[CANCER RESEARCH 62, 5939–5946, October 15, 2002]

Dissecting Karyotypic Patterns in Colorectal Tumors: Two Distinct but Overlapping Pathways in the Adenoma-Carcinoma Transition[1]

Mattias Höglund,[2] David Gisselsson, Gunnar B. Hansen, Torbjörn Säll, Felix Mitelman, and Mef Nilbert

Department of Clinical Genetics, University Hospital, SE-221 85 Lund, Sweden [M. H., D. G., G. B. H., F. M.]; Department of Genetics, Lund University, SE-223 62 Lund, Sweden [T. S.]; and Department of Oncology, the Jubileum Institution, University Hospital, SE-221 85 Lund, Sweden [M. N.]

ABSTRACT

More than 500 colorectal tumors with clonal chromosomal abnormalities have been reported. Although the pattern of aberrations is nonrandom, no specific primary or secondary karyotypic abnormality has been identified. Also, the chronological order in which the aberrations appear during disease progression is not well known. One reason why our understanding of the cytogenetic evolution is unclear is the high degree of karyotypic complexity seen in these tumors. To overcome some of these difficulties we have previously used several statistical methods that allow identification and interpretation of karyotypic pathways as well as establishment of a temporal order of appearance of the imbalances. These methods were applied on 531 colorectal tumor karyotypes. By using a resampling strategy, 1p−, +7, 7q−, and +12p were identified as early events. Two major and two minor cytogenetic pathways were identified by means of principal component analysis. The two major pathways were initiated with 1p− and +7, and the minor pathways were initiated with +12p and 7q−. The +7/+12p tumors were found to be hyperdiploid, whereas those with 1p−/7q− were pseudodiploid. We also show that the adenoma-carcinoma transition in the 1p− pathway is strongly linked to karyoytypic evolution, whereas the +7 pathway is not, and that the cytogenetic pathways are separated at both early and late stages.

INTRODUCTION

Colorectal cancer is the third leading cause of cancer death worldwide (1). Although most patients undergo curative surgery, almost half of the patients will die from the disorder. The major predictor of outcome is disease stage at presentation, most often classified according to Dukes' staging. Prognosis is favorable for patients with stage A and B tumors, who have a 5-year survival of 70–90%, whereas patients with stage C and D tumors have survival rates of 10–40% at 5 years. Variations in precursor lesions, histopathological features, tumor site distribution, and clinical behavior suggest that different evolutionary modes exist for colorectal cancer (2). The molecular genetic model for colorectal carcinogenesis emphasizes a sequential accumulation of genetic alterations including deletions and point mutations of tumor suppressor genes such as *APC* and *TP53* and oncogenic mutations of *KRAS2* (3). At least two forms of genetic instability have been described in colorectal cancer: CIN[3] and MIN (4). About 60% of the tumors develop through the CIN pathway, which is characterized by aneuploidy, multiple chromosomal rearrangements, and clonal heterogeneity. Approximately 15–20% of colorectal tumors show the MIN phenotype. These tumors develop through a defective mismatch repair system, which results in a mutator phenotype with instability of repetitive sequences such as microsatellites. These tumors also acquire somatic frameshift mutations of coding repeats within several cancer-associated genes, and such alterations have been suggested to constitute a MIN-specific tumorigenic mechanism (5).

Cytogenetic studies have revealed few recurrent tumor-characteristic chromosome aberrations in colorectal tumors. The carcinomas mostly show highly complex karyotypes, whereas adenoma karyotypes are less complex (6). Balanced structural rearrangements are rarely seen, and the more complex karyotypes are dominated by various chromosomal imbalances, frequently including losses of the chromosome arms 1p, 8p, 13q, and 17p and gains of 7q, 8q, 11q, 13q, and 7p (7). Furthermore, more than one abnormal clone has been detected in almost half of the tumors, suggesting that clonal evolution is important in colorectal tumor progression.

Studies using CGH have independently confirmed the presence of several of the chromosomal imbalances identified by cytogenetic studies, such as gains of 20q, 13q, 7p, and 8q, and losses of 18q and 17p (8–11). CGH has also revealed a higher number of imbalances in carcinomas compared with adenomas and an increased number of aberrations in advanced tumors (Dukes' stages C and D) compared with less advanced tumors (Dukes' stages A and B; Ref. 10). Imbalances of 8p, 18q, and 20q have been associated with tumor progression, and imbalances affecting 8p and 18q have also been shown to predict poor prognosis in early-stage tumors (11, 12).

Because of the karyotypic complexity, analysis of karyotypic evolution is complicated, and an extensive amount of cytogenetic and/or CGH data are needed to reveal possible karyotypic pathways for tumor development. In the present investigation, we have identified all cytogenetically aberrant colorectal tumors (a total of 569 karyotypes) present in the Mitelman Database on Chromosome Aberrations in Cancer.[4] To identify the most frequent imbalances, we constructed a genomic imbalance map from all of the recorded cases. Tumor cases were then classified with respect to the presence or absence of these imbalances and analyzed statistically (13, 14) to assess the order of appearance of imbalances, possible karyotypic pathways, and cytogenetic subtypes.

