety: in the majority of samples with 8p alterations, the loss of this arm did not include a small region close to the centromere. This region, which includes 5.5 Mbp of the short arm, was either present in normal copy number, or in fact gained to the same extent as the long arm. The summary of this analysis is presented in Fig. 1A and B. Interestingly, when we then tried to understand why chromosome 8p was prone to chromosomal breaks to such an extent, we noticed that in 9 out of 14 cases, the breakpoints coincided with sites of known structural variants identified within the human population, either CNVs or segmental duplication (Supplementary Table

S1). Figure 1B summarizes the BAC array data of the 8p aberration patterns in individual cases.

### High-resolution genome-wide mapping of DNA copy number changes

In order to more precisely map these breakpoints and to investigate whether the observed predilection for chromosomal breaks at sites of known structural variants applies to regions other than 8p, we profiled 31 of the 51 colon cancers analyzed with the BAC arrays by aCGH on a 185K oligonucleotide, genome-wide platform (see Table 1 for the respective cases). Regions of genomic imbalances in these tumors were determined using CBS (26). Taking the different resolution limits of the platforms into consideration, we observed an excellent congruence between the techniques, and the aberration patterns on 8p were confirmed. Our analyses also confirmed the recurrent low-level copy number changes of chromosomes 7, 8, 13, 18, and 20, which are specific for sporadic colorectal cancers (27). However, attributable to the increased resolution of this platform for aCGH, additional novel sites of chromosomal gains and deletions could be identified. Specifically, we detected 393 chromosomal breakpoints (defined as segments of copy number change) in 31 cases, for an average of 12.7 breakpoints per case (0–34). One hundred and sixty-nine breakpoint segments (including those that affected entire chromosomes) resulted in copy number increases, whereas 202 regions of copy number loss were present. Segments with copy number increase were recurrently mapped to chromosomes and chromosome arms 7, 8q, 13, and 20q, whereas losses occurred most frequently on 1p, 5q, 8p, 14, 15, 17p, 18, 21, and 22. A summary of these results is presented as cumulative gain or loss in Fig. 1C, and as a frequency distribution in Supplementary Fig. S1. Gains on chromosomes 13 and 20 were most commonly observed, and also revealed the highest level of genomic amplification, followed by copy number increases of chromosomes 7 and 8q. Chromosome arms 18q, 17p, and 8p showed the highest degree of genomic loss (both in terms of cases and actual copy number reduction).

We detected several regions whose recurrent copy number changes were not appreciated in our previous analyses of colorectal carcinomas using conventional CGH analysis (28). In addition to the above-described retention on 8p, we observed a similar pattern on chromosome 20: the breakpoint that results in copy number increase resides in the euchromatic region of 20p, and not in the centromere. In addition, we observed interstitial deletions of chromosome band 4q34.3-35.2 in three cases (CC-P19, CC-P20, and CC-P65), and a deletion that included the terminal band of the short arm of chromosome 11 (11p15.5) in two cases (CC-P23 and CC-P38). Bands 13q21.32 to 13q31.2 were deleted on this commonly gained chromosome in CC-P23, and remained in normal copy number (with the rest of the chromosome gained) in CC-P65. A few localized high-level amplifications were mapped to chromosome bands 4q13.2-13.4, 5q32-33.2, and 6p21.1 (CC-P14), and 16q12.2 in CC-P65. In CC-P23, we observed the genomic amplification of the *ANKRD10* gene, which maps to distal chromosome 13.

### Comparison between lymph node–negative and -positive cancers

The presence of synchronous lymph node metastases dictates the inclusion of chemotherapy in the treatment of patients with colon cancer. In order to explore whether lymph node status

could be reflected by specific copy number changes on chromosome 8, as previously suggested (11), or elsewhere in the genome, we compared the distribution of genomic imbalances as determined in both groups using the oligonucleotide array platform. The percentage of chromosomal gains and losses was not different between the lymph node–positive (average, 12.9 per case) and lymph node–negative (average, 11.7 per case) carcinomas. The average number of gained or lost segments in the UICC-II tumors was 6.8 and 6.1, respectively, and for the UICC-III tumors, it was 4.8 and 7.5, respectively. In order to further analyze whether tumors associated with lymph node metastases carry distinct genomic aberration profiles, we analyzed the frequency of all CBS units in the two groups: we could not detect any CBS units that were uniquely gained or lost in either the UICC-II or UICC-III samples, nor did we detect a differential distribution of CBS units between the groups that exceeded a 30% difference threshold. The summary plots of the UICC-II and UICC-III tumors are displayed in Supplementary Fig. S2A and B.

### **Influence of genomic imbalances on gene expression**

Genomic copy number changes are arguably one of the most recurrent features of solid tumors of epithelial origin. Consequently, numerous groups attempted to clarify the relationship between genomic copy number changes and gene expression levels; however, most of these studies focused either on the effect of whole chromosomes, or on regional amplicons (25, 29–33). We now analyzed, for the entire colon cancer genome, this correlation by plotting the average gene expression values for all CBS units against their genomic copy number (we only included those 17 cases for which we had gene expression results in both the tumor and matched normal mucosa, and those CBS segments that contained more than five genes). The analysis, shown in Fig. 2, revealed a significant correlation of genomic copy number with average gene expression levels, therefore suggesting a direct effect of gene copy on relative message levels ( $R = 0.66709$ ,  $P = 2.2e^{-16}$ ).

### **CNVs**

In addition to low level copy number changes, the CBS analysis revealed numerous recurrent loci of localized high-level copy number increases or decreases relative to the reference DNA. Such changes could be indicative of structural variations in the genome, either germ line or somatic. Structural variations, including CNVs, have recently emerged as a novel class of DNA segments that differ from one individual to another (20). The systematic mapping of CNVs in 270 individuals that constitute the human HapMap collection (34) suggests that ~12% of the human genome could be subject to copy number variation (20), with as much as 3% of these regions (~0.3% of the total genome) varying from one individual to another (35). CNVs therefore contribute significantly to human sequence variation. Applying the CBS algorithm, we could identify 120 sites that were suggestive of CNVs (i.e., sites of high-level copy number increase or decrease of no more than 200 kbp). The comparison of the variants detected in our data set derived from 31 tumors with the database on genomic variation<sup>8</sup> indicated that 81 of those variants (67.5%)

---

<sup>8</sup><http://projects.tcag.ca/variation/>

overlapped with known CNVs, whereas 39 (32.5%) were potentially novel sites of CNVs. A complete list is provided in Supplementary Table S2.

In order to assess whether these alterations were genomic copy number changes that emerged *de novo* in the tumors, i.e., somatic, or whether they would have to be considered germ line events, we hybridized tumor DNA from five patients against DNA prepared from matched normal mucosa tissue. CNVs detectable in such experiments can be considered bona fide somatic events. The initial CGH experiments revealed 54 known CNVs in these five patients (9–13). We now observe that 13 of these CNVs remained when tumor DNA was hybridized against DNA from matched normal mucosa (1–4). Based on these observations, we conclude that 24% of the CNVs are actual variants that emerged in the tumor tissue, and hence, somatic CNVs. Examples of these variant regions are shown in Fig. 3A.

