

Association of survival and disease progression with chromosomal instability: A genomic exploration of colorectal cancer

Michal Sheffer^{a,1}, Manny D. Bacolod^{b,1}, Or Zuk^c, Sarah F. Giardina^b, Hanna Pincas^b, Francis Barany^b, Philip B. Paty^d, William L. Gerald^{e,3}, Daniel A. Notterman^{f,2}, and Eytan Domany^{a,2,4}

^aDepartment of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel; ^bDepartment of Microbiology, Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021; ^cBroad Institute, Massachusetts Institute of Technology and Harvard University, 7 Cambridge Center, Cambridge, MA 02142; Departments of ^dPathology and ^eSurgery, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065; and ^fDepartment of Molecular Biology, Princeton University, Princeton, NJ 08544

Communicated by Thomas E. Shenk, Princeton University, Princeton, NJ, February 27, 2009 (received for review October 2, 2008)

During disease progression the cells that comprise solid malignancies undergo significant changes in gene copy number and chromosome structure. Colorectal cancer provides an excellent model to study this process. To identify and characterize chromosomal abnormalities in colorectal cancer, we performed a statistical analysis of 299 expression and 130 SNP arrays profiled at different stages of the disease, including normal tissue, adenoma, stages 1–4 adenocarcinoma, and metastasis. We identified broad (> 1/2 chromosomal arm) and focal (< 1/2 chromosomal arm) events. Broad amplifications were noted on chromosomes 7, 8q, 13q, 20, and X and broad deletions on chromosomes 4, 8p, 14q, 15q, 17p, 18, 20p, and 22q. Focal events (gains or losses) were identified in regions containing known cancer pathway genes, such as *VEGFA*, *MYC*, *MET*, *FGF6*, *FGF23*, *LYN*, *MMP9*, *MYBL2*, *AURKA*, *UBE2C*, and *PTEN*. Other focal events encompassed potential new candidate tumor suppressors (losses) and oncogenes (gains), including *CCDC68*, *CSMD1*, *POLR1D*, and *PMEPA1*. From the expression data, we identified genes whose expression levels reflected their copy number changes and used this relationship to impute copy number changes to samples without accompanying SNP data. This analysis provided the statistical power to show that deletions of 8p, 4p, and 15q are associated with survival and disease progression, and that samples with simultaneous deletions in 18q, 8p, 4p, and 15q have a particularly poor prognosis. Annotation analysis reveals that the oxidative phosphorylation pathway shows a strong tendency for decreased expression in the samples characterized by poor prognosis.

colon cancer | DNA copy number | gene expression | SNP arrays

The development of colorectal cancer from normal epithelial cells through benign adenomas to malignant carcinomas and metastasis is a lengthy, multistep process, involving accumulation of mutations of key regulatory genes (1). Two forms of genetic instability are described in colon cancer. Samples displaying Chromosomal *IN*stability (*CIN*) comprise 85% of the cases, and the rest display *Microsatellite IN*stability (*MIN*)—deletions and amplifications of short sequences of nucleotides. *MIN* tumors are usually euploid, while *CIN* tumors are aneuploid; these genetic differences are reflected in a different pathological and clinical behavior (2). Numerous array comparative genomic hybridization (*CGH*) and SNP chip studies (3–5) (*SI References*) have shown that the principal large-scale DNA copy number changes in *CIN*-colorectal cancer are gains of chromosomes 7, 8q, 13q, and 20 and losses of 4, 8p, 17p, 18, and 20p.

Genomic aberrations that recur in many different samples may reflect an underlying selection advantage. Therefore, they often include genes that are important for tumor development, propagation, and spread (6). Tumor suppressor genes (*TSG*) are generally inactivated through mutation or deletion (generally of both copies) while oncogenes become active through mutation or am-

plification (generally of 1 copy). Tsafir et al. (7) showed, contrary to previous claims (8), that changes in expression level were correlated with alterations in gene dose, suggesting that coherent transcriptional changes associated with large chromosomal segments can be attributed to corresponding changes in chromosome structure, such as copy number variation. They also showed that aberrations became more frequent as disease advanced.