MATERIALS AND METHODS

Selection of Data. A total of 569 karyotypes was retrieved from the Mitelman Database on Chromosome Aberrations in Cancer[4] and used to construct imbalance maps at the 368-band level. Segments affected by imbalances in >5% of the cases were identified, and for each region, the most affected segment was selected as the characteristic imbalance. A total of 42 imbalances, excluding –Y, was identified in this way (Table 1). Each karyotype was then assessed for the presence or absence of the selected imbalances. The NIPT was calculated, and karyotypes with at least one of the chosen imbalances were further analyzed. The resulting 531 cases were subdivided into adenomas (118) and carcinomas (348) according to the entries in the database,[4] and the following clinical parameters were recorded: (*a*) tumor size (< 1 cm >); (*b*) location (proximal or distal; tumors localized to the ascending or transverse colon were diagnosed as proximal, and tumors in the descending colon and rectum were classified as distal); (*c*) histopathological classification (adenomas: hyperplastic, tubular, or tubulovillous) or differentiation (carcinomas: low, moderate, or high grade); and (*d*) Dukes' stage.

Received 4/8/02; accepted 8/15/02.

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.

[1] Supported by the Swedish Cancer Society, the Crafoord Foundation, and the Nilsson Family Foundation.

[2] To whom requests for reprints should be addressed. Phone: 46-46-173739; Fax: 46-46-131061; E-mail: mattias.hoglund@klingen.lu.se.

[3] The abbreviations used are: CIN, chromosomal instability; MIN, microsatellite instability; NIPT, number of imbalances/tumor; TO, time of occurrence; PCA, principal component analysis; CGH, comparative genomic hybridization.

[4] http://cgap.nci.nih.gov/Chromosomes/Mitelman.

Table 1 *Summary of the frequency of imbalances*[a] *in colorectal tumors*

Imbalances	Abbreviations	Total tumor population (N = 531)	Adenomas (N = 118)	Adenocarcinomas (N = 348)
+1q10-q44	+1q	7	1	9
+2		9	8	9
+3		9	10	9
+5p10-p15	+5p	9	9	10
+6p10-p25	+6p	8	8	8
+7		37	48	34
+8q10-q24	+8q	20	16	21
+9		9	12	8
+11		7	2	8
+12p10-p13	+12p	9	5	11
+13q10-q34	+13q	28	34	25
+14q10-q32	+14q	6	10	5
+16		8	9	8
+17q10-q25	+17q	9	5	11
+19		8	8	7
+20q10-q13	+20q	20	23	19
+21q10-q22	+21q	6	2	5
+X		9	1	13
−1p36-p10	1p−	19	18	20
−3p26-p10	3p−	8	2	10
−3q26-q29	3q−	6	1	8
−5q10-q34	5q−	17	6	20
−6q22-q27	6q−	10	1	14
−7q32-q36	7q−	7	1	8
−8p23-p10	8p−	20	7	25
−9p24-p21	9p−	6	0	8
−9q22-q34	9q−	6	1	8
−10		11	2	13
−11q23-q25	11q−	7	3	8
−12q13-q24	12q−	9	1	11
−13p13-p10	13p−	11	1	14
−14p13-p10	14p−	15	4	17
−15		14	4	16
−16		9	1	10
−17p13-p10	17p−	30	11	36
−18q21-q23	18q−	30	19	35
−19	−19	8	1	10
−20		8	3	10
−21		12	4	16
−22		15	6	16
−X		8	1	9

[a] Frequencies are given as percentages. Only imbalances seen in more than 5% of the total tumor population are included.

Temporal Analysis. Early and late imbalances were defined as described in Höglund *et al.* (13). In brief, to obtain a value for the time of appearance of an imbalance, all tumors with each given imbalance were selected, and the distribution of NIPT was plotted. The modes of these distributions were then used as an estimate of lateness and referred to as the TO. TO is thus a function of karyotypic complexity (14). The selected distributions were then resampled with replacement (bootstrapped) 1000 times, and the TO was scored after each resampling (15, 16). The mean of the bootstrapped TO values was then used as the TO for the given imbalance. The bootstrapped 2.5, 25, 75, and 97.5 percentiles were also calculated. For the bootstrap estimates, resampling software from Resampling Stats (Arlington, VA) was used. Due to the small number of adenomas (118), the TO for the imbalances in this group was estimated from the median NIPT values, without bootstrap analysis. To evaluate whether an imbalance occurred earlier or later than expected from random events, a simulation procedure was used in which the distribution of NIPT for the whole tumor population and the frequencies of the imbalances were identical to the observed ones. The simulation was performed as described in Höglund *et al.* (13). An imbalance was considered to occur either significantly earlier (observed TO $<$ simulated TO) or significantly later (observed TO $>$ simulated TO) than expected when $>$90% of the simulated TO values deviated from the observed values.