Similar to fragile sites, regions of genomic copy number variation could trigger genomic rearrangements (20, 22). In order to establish to which extent genomic regions containing CNVs contribute to the emergence of chromosomal translocations (as deduced from the presence of segments of genomic copy number change by CGH), we asked how frequently chromosomal translocations coincided with the location of previously identified CNVs. Given the high resolution of our platform (16 kbp), we could manually annotate the breakpoint sequence for each segment using the Database of Genomic Variation<sup>8</sup> in order to search for structural genomic variants at these genome coordinates. In the 31 cases analyzed with the oligonucleotide platform, we mapped 393 sites of genomic copy number change, 161 of which occurred at the site of known CNVs (Fig. 3C). Taking into account that ~18% of the genome consists of segments identified as CNV, the probability that 41% of all translocations mapped to CNVs by coincidence is exceedingly low ( $P < 2.2e^{-16}$ ). This suggests that CNV loci (including segmental duplications) contribute significantly to the emergence of chromosomal breaks in colon cancer, and hence, to the development of genomic imbalances. CNVs that colocalized to chromosomal breakpoints in our data set are listed in Table 2. Figure 3B presents an example of a subchromosomal genomic deletion that eliminates one copy of the tumor suppressor gene *APC* and shows the association between the site of the chromosomal break with a known CNV. Figure 4 shows the possible emergence of genomic copy number changes in CC-P10. In this tumor, we observed chromosomal breakpoints that coincided with two segmental duplications, DC3225 on chromosome 17p and DC2472 on chromosome 20p. A sequence homology of 94.51% between these two sites suggests that homologous recombination events could have contributed to a chromosomal translocation, which eventually leads to the observed pattern of DNA gain and loss.

## Discussion

### Patterns of imbalances

Here, we present a comprehensive map of genomic imbalances in primary colon carcinomas generated using high-resolution aCGH on a genomic tiling array for chromosome 8 and a 185K oligonucleotide platform. The results are, in general, congruent with previous analyses using chromosome banding techniques (2), CGH on metaphase chromosomes (28), and



aCGH with a genomic BAC platform (6–10). In fact, the summary pattern of chromosomal gains and losses in our data set and data sets reported in the literature suggest a striking conservation of genomic imbalances, and underlines the biological significance of these recurrent aneuploidies. We observed, however, a few dissimilarities between the data set presented here and previously published results. For instance, the short arm of chromosome 8 is not always lost in its entirety (as suggested by cytogenetic analyses using chromosome templates), but a minimally retained region that escapes this loss comprises chromosome band 8p11.1-11.2, which is consistent with previous aCGH analyses on genomic platforms (9, 36). A similar phenomenon on the short arm of chromosome 20 was detected. Second, we observed several regions of subtle copy number changes that were clearly below the resolution of conventional cytogenetic or CGH analyses. In patient CC-P9, we observed a localized amplification of chromosome band 6p21.1, which resulted in the significant overexpression of histone gene *HIST1H2BM* in this tumor. Other examples include a common deletion mapped to chromosome band 4q34-35. The most notable difference between chromosome CGH analysis, the use of an overlapping BAC array for chromosome 8, and the high-resolution oligonucleotide platform was the identification of frequent sites of small, high-level gains and losses, many of which coincided with loci of known structural variants in the human genome,<sup>8</sup> which could only be mapped using the 185K oligonucleotide platform. This will be discussed separately below.

### **Correlation of genomic copy number and gene expression changes**

The results presented here underscore the dominant role of specific and recurrent genomic imbalances, which arguably, are one of the defining features of genetic insults in colon cancer cells. We and others have therefore tried to understand the consequences of such genomic imbalances on the cancer transcriptome (25, 29–33). In general, the data are consistent with the interpretation that genomic copy number is positively correlated to transcript levels. The data set generated here now affords us the possibility to interrogate the relationship of genomic imbalances, as detected by segments of copy number change based on the CBS analysis (185K oligonucleotide arrays), with the expression levels of resident genes for the entire genome (Fig. 2). The data show that there is indeed a general, and statistically significant correlation of genomic copy number and gene expression levels and thus provide further evidence that these imbalances exert a direct effect on the cancer transcriptome, and hence, result in a massive and complex deregulation of the transcriptional equilibrium of malignant epithelial cells. This observation underlines the importance of the question as to which extent such rather global gene expression changes contribute to tumorigenesis vis-à-vis the targeted deregulation of specific genes by mutation, deletion, amplification, or epigenetic deregulation.

### **CNVs and potential mechanisms of induction of chromosome breakage**

CNVs constitute a subset of structural variants that represent a substantial amount of interindividual genetic variation (20). The most comprehensive catalogue of structural variants in the human genome can be found at <http://projects.tcag.ca/variation/>. The data summarized there was generated by analyzing the genomes of 270 individuals from the human HapMap consortium using both aCGH and genome-wide single nucleotide polymorphism platforms. These variants are rather ubiquitous, comprising ~ 12% of the

human genome. Some of them have been shown to be associated with a particular phenotype and with disease (20). Based on a comprehensive evaluation of chromosomal breakpoints and associated genomic copy number changes in cell lines derived from solid tumors (i.e., bladder, prostate, cervix, pancreas, and breast), we could previously show that a considerable fraction of chromosomal translocations (in that case referred to as jumping translocations) originated in the pericentromeric heterochromatin of several chromosomes (21). Such heterochromatin is enriched for segmental duplications, and these show a 6:1 ratio of interchromosomal to intrachromosomal duplications. These regions can also vary in copy number between individuals, and if so, could be classified as CNVs (22). We were therefore curious as to which extent chromosomal breakpoints (as defined by sites of genomic copy number change using aCGH) colocalize with such structural variants in the genome of primary colon cancers. Surprisingly, ~ 41% of all translocations resided at sites of known CNVs, including segmental duplications (Fig. 3; Table 2). Such an association is highly significant ( $P < 2.2e^{-16}$ ). Figure 4 suggests a possible scenario on how the observed pattern of genomic gain and loss could be explained in one of the tumors analyzed here (CC-P10). It is, however, not possible to perform cytogenetic analysis on this very sample, and therefore, one cannot formally prove that the observed pattern of imbalance is indeed caused by translocations between chromosomes 17 and 20 despite the high degree of homology (95%) between the segmental duplications that colocalize with these breakpoints. Alternatively, CNVs and segmental duplications are simply regions more prone to chromosome breakage, which can result in loss of genomic segments due to the lack of a centromere, or translocation with other regions in the genome without homology. The difference in copy number of these regions between individuals, however, is perhaps an indication that they are particularly susceptible to homology-mediated recombination, i.e., formation of chiasmata, in meiotic cells. In cells experiencing DNA damage, one could easily envision that aberrant homology-mediated repair of segmentally duplicated regions might also lead to chromosome aberrations in somatic cells, such as deletions, inversions, and translocations. Such analyses will have to be conducted using cell lines established from primary tumors. The mere fact that homologous chromosomes in an interphase nucleus rarely tend to be in the same topographical neighborhood (37) makes it more likely that a homology search will identify a duplicated region on a different chromosome. This may explain the relatively high frequency of whole chromosome arm gains and losses in aneuploid tumors. Why might these regions be more susceptible to DNA damage? First, CNVs are often found in association with gene coding regions and therefore might be expected to be in an open configuration, making them more susceptible to DNA damage. Alu sequences, satellite repeats, and regions with hallmarks of DNA fragility are found to be enriched at the boundaries of these regions, supporting the hypothesis that these areas are preferential sites of DNA double-strand breaks, making them ideal substrates for repair pathways with the potential for causing increased copy number or rearrangements. Gorgoulis et al. (38) and Bartkova et al. (39) observed an early activation of DNA damage response pathways in precancerous lesions. Serrano and colleagues showed that high expression of oncogenes triggers a permanent block in replication, termed oncogene-induced senescence (40). Oncogene-induced senescence has recently been shown to induce a DNA damage response in tissue culture models (41, 42) as well as *in vivo* during the development of thymocytes (43), and is able to restrict the growth of human and murine precancerous