Using novel methods of analysis and a much larger sample size than available to Tsafir et al. (7), we planned to create a comprehensive map of the common genomic aberrations in colon cancer and identify the genes encompassed by these events. Data that were collected from patients at a variety of clinical stages allowed us to associate chromosomal changes with survival and disease progression and to identify genetic and biochemical pathways that are affected by these aberrations. Information from DNA copy number (from Affymetrix 50 k SNP arrays, see *Methods*) and gene expression (from Affymetrix U133A arrays) was integrated. First, we used the SNP data to identify amplification and deletion events that occur in a large subset of samples. We refer to these collectively as *CINons*. Next, data from tumor samples for which we had both copy number and expression were used to identify genes, the expression of which was significantly correlated with the copy number implied by their cognate SNPs. We term these “correlated genes” and focus our attention on them as potential cancer pathway candidates. Finally, we used the tumors with both SNP (i.e., copy number) and expression data to establish that for some *CINons* we can impute the copy number change of individual samples from the associated expression data. This allowed us to construct an *in silico* chromosomal alteration map for the much larger number of tumors for which only expression data were available. This procedure significantly enhanced the statistical power of our search for linkage between survival and amplifications/deletions. Ultimately, we identify several new candidate cancer pathway genes in colon cancer, and demonstrate a strong association between deletions of 18q, 8p, 4p, and 15q and outcome. This association may be related to reduced expression of genes annotated to the oxidative phosphorylation pathway.

Results and Discussion

Amplifications and Deletions in Colon Cancer (SNP Data). Fig. 1A shows the smoothed $CR_{n,s}$ (log copy number ratio for SNP n and

Author contributions: M.S., M.D.B., F.B., D.A.N., and E.D. designed research; M.S., M.D.B., S.F.G., H.P., F.B., P.B.P., W.L.G., D.A.N., and E.D. performed research; M.S., O.Z., and E.D. analyzed data; and M.S., D.A.N., and E.D. wrote the paper.

The authors declare no conflict of interest.

¹M.S. and M.D.B. contributed equally to this work.

²E.D. and D.A.N. contributed equally to this work.

³Deceased September 14, 2008.

⁴To whom correspondence should be addressed. E-mail: eytan.domany@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902232106/DCSupplemental.

Table 1. Broad, focal, and peak CINons found by the 2 configurations of GISTIC

Id	Chr	Start, Mb	End, Mb	Type	Samples, %	Noncoding RNA	Correlated* probe sets	Known and candidates oncogenes/TSG
1	2	q37.1	235.26	Focal	27	1	6/25	
2	6	p21.1	42.43	Focal	22	1	21/35	VEGFA
3	7	p	1.50	Broad	53	11	72/156	
4	7	q	61.71	Broad	42	22	157/355	
5	7	q31.2	115.92	Peak in 4	42	0	1/5	MET
6	7	q36.1	147.56	Peak in 4	44	0	2/3	
7	8	p	40.65	Focal	33	1	6/20	
8	8	q	47.40	Broad	56	12	111/216	
9	8	q11.22	52.31	Peak in 8	51	1	11/18	LYN
10	8	q24.13	125.01	Peak in 8	60	0	8/9	MYC
11	12	p13.32	4.33	Focal	18	0	2/7	FGF6, FGF23
12	13	Q	18.35	Broad	73	14	121/206	
13	13	q12.2	27.13	Peak in 12	78	0	1/1	POLR1D
14	16	q12.1	50.21	Focal	18	0	0/3	
15	17	q11.1	22.51	Focal	29	0	0/3	
16	17	q25.3	73.52	Focal	31	2	21/55	
17	20	p	0.10	Broad	47	8	46/93	
18	20	q	29.31	Broad	91	22	143/209	
19	20	q12–13.12	40.46	Peak in 18	93	0	22/40	MMP9, MYBL2, UBE2C
20	20	q13.2	53.72	Peak in 18	91	3	14/23	AURKA, PMEPA1
21	X	p	2.83	Broad	35	14	62/174	
22	X	q	62.48	Broad	38	63	67/265	
23	1	p	0.84	Focal	36	19	91/309	
24	1	p36.22	11.21	Peak in 23	36	9	50/157	
25	4	p	0.40	Broad	27	8	38/117	
26	4	q	52.38	Broad	25	14	88/304	
27	8	p	0.18	broad	58	9	58/153	
28	8	p23.2	5.76	Peak in 27	60	0	0/1	CSMD1
29	10	q23.31	89.89	Focal	20	0	1/2	PTEN
30	14	q	19.30	Broad	35	102	164/359	
31	15	q	19.92	Broad	42	97	109/342	
32	15	q11.2	21.60	Peak in 31	44	76	1/6	
33	15	q21.3	55.12	Peak in 31	45	0	3/10	
34	15	q26.2	93.07	Peak in 31	40	0	1/2	
35	16	p13.3	6.16	Focal	9	0	0/1	
36	17	p	0.45	Broad	47	24	73/177	
37	18	p	0.15	Broad	56	0	26/46	
38	18	q	16.84	Broad	71	10	55/111	
39	18	q21.2	50.79	Peak in 38	73	0	1/2	CCDC68
40	20	p	0.10	Broad	22	7	37/74	
41	20	p13	0.10	Peak in 40	22	6	20/36	
42	20	p12.1	14.45	Peak in 40	25	0	0/1	
43	20	p12.1	15.20	Peak in 40	22	0	1/2	
44	20	p12.1	17.74	Peak in 40	20	1	5/9	
45	22	q	15.51	Broad	24	14	51/266	