PCA. PCA was performed using the Statistica software package (Statsoft, Tulsa, OK). PCA is a standard multivariate method frequently used to search for underlying structures in data sets (14, 17). In short, principal components are linear combinations of the original variables, orthogonal, and ordered with respect to their variance so that the first principal component has the largest variance. To analyze imbalances, these were used as variables, and the individual tumors were used as the observations; this will group imbalances frequently seen in the same tumors. The factor models arrived at were evaluated at three levels (total variance accounted for, communalities, and the residual correlation matrix). The fraction of the original variance accounted for by the specific factor model is given by the cumulative percentage of the explained variance for each component in the factor model. The communality of a given variable (imbalance) is an estimate of how well the factor model predicts the behavior of the given variable. The communality ranges from 0 to 1, where a value of 1 indicates full explanation of the variance. The residual matrix is obtained by subtracting the original correlation matrix from the one produced by the given factor model. This analysis indicates how well the factor model accounts for relationships between specific pairs of imbalances. To analyze the tumor population, the tumors were used as variables, and the imbalances were used as the observations; this will group tumors with similar sets of imbalances. To relate the individual tumors to karyotypic complexity, the NIPT for each individual tumor was used as an observation in addition to the imbalances (14). Because this parameter shows the greatest variation, the position of the tumors along the first principal component will be almost identical to karyotypic complexity. Classifying imbalances (14) were identified by testing for the presence of each imbalance in all members of a given cluster produced by the PCA. Communalities and the residual correlation matrix were not considered when evaluating the PCA of the tumors.

RESULTS

The temporal analysis (Fig. 1) revealed 1p−, +7, 7q−, and +12p to be early imbalances (TO = 1–5) in the whole colorectal tumor set. The observed TO value for +12p was lower than the TO value obtained from the simulations, *i.e.*, +12p occurred earlier than expected from the simulation. Intermediate imbalances were +5p, +8q, +17q, 17p−, +14q, +9, +13q, +1q, 5q−, and 6q−, with TO = 5–8, and late imbalances were, among others, 3p−, 9p− and 18q−, with TO $>$ 8. The observed TO values for the imbalances 9p−, 9q−, 17p−, 18q−, and –22 were significantly higher than those obtained from the simulation, and these imbalances thus occurred later than expected.

The frequencies of the imbalances in adenomas, carcinomas, and the total population of tumors are summarized in Table 1. The most frequent imbalances in adenomas were +7, +13q, 18q−, +20q, +8q, and 1p−. The majority of gains in the adenomas were caused by whole chromosome changes. The number of acquired imbalances correlated with the level of dysplasia ($P < 0.01$), and the late imbalance 18q− specifically correlated with severe dysplasia ($P < 0.01$). The presence of +13q, +20q, and 18q− correlated ($P < 0.01$) with large tumor size. The most frequent imbalances in carcinomas were +7, +8q, 1p−, 5q−, 8p−, 17p−, +13q, and 18q− (Table 1). Several of the imbalances that occurred at low frequencies (1%) in the adenomas (*e.g.*, +X, 6q−, 12q−, 13p−, −16, and −19) were found at moderate frequencies (10–20%) in the carcinomas. These imbalances occurred late (TO $>$ 6) in tumor progression, as shown by the temporal analysis. Eight imbalances (namely, +2, +12p, +16, 3q−, 6q−, 8p−, 11q−, and 13p−) correlated ($P < 0.05$) with poorly differentiated carcinomas. Of these, all were late (TO $>$ 6) except +12p (TO = 2.5).

Adenomas. Because the adenomas constitute the first step in the adenoma-carcinoma transition, the temporal order of the imbalances was established separately in this group. This analysis revealed +7, 1p−, and 8p− to be early imbalances (TO $<$ 3); +13q, +14q, +17q, −5q, 17p−, 18q−, and −22− to be intermediate imbalances; and +2, +3, +5, +6, +8, +9, +12p, +16, +19, +20, and 11q− to be late imbalances (Table 2). The first principal component in the PCA of the imbalances (Fig. 2*a*) showed a moderate correlation ($r = 0.62$) with the TO values and could thus be approximated as a time axis. The second component separated losses from gains and particularly 1p− from +7; these two imbalances showed significant negative correlation (Table 2). Gain of chromosome 7 clustered with the whole