tissues (44–48). These early incidents set the stage for the events outlined above. Further progression to more advanced dysplastic lesions and to invasive carcinomas was associated with p53 inactivation and reduction of apoptosis. Interestingly, allelic loss of loci prone to DNA double-strand break formation, i.e., fragile sites was common. The authors put forward a model in which, at early stages of tumorigenesis, replicative stress triggers the formation of double-strand breaks, which in turn results in genomic instability, and through that to inhibition of apoptosis and cell cycle arrest. One could therefore reasonably speculate that CNV-induced double-strand breaks are among the earliest gross chromosomal aberrations in cancer genomes. The resulting unbalanced trans-locations could then, in addition to aneuploidies of entire chromosomes (which are also observed in premalignant, early dysplastic lesions), contribute to the emergence of patterns of genomic imbalances that define different tumors of epithelial origin. These speculations are potentially substantiated by our observation that ~ 24% of the observed CNVs are actually *de novo* events, i.e., are detectable when tumor DNA was compared with DNA prepared from matched normal mucosa tissue. These data suggest that regions of copy number variation observed in the normal population continue to be subject to hypervariability and are foci of genomic instability in the tumor.

It remains to be seen whether the striking colocalization of sites of structural variants in the genome and cancer-associated chromosomal breakpoints that we observed here in colon carcinomas occurs in other epithelial neoplasms as well. It will be equally interesting to determine whether the distribution and frequency of specific CNVs is associated with population-based cancer risk.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Grant support:** Intramural Research Program of the NIH, National Cancer Institute, the Deutsche Forschungsgemeinschaft (KFO 179), and a stipend from the Deutsche Krebshilfe (M. Grade).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

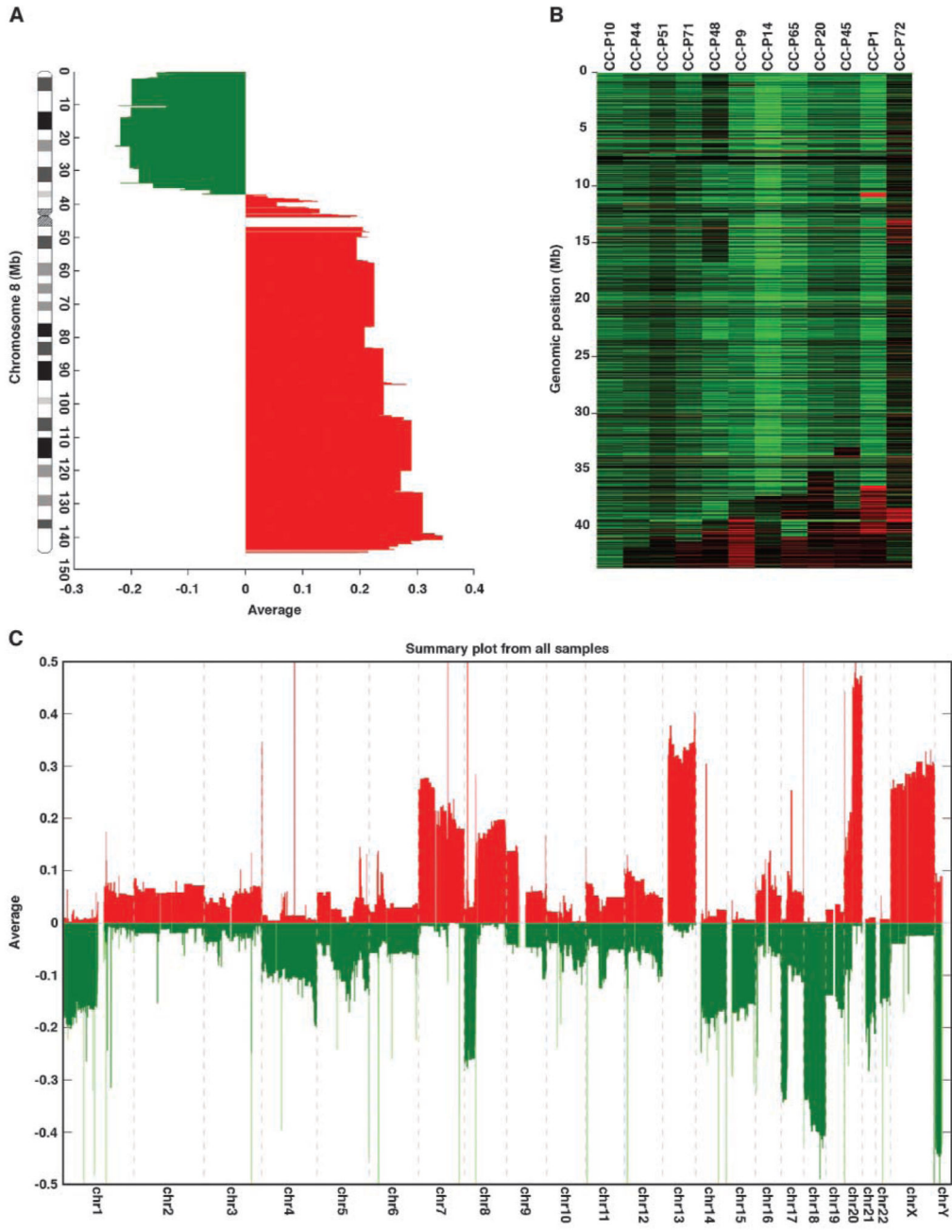
The authors thank Hesus M. Padilla-Nash, Buddy Chen, Joseph Cheng, and Jessica Eggert for helpful discussion, technical, and editorial assistance.

## References

1. Midgley R, Kerr D. Colorectal cancer. *Lancet*. 1999; 353:391–9. [PubMed: 9950460]
2. Heim, S.; Mitelman, F. *Cancer cytogenetics*. 2nd ed.. Wiley-Liss; New York: 1995.
3. Shih W, Chetty R, Tsao MS. Expression profiling by microarrays in colorectal cancer (review). *Oncol Rep*. 2005; 13:517–24. [PubMed: 15706427]
4. Bustin SA, Dorudi S. Gene expression profiling for molecular staging and prognosis prediction in colorectal cancer. *Expert Rev Mol Diagn*. 2004; 4:599–607. [PubMed: 15347254]
5. Augenlicht LH, Velcich A, Klampfer L, et al. Application of gene expression profiling to colon cell maturation, transformation and chemoprevention. *J Nutr*. 2003; 133:2410–6S.