CINons 1–22 are amplicons and 23–45 deletions. The broad CINons (including the focal CINon 7, see text) were identified by configuration 1, the focal CINons 11, 23, and 29 were identified by both configurations, and the rest by configuration 2. *, see *Methods*.

pressed in many different types of cancers, and is associated with the degree of tumor differentiation in different types of carcinomas (19). Another peak amplification, on 20q13.2, contains *AURKA*, a serine/threonine kinase known as a key regulator of the mitotic cell division process, which has been identified as an oncogene overexpressed in many human cancers (20).

Candidate Cancer Pathway Genes Correlated with Broad and Focal Events. Some of the focal and peak CINons that we found contained candidate cancer pathway genes not previously known to play a role in colon cancer (a list of the correlated genes in these CINons is found in [Table S1](#)). One such candidate gene is *CCDC68* (18q21.2), also termed cutaneous T cell lymphoma-associated antigen se57–1. The role of *CCDC68* in cancer is not known. In colon cancer, *CCDC68* is expressed on the cancer cell surface in about 15% of patients (21). In our data this gene is downregulated in 89% of primary tumors (compared with normal tissue) and its expression is highly correlated with the associated gene copy number ($r = 0.51$; $P = 3.6e-5$). These observations suggest that downregulation of this

surface protein is an important selective advantage for colon cancer cells and that it should be studied as a potential TSG. *CCDC68* is the gene closest to the peak deletion identified on chromosome 18 (Fig. 1C), which is also close to *DCC*, a known TSG in colon cancer, whose expression level was absent in our samples. *CSMD1*, located near the peak of the 8p deletion (8p23.2), is not represented on the microarray. Farrell et al. (22) also suggested this gene as a candidate TSG inactivated in colon cancer.

Several candidates are located in amplified regions and are overexpressed. *PMEPA1*, located on 20q13.31, is an androgen-induced gene expressed in normal prostate tissue. The role of *PMEPA1* in cancer is unclear because it shows overexpression in some colon, breast, and ovarian cancers but is underexpressed in prostate cancer. *PMEPA1* is known to be induced by the *TGF-β* signaling pathway, *EGF*, *IGF*, and *PDGF* (23), and it serves to inhibit *TGF-β* signaling by interfering with *TGF-β* type I receptor (*TGFBRI*)-induced *R-Smad* phosphorylation. Because *TGF-β* mediates an apoptotic pathway in colon cancer, the effect of *PMEPA1* overexpression could be to promote cancer cell growth by fore-