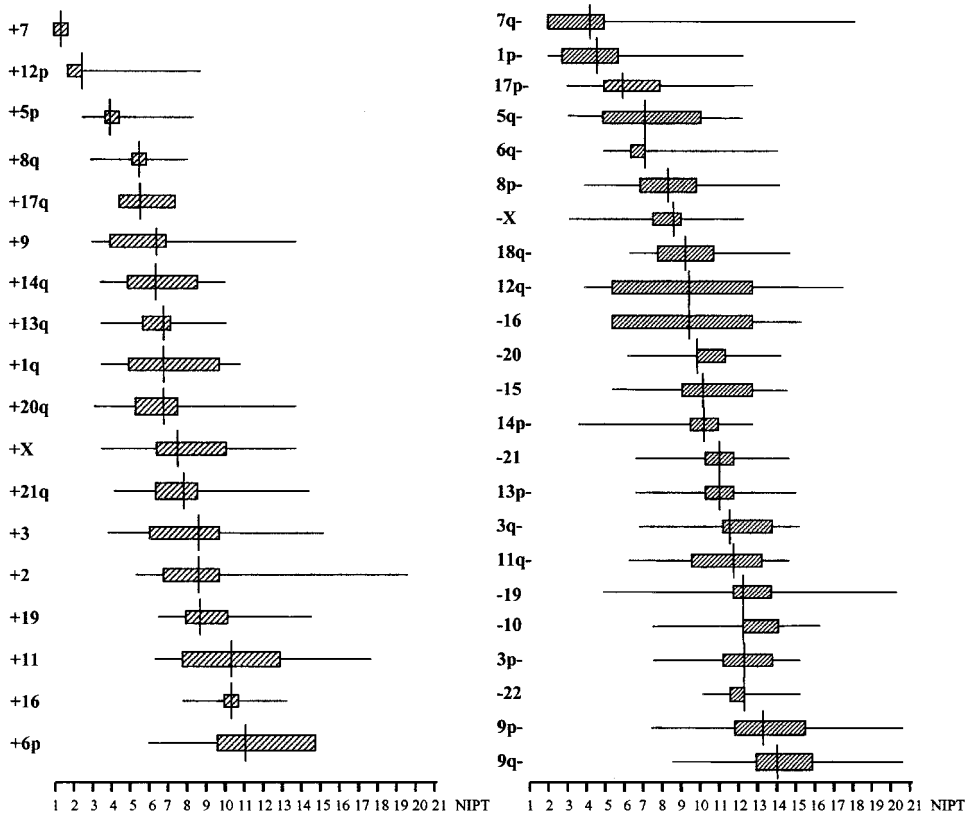

Fig. 1. Temporal analysis of imbalances (gains and losses are shown separately) in colorectal tumors. *Vertical lines in hatched boxes*, the mean of the bootstrapped TO; *hatched boxes*, boundaries of the 25 and 75 percentiles; *thin lines*, boundaries of the 2.5 and 97.5 percentiles. Imbalances were first sorted according to the TO values, and then imbalances with the same TO were organized with respect to the value of the 25 percentile. For abbreviations, see Table 1.

chromosome gains, whereas 1p− was located closer to the losses. The PCA of the adenoma tumor cases produced four clusters (Fig. 3*a*). By testing these for the presence or absence of individual imbalances, it was shown that +7 and 1p− were cluster-determining imbalances. Hence, all clusters could be defined by the presence or absence of these two imbalances. The cluster characterized by the simultaneous presence of 1p− and +7 was the smallest of the four (Fig. 3*a*). Thus, the data imply two cytogenetic pathways in adenomas: one initiated by +7 that is predominantly followed by whole chromosome gains; and one initiated with 1p− that is followed by losses, *i.e.*, one hyperdiploid and one pseudodiploid subtype of tumors. No significant differences between tumors containing 1p− and tumors containing +7 were seen regarding tumor location (proximal/distal colon), degree of dysplasia (minimal/moderate/severe), tumor size (< 1 cm >), or tumor type (tubulovillous, tubular, or hyperplastic).

Table 2 *Summary of the statistical characteristics of the imbalances present in adenomas*[a]

				Correlations[f]	
Imbalance[b]	TO[c]	1p−[d]	+7[e]	1p−	+7
+7	2.0	19	100	**−0.26**	1.00
1p−	2.0	100	7	1.00	**−0.26**
8p−	2.5	0	7	−0.13	0.01
+13q	4.0	29	45	−0.05	**0.22**
+17q	4.0	5	4	−0.01	−0.07
5q−	4.0	10	2	0.07	−0.15
17p−	4.0	19	9	0.12	−0.06
18q−	4.5	38	20	**0.23**	0.02
−22	4.5	14	2	**0.26**	−0.17
+14q	5.0	5	9	−0.08	−0.04
+20q	5.0	0	32	**−0.25**	**0.21**
+9	5.5	0	16	−0.17	0.12
+5p	6.0	10	11	0.00	0.05
+8q	6.0	5	16	−0.14	0.00
+3	6.5	0	13	−0.16	0.07
+12p	6.5	0	5	−0.11	0.01
+6p	7.0	5	11	−0.06	0.08
11q−	7.0	10	3	0.16	0.01
+2	7.5	5	11	−0.06	0.08
+16	8.0	5	14	−0.07	0.16
+19	8.0	0	9	−0.13	0.05

[a] Frequencies are given as percentages. The table only includes imbalances present in 5% or more of the adenomas.
[b] For abbreviations, see Table 1.
[c] Calculated as the median value for NIPT for all tumors with the given imbalance.
[d] Adenomas containing 1p−.
[e] Adenomas containing +7.
[f] Individual correlations with +7 and 1p−, respectively. Significant correlations, $P < 0.01$, are shown in bold.