6. Lassmann S, Weis R, Makowiec F, 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; 85:289–300.
7. Kim MY, Yim SH, Kwon MS, et al. Recurrent genomic alterations with impact on survival in colorectal cancer identified by genome-wide array comparative genomic hybridization. *Gastroenterology*. 2006; 131:1913–24. [PubMed: 17087931]
8. Gaasenbeek M, Howarth K, Rowan AJ, et al. Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res*. 2006; 66:3471–9. [PubMed: 16585170]
9. Nakao K, Mehta KR, Fridlyand J, et al. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis*. 2004; 25:1345–57. [PubMed: 15001537]
10. Camps J, Armengol G, del Rey J, et al. Genome-wide differences between microsatellite stable and unstable colorectal tumors. *Carcinogenesis*. 2006; 27:419–28. [PubMed: 16272173]
11. Ghadimi BM, Grade M, Liersch T, et al. Gain of chromosome 8q23–24 is a predictive marker for lymph node positivity in colorectal cancer. *Clin Cancer Res*. 2003; 9:1808–14. [PubMed: 12738738]
12. Saha S, Bardelli A, Buckhaults P, et al. A phosphatase associated with metastasis of colorectal cancer. *Science*. 2001; 294:1343–6. [PubMed: 11598267]
13. Haiman CA, Le Marchand L, Yamamoto J, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet*. 2007; 39:954–6. [PubMed: 17618282]
14. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*. 2007; 39:984–8. [PubMed: 17618284]
15. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet*. 2007; 39:989–94. [PubMed: 17618283]
16. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004; 305:525–8. [PubMed: 15273396]
17. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004; 36:949–51. [PubMed: 15286789]
18. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006; 444:444–54. [PubMed: 17122850]
19. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007; 315:848–53. [PubMed: 17289997]
20. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006; 7:85–97. [PubMed: 16418744]
21. Padilla-Nash HM, Heselmeyer-Haddad K, Wangsa D, et al. Jumping translocations are common in solid tumor cell lines and result in recurrent fusions of whole chromosome arms. *Genes Chromosomes Cancer*. 2001; 30:349–63. [PubMed: 11241788]
22. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet*. 2006; 7:552–64. [PubMed: 16770338]
23. Sobin, LH.; Wittekind, C. UICC: TNM classification of malignant tumors. 6th ed.. Wiley-Liss; New York: 2002.
24. Rodriguez V, Chen Y, Elkahlon A, Dutra A, Pak E, Chandrasekharappa S. Chromosome 8 BAC array comparative genomic hybridization and expression analysis identify amplification and overexpression of TRMT12 in breast cancer. *Genes Chromosomes Cancer*. 2007; 46:694–707. [PubMed: 17440925]
25. Grade M, Hörmann P, Becker S, et al. Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. *Cancer Res*. 2007; 67:41–56. [PubMed: 17210682]
26. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004; 5:557–72. [PubMed: 15475419]

27. Ried T, Knutzen R, Steinbeck R, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer*. 1996; 15:234–45. [PubMed: 8703849]
28. Grade M, Becker H, Liersch T, Ried T, Ghadimi BM. Molecular cytogenetics: Genomic imbalances in colorectal cancer and their clinical impact. *Cell Oncol*. 2006; 28:71–84. [PubMed: 16823176]
29. Upender MB, Habermann JK, McShane LM, et al. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer Res*. 2004; 64:6941–9. [PubMed: 15466185]
30. Monni O, Barlund M, Mousses S, et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A*. 2001; 98:5711–6. [PubMed: 11331760]
31. Pollack JR, Sorlie T, Perou CM, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A*. 2002; 99:12963–8. [PubMed: 12297621]
32. Tsafrir D, Bacolod M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res*. 2006; 66:2129–37. [PubMed: 16489013]
33. Grade M, Ghadimi BM, Varma S, et al. Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/hcaterin signaling pathway in human rectal carcinomas. *Cancer Res*. 2006; 66:267–82. [PubMed: 16397240]
34. The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005; 437:1299–320. [PubMed: 16255080]
35. Wong KK, de Leeuw RJ, Dosanjh NS, et al. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet*. 2007; 80:91–104. [PubMed: 17160897]
36. Pole JC, Courtay-Cahen C, Garcia MJ, et al. High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. *Oncogene*. 2006; 25:5693–706. [PubMed: 16636668]
37. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. 2001; 2:292–301. [PubMed: 11283701]
38. Gorgoulis VG, Vassiliou LV, Karakaidos P, et al. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature*. 2005; 434:907–13. [PubMed: 15829965]
39. Bartkova J, Horejsi Z, Koed K, et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*. 2005; 434:864–70. [PubMed: 15829956]
40. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*. 1997; 88:593–602. [PubMed: 9054499]
41. Di Micco R, Fumagalli M, Cicalese A, et al. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature*. 2006; 444:638–42. [PubMed: 17136094]
42. Bartkova J, Rezaei N, Liontos M, et al. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*. 2006; 444:633–7. [PubMed: 17136093]
43. Xu M, Yu Q, Subrahmanyam R, et al. B-Catenin expression results in p53-independent DNA damage and oncogene-induced-senescence in pre-lymph-omagenic thymocytes *in vivo*. *Mol Cell Bio*. 2008 In press.
44. Braig M, Lee S, Loddenkemper C, et al. Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature*. 2005; 436:660–5. [PubMed: 16079837]
45. Chen Z, Trotman LC, Shaffer D, et al. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature*. 2005; 436:725–30. [PubMed: 16079851]
46. Collado M, Gil J, Efeyan A, et al. Tumour biology: senescence in premalignant tumours. *Nature*. 2005; 436:642. [PubMed: 16079833]
47. Michaloglou C, Vredeveld LC, Soengas MS, et al. BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature*. 2005; 436:720–4. [PubMed: 16079850]
48. Sharpless NE, DePinho RA. Cancer: crime and punishment. *Nature*. 2005; 436:636–7. [PubMed: 16079829]



**Figure 1.** A, summary of the BAC aCGH analysis of chromosome 8 from 51 patients. Chromosome losses (*green*) and gains (*red*) are in log 2 ratios. Note the retention of a portion of the short arm of chromosome 8 from the commonly observed loss of this chromosome arm. *Left*, genome coordinates and cytogenetic bands. *B*, summary of BAC aCGH analysis of individual cases (see patient numbers on top of graph) for those cases that revealed chromosome 8p abnormalities. Note the varying degrees of retention/amplification of chromosome 8p material (*red*). DNA copy number losses (*green*). The red line at 6 Mbp