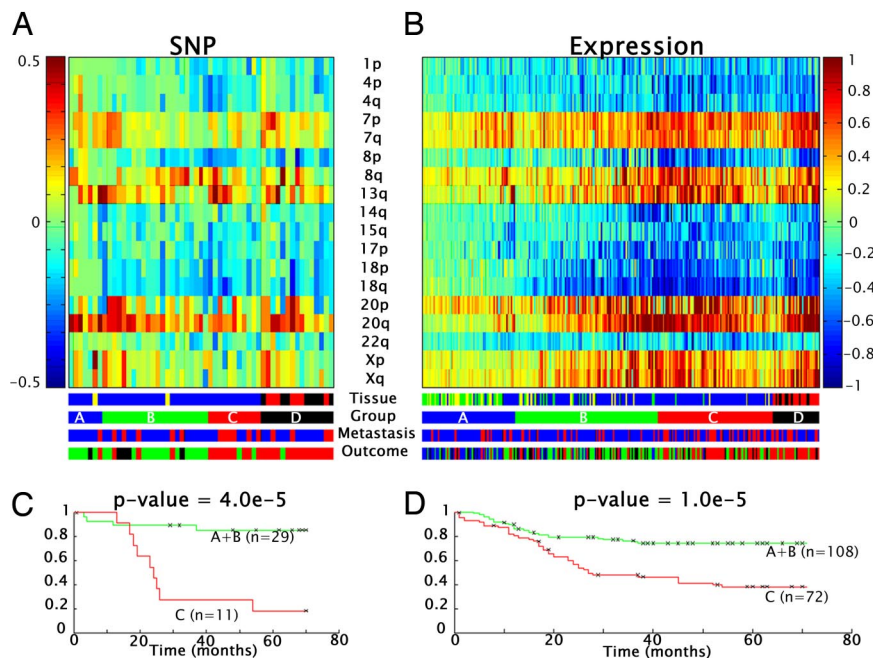


Fig. 2. CINon copy number tables, from SNP and expression. (A) Copy number table of the broad CINons (including the focal CINon 1p) as obtained from SNP data, for 55 aneuploid tumors; each row represents a broad CINon and each column a sample. An entry is calculated as the median of $CR_{n,s}$: the SNPs within the CINon, in the corresponding sample (see color bar to Right of figure). (B) CINon expression table, for 256 adenomas and tumors, similar to the copy number table (A). The entries were calculated as the median of the “correlated genes” of the CINon (see text and Methods). In both A and B the samples are divided into 4 groups as shown in the second color bar at the Bottom (in the text); the first color bar represents the tissue origin of the sample where primary tumor is blue, MIN tumor is yellow, adenoma is green, sample from liver metastasis is red, and sample from lung metastasis is black; in the third color bar the stage 4 tumors (primary samples presenting with metastasis) are in red; the fourth color bar represents outcome, where adenomas are in blue, good outcome samples in green, poor outcome in red, and unknown outcome in black. (C and D) Kaplan-Meier plot of primary tumor samples that belong to groups A + B (green) versus group C (red) of copy numbers and expression, respectively. Follow-up intervals that were greater than 70 months were assigned 70.

stalling apoptosis. In our data set, *PMEPA1* is overexpressed in 84% of the primary tumors (> 2 -fold). Expression of *PMEPA1* exhibits high correlation with the associated copy numbers ($r = 0.43$, $P = 9.9e-4$).

The peak region of 13q contains *POLR1D*, a subunit of both RNA polymerases I and III. RNA polymerase I is involved in the production of 18S, 5.8S, and 28S rRNAs, while RNA polymerase III synthesizes small essential RNAs, such as tRNAs, 5S rRNA, and some snRNAs (24). There is not much information regarding this gene in the literature. *POLR1D* is overexpressed in 42% of the primary tumors (> 2 -fold), showing high correlation between expression and copy number ($r = 0.7$, $P = 8.6e-11$). Note that in (25) another gene, *CDK8*, located 1200 kb upstream *POLR1D*, was associated with this peak amplification.

We found 973 genes the expression of which differs between samples with low and high expression of *PMEPA1*. They were enriched by genes expressing proteins located in the oxidative phosphorylation pathway. Similar analyses were done for *POLR1D* and *CCDC68* (see Table S2).