Carcinomas. The PCA of the imbalances produced two major clusters, one consisting of gains and one consisting of losses (data not shown), indicating, as for adenomas, two major cytogenetic subtypes. Because no clear time axis was seen in this analysis, the TO values were calculated for the imbalances using data from carcinomas only (Table 3), and these values were used in the PCA as an observation. The first component in this analysis organized the imbalances according to TO, and the second component separated gains from losses (Fig. 2*b*). The early imbalances +7 and +12p were clearly separated from the remaining imbalances and were positioned closer to the gains; all gains had negative loadings on the second component. The intermediate imbalances formed two minor clusters: one included the losses 5q−, 6q−, 7q−, 8p−, 17p-, 18q−, and –21; and the other included the gains +5p, +8q, +13q, +17q, +20q, and +X. The late imbalances formed a tight cluster. Of the early changes identified in adenomas, +7 remained an early event in carcinomas as well (TO = 1.5), whereas 1p− predominantly occurred in complex karyotypes (TO = 7.5; Table 3).

Identification of Cytogenetic Pathways. To get an overall picture of the cytogenetic pathways, the adenoma and carcinoma data sets were combined. The PCA of the imbalances of the compiled set produced two major clusters of imbalances (Fig. 4*a*), representing the hyperdiploid and the pseudodiploid tumors, respectively. In the subsequent three-factor PCA, in which the TO values (Fig. 1) were

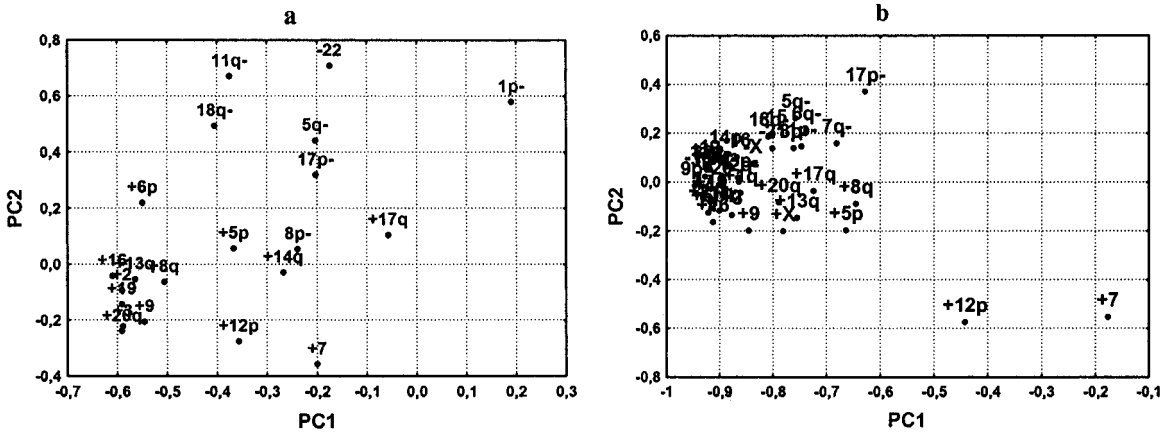

Fig. 2. PCA of the imbalances. The TO values were included as an observation when computing the principal components. *a*, adenomas; *b*, carcinomas.

included as an observation, the variables (imbalances) were well represented because 67% of the variance was explained, and the communalities ranged from 0.34 to 0.89. The first component in this analysis organized the imbalances according to TO, whereas the second component separated gains from losses (Fig. 4*b*). The third component separated the gains further in two cytogenetic pathways: one initiated by +12p, followed by +3, +5p, +9, +14q, and +X; and one initiated by +7, followed by +8q, +13q, +17q, and +20q (Fig. 4*c*). These pathways converged to a common set of late imbalances. By inspection of the correlation matrices, it could be seen that +7 showed significant positive correlation with +13 (+13q) and with +20 (+20q) in both the adenomas (Table 2) and the carcinomas (Table 3) and with +2, +3, +6p, +11, +16, −20, and −21 in the carcinomas. This finding suggests the presence of a +7 cytogenetic pathway including the intermediate imbalances +8q, +17q, +9, +13q, and +20q, in order of appearance, followed by the late imbalances +3, +6p, +11, +16, −20, and −21. Likewise, the PCA, the temporal analysis, and the individual correlations suggest a +12p karyotypic pathway including the imbalances +5p, +14q, +9, and +13q, in order of appearance, followed by +2, +3, +6p, and +16.

The PCA revealed at least two possible cytogenetic routes in the pseudodiploid group (Fig. 4*d*), one initiated by 1p− followed by 5q− and 8p−, and one initiated by 7q− followed by 5q− and 6q−, which converged to a common set of late imbalances. The correlation matrix showed that 1p− was significantly associated with 14 imbalances, of which 17p−, +13q, +1q, +20q, and 18q− were moderately late and appeared in the given order. Loss of 7q showed significant association with five imbalances, of which 8p− was moderately late, and the remainder were very late. The PCA also suggested a pathway initiated by 17p−, with 18q− as a secondary change. However, loss of 17p− differed from 1p− and 7q− because it did not behave as a typical early imbalance; few low-complexity karyotypes with 17p− were seen. This is shown in Fig. 5, in which the pooled NIPT distributions for 1p− and 7q− and the NIPT distribution for 17p− are plotted. Thus, two major cytogenetic pathways are present in the colorectal tumors, one dominated by gains and characterized by +7 and +12p, and one dominated by losses and characterized by 1p− and 7q−.