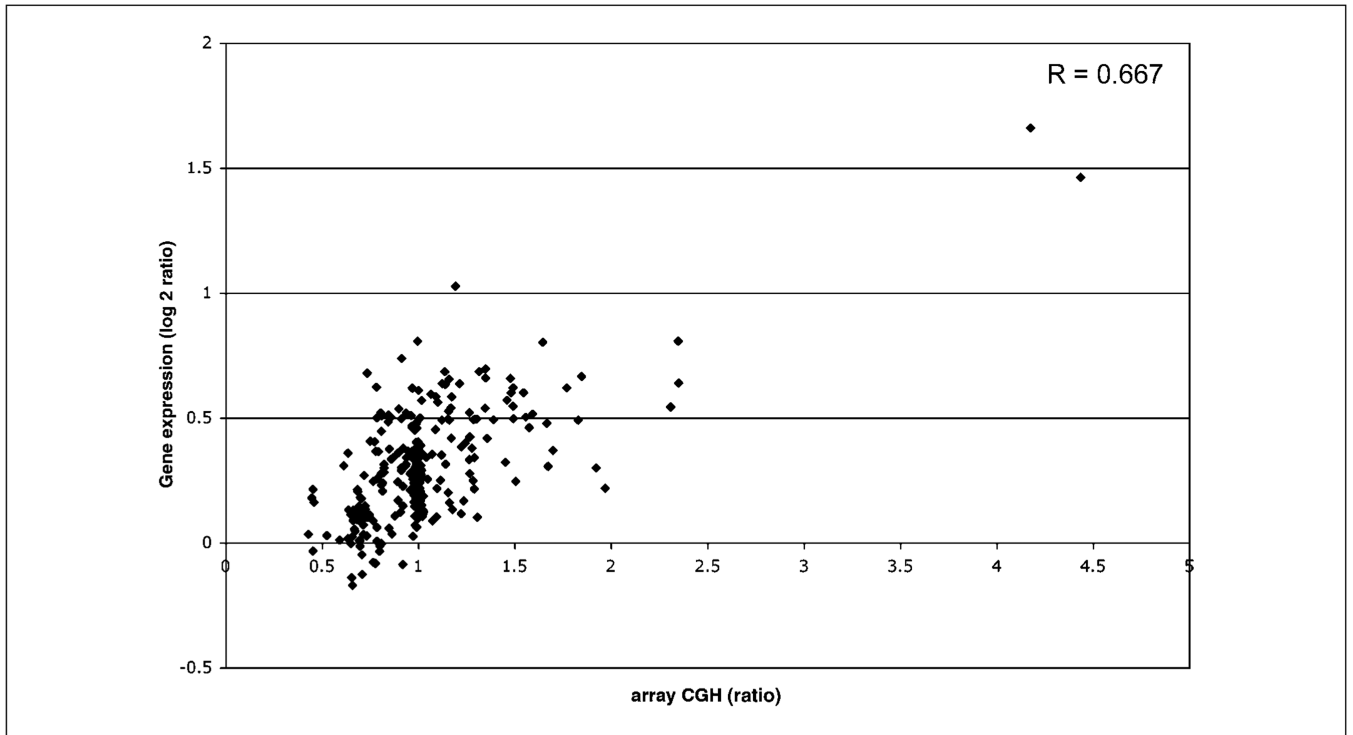
observed in all cases indicates a novel, common CNV which includes the gene *GATA4*. The amplicon at 11 Mbp in CC-P1 includes two genes, *SOX7* and *PINX1*. *Left*, genome coordinates. *C*, summary of oligonucleotide aCGH analysis from 31 cases. Copy number increases (*red*) and decreases (*green*). Sites suggestive of CNVs are indicated as thin lines in light green or red. Numbers below the graph refer to chromosomes. Average values are in log 2 ratios.