Noncoding RNA. Some peak and focal CINons contain small noncoding RNA (Table 1). hsa-mir-103-2, identified on the (deleted) peak region of 20p13, was implicated in colon cancer and was found to be downregulated in tumors (26). The peak CINon on chromosome 15 is part of a region that is subjected to genomic imprinting, associated with Prader-Willi syndrome and Angelman syndrome (27). This peak region consists of 76 C/D snoRNAs, organized in 2 large clusters. Currently, these snoRNA are not known to be related to cancer. Another imprinted locus that consists of large clusters of C/D snoRNAs and microRNAs is located on 14q32 (27). Deletion of C/D snoRNAs could affect methylation of rRNAs or the function of core C/D box snoRNP. Additionally, chromosome

14q is among broad regions that were found to be deleted in colon cancer by others (3, 4) and also by us. Inactivation of imprinted regions requires the silencing of only 1 allele and therefore may serve as an efficient mechanism for inactivating TSGs.

Deriving Copy Number Changes from Expression Data. One of our principal aims was to search for associations between chromosomal events and clinically relevant features of the tumors, such as survival and progression. To enhance statistical power and reliability, this analysis must be based on as large a number of samples as possible. Because expression data are available for many more samples than are SNP data, it would be useful to impute copy number from expression (28, 29). To establish a reliable procedure, we constructed 2 chromosomal instability maps (see Methods), as shown in Fig. S1, for 45 aneuploid tumor samples, which had both expression and SNP data (see Table S3). To minimize variance because of small numbers of genes, we treated only broad CINons and the focal CINon of 1p (because it contained many probe sets). The similarity of the 2 maps is striking, and indeed, the correlations between the entries for each CINon were highly significant ($r > 0.6$ at 10% false discovery rate [FDR]). On the basis of these observations we built an expression-based CINon table (heat map) for all 256 adenoma and tumor samples for which we had expression data (including liver and lung metastasis); this table (heat map) is shown in Fig. 2B and serves as the basis of the analysis described below.

Association of Aberrations with Survival, Disease Progression, and Other Factors. We now searched for connections between the different aberrations of Fig. 2B and survival. As expected, there is strong correlation between clinical state and outcome in colon cancer patients: in our data, almost all stage 4 patients had poor outcome (either death of disease or recurrence within 60 months)

HIF1 even under normoxic conditions (37, 39), as does accumulation of citric acid cycle constituents, succinate, or fumarate (37).

Several explanations have been proffered to explain this “glycolic shift”: (i) the use of glycolytic metabolites such as ribose or citrate to enhance cell proliferation (37); (ii) high concentrations of lactate create an acidic environment that is advantageous to the process of invasion (40); (iii) decreased mitochondrial function reduces reactive oxygen species (ROS) production; however, the role of ROS in hypoxia-induced death is controversial (37); (iv) the shift to glycolysis also makes the tumor mitochondria less susceptible to permeabilization of the outer mitochondrial membrane, and as a result, less sensitive to apoptosis (40); and (v) the shift helps to conserve O₂ for alternative use such as sterol synthesis and oxidative protein folding (37). Our analysis indicates that one mechanism by which the neoplastic cell downregulates oxidative phosphorylation is deletion of the relevant genes and the resulting decrease in the activity of this pathway. This task of identifying all of the specific genes is challenging and not straightforward, perhaps because many different pathways converge on this essential cellular function. For example, based on the observation that *PTEN* loss results in *HIF1* accumulation (37), one might have anticipated an effect of *PTEN* loss on expression of oxidative phosphorylation genes. However, this was not observed (the mean *r* of *PTEN* expression with that of the genes in Fig. 3 was 0.13). Further insight may come from additional analysis of the different regulatory elements of this pathway, their status in the neoplastic cell, and the relations between them.

Methods

Expression Data. The 299 U133A arrays (see Table S3) were subjected to MAS 5, threshold, and log₂ transformation (see SI Methods and Fig. S3). Log-expression ratios $ER_{n,s}$ were calculated by subtracting from the log-transformed expression value of probe set *n* in sample *s* the median log-transformed expression value of probe set *n* in the normal colon samples of the same batch (2 batches were identified, because of protocol change). For chromosome X, all normal samples of the same gender and batch were used.

1. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61:759–767.
2. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396:643–649.
3. Jones AM, et al. (2005) Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene* 24:118–129.
4. Kurashina K, et al. (2008) Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. *Cancer Sci* 99:1835–1840.
5. Leslie A, et al. (2003) Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas. *Cancer Res* 63:4656–4661.
6. Albertson DG, Collins C, McCormick F, Gray JW (2003) Chromosome aberrations in solid tumors. *Nat Genet* 34:369–376.
7. Tsafir D, et al. (2006) Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res* 66:2129–2137.
8. Platzer P, et al. (2002) Silence of chromosomal amplifications in colon cancer. *Cancer Res* 62:1134–1138.
9. Beroukhi R, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc Natl Acad Sci USA* 104:20007–20012.
10. Ferrara N (2004) Vascular endothelial growth factor: Basic science and clinical progress. *Endocr Rev* 25:581–611.
11. Eswarakumar VP, Lax I, Schlessinger J (2005) Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev* 16:139–149.
12. Goel A, et al. (2004) Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res* 64:3014–3021.
13. Buffart TE, et al. (2005) DNA copy number changes at 8q11–24 in metastasized colorectal cancer. *Cell Oncol* 27:57–65.
14. Bates RC, Edwards NS, Burns GF, Fisher DE (2001) A CD44 survival pathway triggers chemoresistance via lyn kinase and phosphoinositide 3-kinase/Akt in colon carcinoma cells. *Cancer Res* 61:5275–5283.
15. Subramaniam V, et al. (2007) CD44 regulates cell migration in human colon cancer cells via Lyn kinase and AKT phosphorylation. *Exp Mol Pathol* 83:207–215.
16. Furge KA, Zhang YW, Vande Woude GF (2000) Met receptor tyrosine kinase: Enhanced signaling through adapter proteins. *Oncogene* 19:5582–5589.
17. Rundhaug JE (2005) Matrix metalloproteinases and angiogenesis. *J Cell Mol Med* 9:267–285.
18. Sala A (2005) B-MYB, a transcription factor implicated in regulating cell cycle, apoptosis and cancer. *Eur J Cancer* 41:2479–2484.
19. Wagner KW, et al. (2004) Overexpression, genomic amplification and therapeutic potential of inhibiting the UbcH10 ubiquitin conjugase in human carcinomas of diverse anatomic origin. *Oncogene* 23:6621–6629.
20. Katayama H, Brinkley WR, Sen S (2003) The Aurora kinases: Role in cell transformation and tumorigenesis. *Cancer Metastasis Rev* 24:451–464.

SNP data are composed (see Table S3) of 130 50K XbaI SNP arrays (41). Log-copy number ratios $CR_{n,s}$ were calculated by subtracting from the log-transformed copy number of SNP *n* in sample *s* the log-transformed copy number of SNP *n* in the normal sample. The $CR_{n,s}$ values were smoothed with GLAD (33). See SI Methods for detailed normalization and preprocessing steps.

Identification of CINons. We used GISTIC (9). Q-values for the statistical significance of the copy number of each SNP were calculated for amplifications and deletions separately. CINons were then defined as regions of at least 5 contiguous SNPs with a *q*-value of <0.25 (for details see SI Methods). Two different configurations of GISTIC were used: *configuration 1* aimed at identifying all broad regions and *configuration 2* aimed at identifying small CINons (see SI Methods).

Correlation Between Expression and Copy Number. Detailed description is found in SI Methods. For each probe set *n*, the PCC was calculated between the $ER_{n,s}$ values of the samples that had both measures of expression and copy numbers (79 samples), and the copy number ratios measured in sample *s* from SNPs located near the probe set *n*. Each correlation was assigned with a *q*-value and FDR of 25% was then used to generate a list of “correlated genes.”