Two investigations were made to examine the relationship between the two pathways. First, a PCA was performed with the six latest gains and losses (+3, +2, +19, +11, +16, and +6p and –19, −10, 3p−, −22, 9p−, and 9q−). This analysis produced two clusters, one consisting of gains and one consisting of losses (data not shown), indicating that the two pathways remain separated throughout tumorigenesis. Second, the frequencies of the imbalances were calculated in tumors with 1p−/7q− and +7/+12p (Table 4). The obtained frequencies revealed that the pathways overlap, with 8p−, +13q, 17p−, 18q−, and +20q found in both subtypes.

To investigate possible differences in the adenoma-carcinoma transition between the two pathways, the +7/+12p and 1p−/7q− tumors were divided into adenomas and carcinomas, and the frequencies for the imbalances were recalculated. Significant differences in frequencies were identified by χ^2 analysis (Table 4). In the +7/+12p tumors, 9q−, +12p, and 17p− were seen at significantly higher frequencies in the carcinomas, whereas 11q− and 3q− were seen at significantly

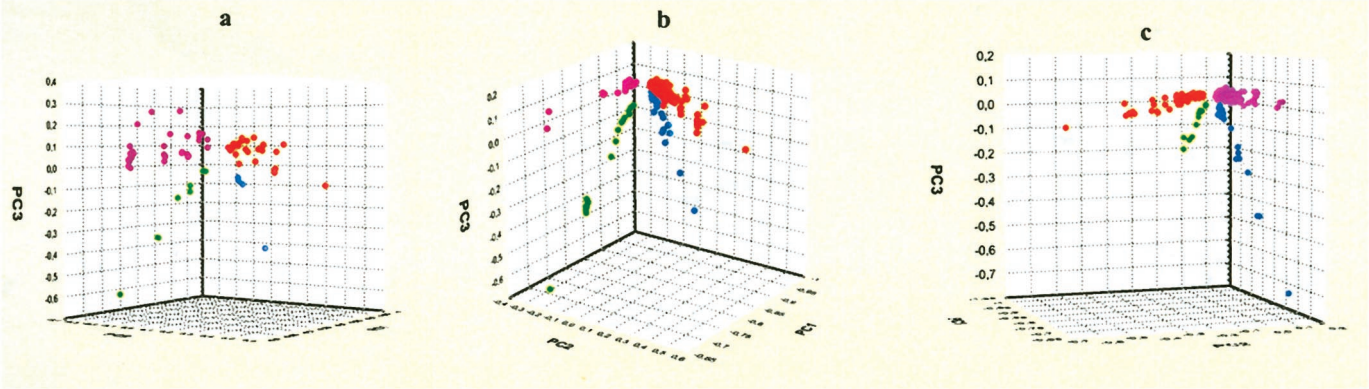

Fig. 3. PCA of colorectal tumor cases. *a*, adenomas. *Red*, tumors with +7 but not 1p−; *blue*, tumors with +7 and 1p−; *green*, tumors with 1p− but not +7, *violet*, tumors showing neither +7 nor 1p−. *b*, hyperdiploid carcinomas. *Green*, tumors with +12p but not +7; *blue*, tumors with +12p and +7; *red*, tumors with +7 but not +12p, *violet*, tumors lacking both +12p and +7. *c*, pseudodiploid carcinomas. *Blue*, tumors with 7q− but not 1p−; *green*, tumors with 7q− and 1p−; *red*, tumors with 1p− but not 7q−; *violet*, tumors lacking both imbalances.

Table 3 *Summary of the statistical characteristics of the imbalances present in carcinomas*

Imbalance[a]	TO[b]	Correlations[c]			
		+7	+12p	1p−	7q−
+7	1.5	1.00	0.14	−0.01	−0.04
+12p	3.0	0.14	1.00	−0.05	−0.08
7q−	4.5	−0.04	−0.08	0.14	1.00
+5p	5.0	0.03	0.12	0.04	−0.07
17p−	5.5	0.02	**−0.16**	**0.23**	0.10
+8q	6.0	0.09	0.06	0.05	0.05
+17q	6.0	0.02	−0.04	0.06	−0.00
6q−	6.5	−0.12	−0.04	0.13	0.03
5q−	7.5	−0.04	−0.09	0.12	0.11
+X	7.5	0.12	0.11	0.06	−0.05
1p−	7.5	−0.01	−0.05	1.00	0.14
−21	7.5	**0.14**	−0.03	**0.17**	0.22
8p−	8.0	0.07	−0.03	0.13	**0.16**
+9	8.0	**0.17**	0.09	−0.05	0.02
+13q	8.0	**0.20**	**0.14**	**0.17**	0.04
−15	8.0	0.01	−0.13	**0.18**	0.04
−X	8.0	0.01	−0.09	0.09	0.12
+1q	8.5	0.03	−0.05	**0.20**	0.02
+20q	8.5	**0.17**	0.08	**0.16**	0.02
−10	8.5	0.11	−0.01	**0.23**	0.17
+14q	9.0	0.06	0.09	−0.05	0.03
−16	9.0	−0.03	−0.06	0.09	**0.14**
+3	9.5	**0.15**	0.05	0.05	0.06
12q−	9.5	0.10	0.02	0.08	0.10
18q−	9.5	0.07	−0.06	**0.31**	0.09
+21q	10.0	0.07	0.08	−0.02	−0.07
+19	10.5	0.13	0.04	0.03	−0.00
13p−	10.5	0.05	0.04	**0.23**	0.06
−20	10.5	**0.15**	0.02	0.11	−0.00
+11	11.0	**0.15**	0.09	0.05	0.06
+16	11.0	**0.25**	**0.18**	0.02	−0.09
14p−	11.0	0.03	−0.07	**0.22**	0.08
3q−	11.5	−0.05	−0.04	**0.16**	**0.15**
+2	12.0	**0.17**	0.05	0.04	0.09
11q−	12.0	−0.09	−0.08	0.11	0.02
−19	12.0	0.05	−0.06	0.14	**0.18**
+6p	12.5	**0.19**	**0.15**	0.11	0.06
3p−	12.5	0.01	−0.03	**0.18**	0.12
−22	12.5	0.10	0.01	**0.21**	0.13
9p−	14.0	0.08	0.03	0.12	**0.15**
9q−	15.0	0.12	0.00	**0.19**	0.12