Author Manuscript

Author Manuscript

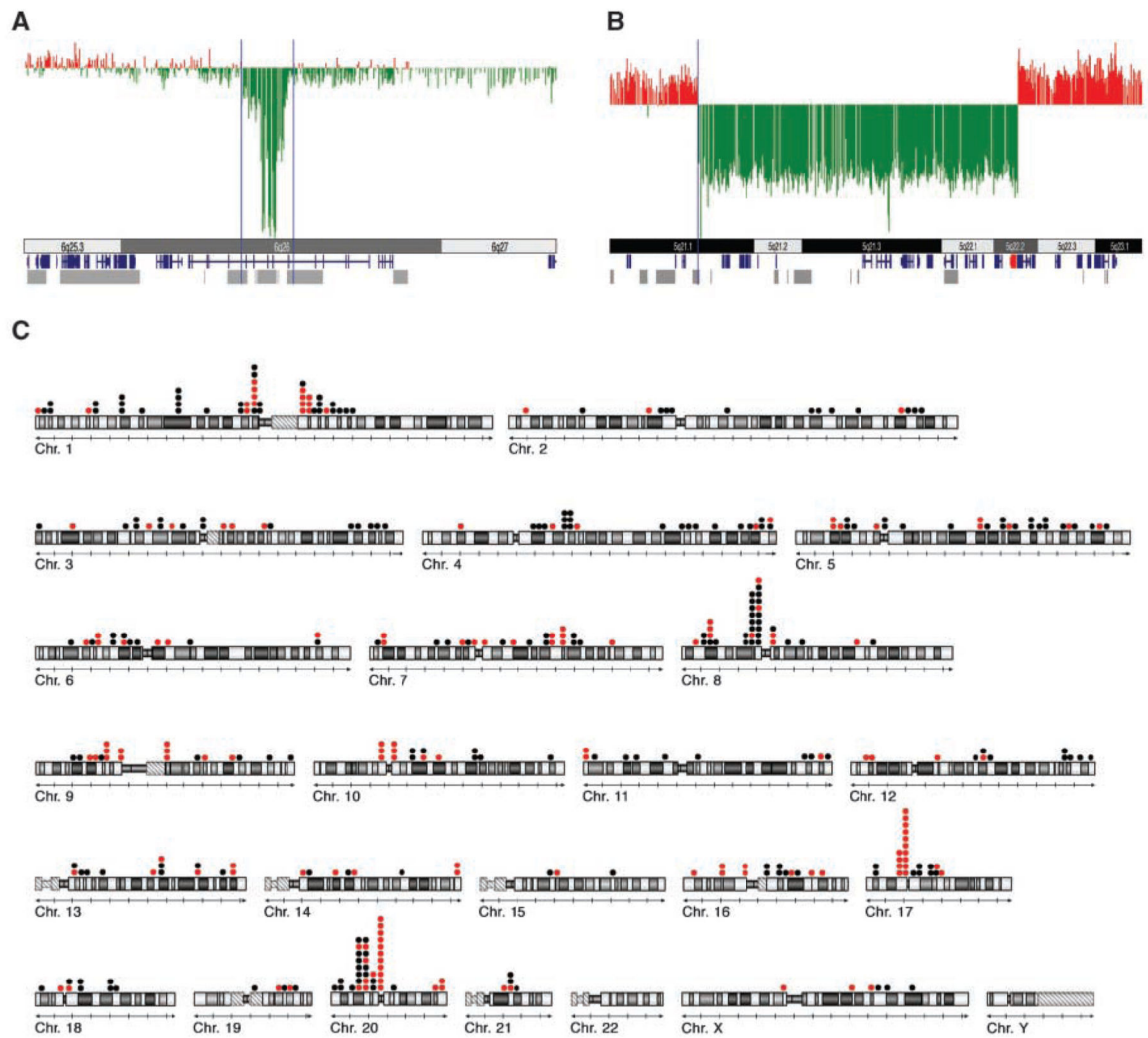
Author Manuscript

Author Manuscript



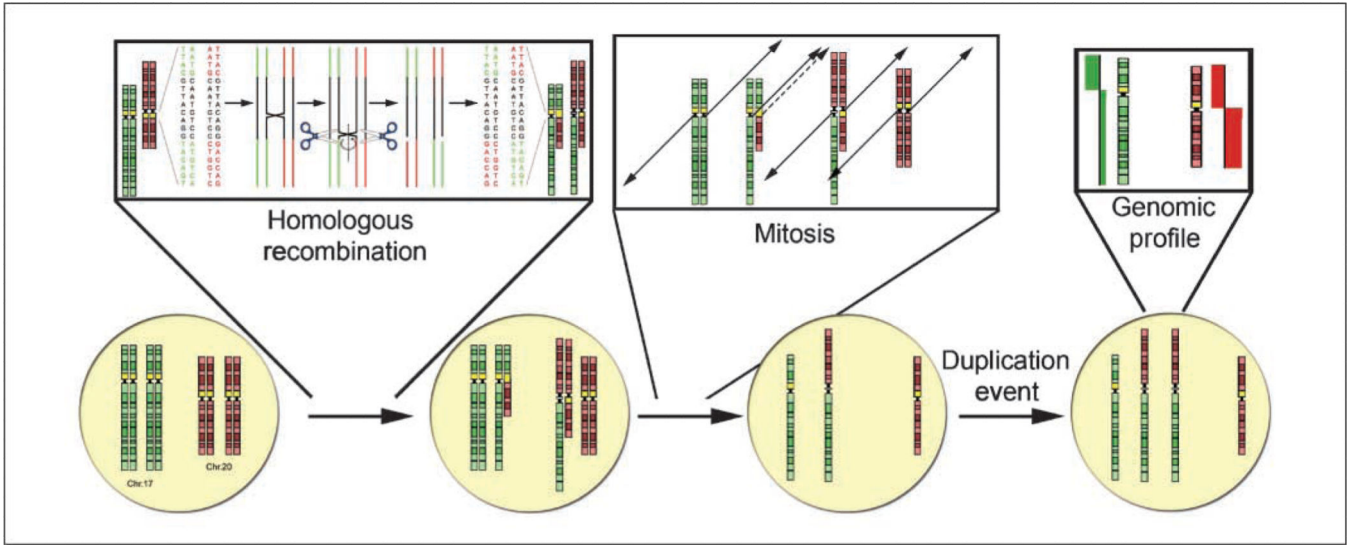
**Figure 2.**  
Correlation of genomic copy number of all CBS units with gene expression levels.





**Figure 3.**

Regional alterations of CNVs and their prevalence at sites of chromosome breakpoints. *A*, example of a somatic CNV in patient CC-P42. The hybridization of matched tumor and normal mucosa DNA revealed a site of a somatic copy number change at the site of a known CNV on chromosome 6q26. *B*, example of an aCGH experiment showing the colocalization of a known CNV with a chromosomal breakpoint in patient CC-P24. The breakpoint occurred at genome coordinate chr5:100699314–100749188. The CNV at this site spans chr5:100,535,625–100,788,621 (18) and is annotated in the Database of Genomic Variants.<sup>8</sup> *C*, genome-wide map of chromosomal breakpoints in colon cancer. Black dots, observed chromosomal breakpoints according to our CBS analysis. *Red dots*, breakpoints that coincide with the map location of structural variants of the human genome. Although CNV-induced breakpoints occur throughout the genome, they cluster near centromeres. The precise genomic coordinates of all CNV-associated breakpoints are provided in Table 2.



**Figure 4.** Structural variation-mediated translocation resulting in genomic imbalance. Interchromosomal segmental duplications with high sequence homology are found on chromosomes 17 (green) and 20 (red) in their pericentromeric regions (white boxes). In the example illustrated above, a single-strand break occurring in one segmental duplication on chromosome 17p finds homology with a segmental duplication with inverted orientation on chromosome 20p (first inset). The strand invasion at the site of homology (black text and lines) results in the formation of a Holliday junction and branch migration. One resolution of the Holliday junction occurs through single-strand breaks of the uninvolved DNA strands and rotation of the structure resulting in a chromosome translocation of 17p to 20p and 17q to 20q. The former structure lacks a centromere whereas the latter is a dicentric chromosome. Spindle attachment and chromosome segregation during mitosis could result in one daughter cell containing a genomic imbalance due to loss of the acentric t(17;20). Duplication of the t(17;20)(q;q) in subsequent cell divisions would result in the genomic imbalances observed in the tumor of patient CC-P10.

Table 1

Clinical information and experimental setup

Patient code	Histopathology	Chromosome 8 BAC microarray	Gene expression microarray	185K oligonucleotide microarray
CC-P1	pT <sub>3a</sub> pN <sub>0</sub> (0/17) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P2	pT <sub>3</sub> pN <sub>0</sub> (0/19) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P3	pT <sub>3</sub> pN <sub>0</sub> (0/29) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P4	pT <sub>3a</sub> pN <sub>0</sub> (0/31) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	
CC-P6	pT <sub>4</sub> pN <sub>0</sub> (0/17) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	(×)
CC-P7	pT <sub>3</sub> pN <sub>0</sub> (0/25) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P8	pT <sub>3</sub> pN <sub>0</sub> (0/44) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P9	pT <sub>3b</sub> pN <sub>0</sub> (0/31) M <sub>0</sub> R <sub>0</sub> G <sub>1-2</sub>	×	×	×
CC-P10	pT <sub>3b</sub> pN <sub>0</sub> (0/20) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P11	pT <sub>3a</sub> pN <sub>0</sub> (0/21) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	(×)	×	×
CC-P12	pT <sub>3</sub> pN <sub>0</sub> (0/27) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P13	pT <sub>3b</sub> pN <sub>0</sub> (0/39) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P14	pT <sub>3</sub> pN <sub>0</sub> (0/23) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P15	pT <sub>3</sub> pN <sub>0</sub> (0/31) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	×
CC-P16	pT <sub>3</sub> pN <sub>0</sub> (0/15) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P19	pT <sub>4</sub> pN <sub>0</sub> (0/57) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P20	pT <sub>3b</sub> pN <sub>0</sub> (0/28) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P21	pT <sub>3b</sub> pN <sub>0</sub> (0/24) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P22	pT <sub>3</sub> pN <sub>0</sub> (0/15) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P23	pT <sub>3</sub> pN <sub>0</sub> (0/21) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	×
CC-P24	pT <sub>3</sub> pN <sub>0</sub> (0/17) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P26	pT <sub>3</sub> pN <sub>0</sub> (0/20) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P27	pT <sub>3</sub> pN <sub>0</sub> (0/26) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P28	pT <sub>3</sub> pN <sub>0</sub> (0/20) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P30	pT <sub>3b</sub> pN <sub>0</sub> (0/35) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P32	pT <sub>3a</sub> pN <sub>0</sub> (0/23) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P34	pT <sub>3</sub> pN <sub>1</sub> (2/17) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P35	pT <sub>4</sub> pN <sub>1</sub> (2/51) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P36	pT <sub>3</sub> pN <sub>2</sub> (15/42) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P37	pT <sub>3</sub> pN <sub>1</sub> (1/25) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P38	pT <sub>2</sub> pN <sub>1</sub> (1/23) M <sub>0</sub> R <sub>0</sub> G <sub>2-3</sub>	×	×	×
CC-P39	pT <sub>3c</sub> pN <sub>1</sub> (1/28) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P42	pT <sub>3a</sub> pN <sub>1</sub> (1/2) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P44	pT <sub>1-3</sub> pN <sub>1</sub> (2/26) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P45	pT <sub>4</sub> pN <sub>2</sub> (4/36) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P46	pT <sub>3b</sub> pN <sub>2</sub> (8/16) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	