CINon Expression Table. This analysis was limited to CINons that contained many probe sets—i.e., the broad CINons and the focal CINon of 1p. Using the “correlated genes,” we constructed a CINon expression table; for each CINon *i* and sample *s* we calculated $CE_{i,s}$, the median ($ER_{n,s}$) evaluated for all probe sets *n* located on CINon *i*. A CINon copy number table was constructed in the same way, with entry $CC_{i,s}$ derived of CINon *i* and sample *s* by taking the median ($CR_{n,s}$) for all SNPs *n* that are located on CINon *i*. The CINon expression table may serve as a predictor of copy number for samples, for which no copy number data were available. To assess the quality of this prediction, PCCs of the $CE_{i,s}$ and $CC_{i,s}$ were calculated for each CINon *i* over the aneuploid tumor samples for which we had both expression and copy number measurements (45 samples, see Table S3 and Fig. S1).

ACKNOWLEDGMENTS. We thank G. Getz for help with implementation of GISTIC, and L. Hertzberg, and T. Shay for sharing their results. This work was funded by a Program Project Grant from the National Cancer Institute (P01-CA65930), a contract from the National Cancer Institute (263 MQ 610681), and by grants from the Ridgefield Foundation, the Ludwig Institute for Cancer Research/Conrad N. Hilton Foundation joint Hilton-Ludwig Cancer Metastasis Initiative, and the Gilbert Family Foundation.

21. Gerhardt A, et al. (2004) Tissue expression and sero-reactivity of tumor-specific antigens in colorectal cancer. *Cancer Lett* 208:197–206.
22. Farrell C, et al. (2008) Somatic mutations to CSMD1 in colorectal adenocarcinomas. *Cancer Biol Ther* 7:609–613.
23. Hirokawa YS, et al. (2007) High level expression of STAG1/PMEPA1 in an androgen-independent prostate cancer PC3 subclone. *Cell Mol Biol Lett* 12:370–377.
24. Martinez-Calvillo S, et al. (2007) Characterization of the RNA polymerase II and III complexes in *Leishmania major*. *Int J Parasitol* 37:491–502.
25. Firestein R, et al. (2008) CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 455:547–551.
26. Cummins JM, et al. (2006) The colorectal microRNAome. *Proc Natl Acad Sci USA* 103:3687–3692.
27. Royo H, Bortolin ML, Seitz H, Cavaille J (2006) Small non-coding RNAs and genomic imprinting. *Cytogenet Genome Res* 113:99–108.
28. Carter SL, et al. (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38:1043–1048.
29. Hertzberg L, et al. (2007) Prediction of chromosomal aneuploidy from gene expression data. *Genes Chromosomes Cancer* 46:75–86.
30. Al-Mulla F, et al. (2006) Metastatic recurrence of early-stage colorectal cancer is linked to loss of heterozygosity on chromosomes 4 and 14q. *J Clin Pathol* 59:624–630.
31. Bardi G, et al. (2004) Tumor karyotype predicts clinical outcome in colorectal cancer patients. *J Clin Oncol* 22:2623–2634.
32. Duensing A, Duensing S (2005) Guilt by association? p53 and the development of aneuploidy in cancer. *Biochem Biophys Res Commun* 331:694–700.
33. Hupe P, et al. (2004) Analysis of array CGH data: From signal ratio to gain and loss of DNA regions. *Bioinformatics* 20:3413–3422.
34. Dennis G, Jr., et al. (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* 4:P3.
35. Hosack DA, et al. (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol* 4:R70.
36. Tsafir D, et al. (2005) Sorting points into neighborhoods (SPIN): Data analysis and visualization by ordering distance matrices. *Bioinformatics* 21:2301–2308.
37. Denko NC (2008) Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat Rev Cancer* 8:705–713.
38. Warburg O (1956) On respiratory impairment in cancer cells. *Science* 124:269–270.
39. Bardos JI, Ashcroft M (2004) Hypoxia-inducible factor-1 and oncogenic signalling. *Bioessays* 26:262–269.
40. Gogvadze V, Orrenius S, Zhivotovsky B (2008) Mitochondria in cancer cells: What is so special about them? *Trends Cell Biol* 18:165–173.
41. Bacolod MD, et al. (2008) The signatures of autozygosity among patients with colorectal cancer. *Cancer Res* 68:2610–2621.