[a] For abbreviations, see Table 1.
[b] Based on carcinomas only.
[c] Individual correlations with +7, +12p, 1p−, and 7q−, respectively. Significant correlations, $P < 0.01$, are shown in bold.

lower frequencies. In the 1p−/7q− tumors, 7q−, 8p−, −10, 13p−, 17p−, and +20q were seen in significantly higher frequencies in the carcinomas. Also, +1q, 5q−, 9p−, and 18q− were seen at higher frequencies in carcinomas, albeit at borderline significance.

Analysis of the Tumor Population. To further evaluate the underlying structure of the two tumor populations, the tumor cases were subjected to PCA. To approximate the hyperdiploid subtype, tumors containing any of the +7, +9, +12p, and +16 imbalances were selected. These imbalances were all seen with at least 2× higher frequency in the +7/+12p tumors compared with the 1p−/7q− tumors (Table 4). A similar subset approximating the pseudodiploid tumors was constructed by selecting tumors containing any of the 1p−, 3p−, 3q−, 6q−, 7q−, 9p−, 9q−, −10, 14p−, −15, −16, −19, −22, and –X imbalances (Table 4). The PCA of the hyperdiploid tumors revealed four well-separated clusters determined by the presence or absence of +12p and +7 (Fig. 3*b*). The tumors could thus be classified as (*a*) +12p, (*b*) +12p and +7, (*c*) +7, and (*d*) tumors showing neither +12p nor +7. The PCA of the pseudodiploid tumors revealed four clusters with the classifying imbalances 1p− and 7q− (Fig. 3*c*).

DISCUSSION

Several lines of evidence have suggested that colorectal tumors develop through different genetic pathways. Based on flow cytometric analysis of DNA content, approximately 60% of colorectal cancers are classified as aneuploid, and the remaining 40% are classified as diploid. Furthermore, genetic instability in colorectal cancer depends on two separate mechanisms, CIN and MIN (4). In the present study, we used the compiled cytogenetic data from colorectal tumors registered in the Mitelman Database of Chromosome Aberrations in Cancer[4] to investigate the karyotypic profile of this tumor type. After transformation of the individual tumor karyotypes to chromosomal imbalances, several statistical methods were applied to the data to determine the temporal order of the occurrence of these changes and to distinguish genetic subsets characterized be specific gains and/or losses. A major characteristic of the analyzed tumors was the presence of two distinct cytogenetic subtypes, one dominated by gains (the hyperdiploid group), and one dominated by losses (the pseudodiploid group). The finding that the early changes in the two subtypes, 1p− and +7, showed significant negative correlation indicates that these two imbalances either are noncompatible or occur in two different cell types. Nevertheless, when the initiating changes have taken place, the nature of the following steps is restricted; one path results in the accumulation of gains, and the other results in the accumulation of losses. The separation in the two subtypes was also seen when the PCA was limited to the very late imbalances, in which the separation of gains and losses remained. Hence, our data suggest that colorectal tumorigenesis occurs through one of these two major pathways throughout the karyotypic evolution and remains delineated even at the late tumor stages, although some overlap between the karyotypic profiles occurs.

The temporal analysis revealed a clear chronological sequence of the chromosomal imbalances. We have previously shown that this analysis correlates well with tumor progression as determined by histopathological criteria (16, 18). In the present study, late imbalances such as +2, 3q−, 6q−, 8p−, 11q−, 13p−, +16, and 18q− showed a significant positive correlation with severe dysplasia and poor tumor differentiation. Furthermore, tumor size correlated with karyotypic complexity. Thus, the deduced order of events most likely represents a good approximation of events taking place *in vivo*. The simulation showed +12p to occur earlier than expected from a random combination of imbalances, and +12p may thus be of particular importance for the early stages of tumor formation, whereas the selective advantage of 9p−, 9q−, −10, 17p−, 18q−, and −22, which occurred later than expected from the simulation, may be dependent on preceding imbalances. Although discrete TO values for the individual imbalances were obtained, the bootstrapped confidence intervals revealed imbalances with extensive overlap. Thus, the obtained temporal sequence of changes should not be treated as a fixed order of events but rather as a biologically favored one.