Patient code	Histopathology	Chromosome 8 BAC microarray	Gene expression microarray	185K oligonucleotide microarray
CC-P47	pT <sub>3</sub> pN <sub>2</sub> (12/13) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	(×)	×	×
CC-P48	pT <sub>3a</sub> pN <sub>2</sub> (5/23) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P49	pT <sub>4</sub> pN <sub>2</sub> (9/21) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	(×)	×	×
CC-P51	pT <sub>3c</sub> pN <sub>2</sub> (4/23) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P53	pT <sub>4</sub> pN <sub>2</sub> (11/26) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	(×)	×	×
CC-P54	pT <sub>3</sub> pN <sub>1</sub> (3/22) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	(×)	×	×
CC-P56	pT <sub>3b</sub> pN <sub>1</sub> (2/20) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P58	pT <sub>3</sub> pN <sub>2</sub> (1/32) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P60	pT <sub>3</sub> pN <sub>1</sub> (2/24) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	×
CC-P65	pT <sub>3</sub> pN <sub>1</sub> (2/22) M <sub>0</sub> R <sub>0</sub> G <sub>2,3</sub>	×	×	×
CC-P66	pT <sub>2</sub> pN <sub>2</sub> (4/20) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P68	pT <sub>3c</sub> pN <sub>2</sub> (12/22) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	
CC-P70	pT <sub>3</sub> pN <sub>2</sub> (12/21) M <sub>0</sub> R <sub>0</sub> G <sub>2</sub>	×	×	
CC-P71	pT <sub>3</sub> pN <sub>1</sub> (1/18) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	×
CC-P72	pT <sub>2</sub> pN <sub>1</sub> (2/18) M <sub>0</sub> R <sub>0</sub> G <sub>3</sub>	×	×	×

NOTE: (×), not included.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Colocalization of chromosomal breakpoints and structural variants in the genome

BP ID	Cytoband	Breakpoint start	Breakpoint stop	CNV locus ID	Segmental duplications	Interchromosomal/intrachromosomal	Patient ID (CC-P no.)	Genes mapping to the breakpoint
1.1	1p36.33	1,594,502	1,686,141	0003	DC0021, DC0022	Intra	14	SSU72, CDC2L2
1.2	1p35.2	31,435,129	31,444,435	0053	—	—	14	WDR57, ZCCHC17
1.3	1p13.2–1p13.1	113,299,763	113,313,396	0141	—	—	56	—
1.4	1p11.2–1p12	120,357,104	120,525,002	0144	DC0153, DC0154	Intra	10, 11, 16, 42	—
1.5	1q12	120,961,845	141,468,205	0145	DC0165	Intra	20	—
1.6	1q21.1–1q12	143,526,765	143,543,580	0146	DC0185	Intra	1, 44, 53, 56, 71	NBPF11
1.7	1q21.1–1q12	145,338,927	145,359,805	0148	DC0206	Inter	1	NBPF15
1.8	1q21.1	146,599,106	146,628,218	0148	DC0218	Inter	1	HIST2H4, H2BE, HIST2H3C
1.9	1q23.1	154,677,866	154,699,522	—	DC0238	Inter	1	—
2.1	2p25.1	8,053,090	8,078,648	0278	—	—	65	LOC339789
2.2	2p13.1	73,638,384	73,655,615	0354	—	—	14	ALMS1
2.3	2q34	211,358,859	211,368,373	0485	—	—	8	CPS1
3.1	3p24.3	20,205,874	20,219,511	0544	—	—	48	—
3.2	3p14.2	60,979,549	60,993,090	0597	—	—	53	FHIT
3.3	3p12.3	74,382,590	74,402,458	0612	—	—	9	CNTN3
3.4	3q11.2	95,112,766	95,126,868	—	DC0913	Intra	47	PROSI
3.5	3q13.11	105,937,188	105,980,769	0639	—	—	42	—
3.6	3q13.33	121,044,859	121,064,116	0658	—	—	8	GSK3B
4.1	4p15.31	20,205,995	20,212,010	0784	—	—	16	SLIT2
4.2	4q13.2	69,310,017	69,319,707	0867	DC1229	Intra	24	—
4.3	4q13.3	71,229,595	71,238,968	0869	—	—	9	—
4.4	4q22.3	94,893,243	94,904,635	0906	—	—	9	GRID2
4.5	4q35.1	184,012,774	184,022,817	1051	—	—	20	—
5.1	5p14.3	20,507,734	20,530,987	1090	—	—	8	—
5.2	5p14.3	20,507,734	20,530,987	1090	—	—	19, 24	—
5.3	5p12	45,125,704	45,157,897	1116	—	—	14	—

BP ID	Cytoband	Breakpoint start	Breakpoint stop	CNV locus ID	Segmental duplications	Interchromosomal/intrachromosomal	Patient ID (CC-P no.)	Genes mapping to the breakpoint
5.4	5q21.1	99,432,472	99,447,496	1183	DC1732, DC1733	Intra, Intra	51	—
5.5	5q21.1	100,703,233	100,734,459	1184	—	—	24	—
5.6	5q23.1	115,529,614	115,544,817	1203	—	—	42	COMMD10
5.7	5q33.2	153,090,806	153,106,089	1242	—	—	14	GRIA1
5.8	5q34	165,144,873	165,198,317	1251	—	—	8	—
6.1	6p21.33-6p22.1	29,949,864	29,967,114	1311	DC1955	Intra	20	—
6.2	6p21.33	30,085,931	30,095,973	—	DC1958	Intra	23	—
6.3	6p21.1	46,134,331	46,141,981	1330	—	—	8	—
6.4	6q12	65,074,130	65,145,208	1353	—	—	8, 20	—
6.5	6q13	73,616,917	73,628,301	1369	—	—	23	KCNQ5
6.6	6q25.1	149,905,776	149,946,488	—	DC2020	Intra	48	C6orf72, PP1L4
7.1	7p21.3	7,153,539	7,187,740	1491	—	—	1, 47	C1GALT1
7.2	7p11.2	55,572,233	55,593,091	—	DC2177	Inter	47, 48	—
7.3	7q11.21	62,195,037	62,271,252	1558	DC2233	Inter	65	—
7.4	7q11.23	76,397,789	76,426,065	1572	DC2356, DC2357	Intra	20, 48	KIAA1505
7.5	7q21.3	97,132,792	97,132,822	1596	DC2371	Inter	48	ASNS
7.6	7q22.1	101,837,056	101,878,089	1604	DC2398, DC2399	Intra, Inter	48	RASA4, POLR2J2
7.7	7q22.1	101,914,350	101,929,241	1604	DC2399	Inter	1, 20	—
7.8	7q32.2	130,012,776	130,017,714	1642	—	—	42	—
8.1	8p23.1	7,709,141	7,735,365	1691	DC2508	Inter	48	DEFB106A, DEFB104A, DEFB105A,
8.2	8p22	12,916,574	12,922,059	1698	—	—	72	FLJ36980, KIAA1456
8.3	8p22	16,642,931	16,684,703	1704	—	—	48	—
8.4	8p11.23-8p12	38,471,875	38,487,158	1733	—	—	65	—
8.5	8p11.22-8p11.23	39,356,595	39,369,686	1734	—	—	9, 51, 65	ADAM18
8.6	8q11.1	47,658,706	47,680,223	1742	—	—	10	—
8.7	8q11.21	51,143,041	51,151,309	1749	—	—	10	SNTG1
8.8	8q22.1	93,686,213	93,707,355	1811	—	—	1	—
9.1	9p21.1	30,929,344	30,963,839	1931	—	—	14	—
9.2	9p13.1	38,758,232	38,799,072	1944	DC2781, 2806, 2807	Intra, Inter	8, 9, 12, 42, 47, 71	—
9.3	9q12	68,141,956	68,151,359	1945	DC2382	Inter	8, 12, 14	KGFLP1, FOXPDL3

Camps et al.

Page 22

BP ID	Cytoband	Breakpoint start	Breakpoint stop	CNV locus ID	Segmental duplications	Interchromosomal/intrachromosomal	Patient ID (CC-P no.)	Genes mapping to the breakpoint
9.4	9q21.33-9q22.1	87,362,045	87,376,119	1960	—	—	8	<i>DAPK1</i>
9.5	9q31.1	104,443,646	104,448,713	1980	—	—	8	<i>OR13C2</i>
10.1	10p11.21	37,523,207	37,536,450	2086	DC3033	Inter	71	<i>ANKRD30A</i>
10.2	10p11.21	37,561,205	37,573,675	2086	—	—	8	<i>ANKRD30A</i>
10.3	10q11.21	42,066,866	42,097,773	2093	DC3061, DC6062	Inter	71	—
10.4	10q11.21	42,676,347	42,724,171	2093	DC3077, DC3078	Inter	8	—
10.5	10q11.22	45,489,352	45,507,880	2095	DC3091	Intra	14, 49	—
10.6	10q21.1	58,591,323	58,643,177	2111	—	—	8	—
10.7	10q21.3	66,455,837	66,527,887	2124	—	—	14	—
11.1	11p15.5	1,341,557	1,478,016	2201	—	—	38	<i>HCCA2, BRSK2</i>
11.2	11p15.4	3,625,683	3,638,563	2204	—	—	38	<i>ART1</i>
11.3	11q24.2	124,745,557	124,758,015	2349	—	—	8	<i>PKNOX2</i>
12.1	12p13.31	8,908,348	8,909,373	2370	—	—	8	—
12.2	12p13.2	11,040,946	11,045,834	2374	—	—	42	<i>TAS2R49</i>
12.3	12p13.2	11,113,176	11,274,374	2374	DC3675	Intra	45	<i>TAS2R46, TAS2R43, TAS2R44</i>
12.4	12q13.11	47,046,663	47,074,456	2418	—	—	49	—
13.1	13q12.11	21,020,530	21,029,004	2540	—	—	20	<i>EFHA1</i>
13.2	13q21.31	62,773,029	62,811,083	2578	—	—	65	—
13.3	13q21.33	70,415,095	70,460,603	2590	—	—	47	—
13.4	13q31.3	89,088,605	89,156,716	2615	—	—	23	—
13.5	13q33.3	108,193,190	108,202,800	2631	—	—	47	<i>RPI1-54H7.1</i>
13.6	13q34	111,053,779	111,083,506	2633	—	—	47	—
14.1	14q11.2	20,923,484	20,941,004	2641	—	—	42	<i>CHD8</i>
14.2	14q21.1	38,759,886	38,773,059	2668	—	—	19	<i>MIA2</i>
14.3	14q21.1	40,612,280	40,647,692	2671	—	—	51	—
14.4	14q21.2	43,688,550	43,723,219	2676	—	—	9	—
14.5	14q32.33	105,881,961	105,902,383	2747	—	—	9	<i>KIAA0125</i>
15.1	15q21.1	43,050,880	43,057,665	2772	DC3877	Intra	53	<i>C15ORF43</i>
16.1	16p13.3	5,085,210	5,132,553	2870	DC3528	Inter	20	<i>FAM86A</i>
16.2	16p12.1	22,578,173	22,625,780	2893	—	—	45, 65	—

Camps et al.

Page 23

BP ID	Cytoband	Breakpoint start	Breakpoint stop	CNV locus ID	Segmental duplications	Interchromosomal/intrachromosomal	Patient ID (CC-P no.)	Genes mapping to the breakpoint
16.3	16p11.2	28,141,539	28,154,008	2899	—	—	20	—
16.4	16p11.2	32,181,810	32,206,373	2905	DC3586	Inter	14	—
16.5	16q13	56,228,731	56,242,451	2924	—	—	47	<i>GPR56</i>
16.6	16q22.1	68,534,408	68,575,859	2935	DC3625	Inter	45	<i>LOC348174</i>
16.7	16q22.3	73,128,535	73,148,905	2940	DC3635	Intra	45	<i>GLG1</i>
17.1	17p11.2	17,935,146	17,945,526	2998	—	—	71	<i>DRG2</i>
17.2	17p11.2	18,823,700	18,845,631	3001	—	—	51	<i>SLC5A10</i>
17.3	17p11.2	21,471,897	21,627,596	3005	DC3225, DC3226, DC3227	Inter	10, 48, 72	—
17.4	17p11.2	22,324,326	22,367,302	3006	DC3241, DC3242, DC3243	Intra, Inter	1, 16, 20, 24, 42, 45, 53, 56, 71	—
17.5	17q21.2	35,991,941	36,017,353	3020	—	—	56	—
17.6	17q21.2	37,142,507	37,148,208	3021	—	—	56	<i>HAP1</i>
18.1	18p11.21	14,968,075	15,042,839	3109	DC2923, DC2924, DC2925	Inter	1	—
18.2	18q11.2	18,905,068	18,910,922	3112	—	—	1	—
19.1	19q13.12	42,671,304	42,678,874	3252	—	—	42	—
19.2	19q13.32	53,178,403	53,192,305	3271	—	—	71	<i>ELSPBP1</i>
20.1	20p13	98,836	111,637	3294	—	—	47	—
20.2	20p13	1,546,858	1,557,784	3300	—	—	47	<i>SIRPB1</i>
20.3	20p12.1	14,802,986	14,816,751	3327	—	—	9, 44	<i>C20orf133</i>
20.4	20p12.1	15,234,450	15,249,882	3330	—	—	1, 9, 47	<i>C20orf133</i>
20.5	20p11.22	22,297,096	22,316,323	3347	—	—	65	—
20.6	20p11.1	26,023,784	26,096,870	3352	DC2471, DC2472	Inter	10, 42, 49	—
20.7	20q11.1	28,209,786	28,225,117	3353	DC2479, DC2480	Inter	12, 14, 16, 20, 44, 48, 53, 71	—
20.8	20q13.33	58,950,616	59,000,615	3414	—	—	39	—
20.9	20q13.33	62,234,272	62,235,717	3419	—	—	39	—
21.1	21q21.1	21,148,687	21,166,248	3430	—	—	44	<i>NCAM2</i>
21.2	21q21.2	24,275,094	24,312,332	3435	—	—	14	—

Camps et al.

Page 24



BP ID	Cytoband	Breakpoint start	Breakpoint stop	CNV locus ID	Segmental duplications	Interchromosomal/intrachromosomal	Patient ID (CC-P no.)	Genes mapping to the breakpoint
X.1	Xp11.22	52,775,986	52,782,545	—	DC1465	Intra	42	<i>GAGED4</i>
X.2	Xq21.31	91,052,975	91,151,194	—	DC1550, DC1551	Inter	71	<i>PCDH11X</i>
X.3	Xq22.1	101,404,534	101,556,005	—	DC1569	Intra	8	<i>MXF2</i>

Camps et al.