To screen for possible cytogenetic pathways, we reasoned that imbalances frequently present in the same cases would belong to the same pathway, so that the pattern of correlations among imbalances would reveal possible cytogenetic pathways. To condense the large correlation matrix produced by the 41 variables (imbalances) and extract the central features of the matrix, we performed PCA. We have previously shown the adequacy of this method by comparing it with correspondence analysis, a method similar to PCA but based on normalized weighted χ^2 distances (19). Results obtained with this method conform completely to results obtained by PCA (16).

Two cytogenetic pathways, characterized by the early changes +7 and +12p, were identified in the hyperdiploid tumors. The secondary, intermediate imbalances differed between the groups, with +8q +9, +17q, 17p−, and +20q common in the +7 group, and +2, +5p, +9, and +14q frequently seen together with +12p (Fig. 6). However, the pathways converged at later stages with +3, +13q, and 18q−, among others, constituting late changes found in both tumor subsets (Fig. 6).

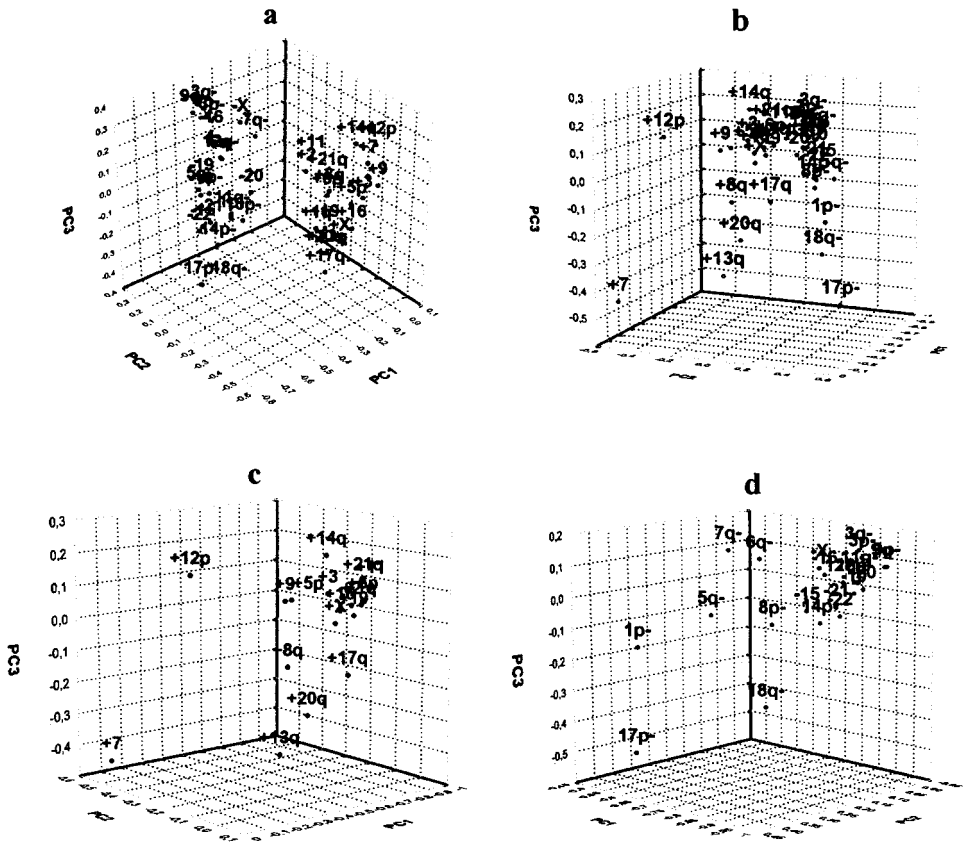

Fig. 4. PCA of the imbalances in colorectal tumors. *a*, PCA of the imbalances showing two clusters of imbalances, one dominated by gains ,and one dominated by losses. *b*, PCA of the imbalances including the TO values as observations when calculating the principal components. *c*, to make the relationships among the gains more clear, the losses have been deleted from Fig. 3*b*, and the resulting representation was rotated so that the time axis goes from left to right. *d*, to make the relationships among the losses more clear, the gains have been deleted and Fig. 3*b*, and the representation was rotated so that the time axis goes from left to right.

The proportion of tumors classified as carcinomas was significantly higher ($P = 0.013$) for +12p tumors (36 of 44) than for +7 tumors (57 of 175), and +12p was the only early imbalance that correlated with poor tumor differentiation. This suggests that of the two hyperdiploid pathways, the +12p route may be the more aggressive one.

At least two cytogenetic pathways were also identified in the pseudodiploid group of tumors. The major pathway was initiated by 1p−, which was followed by the moderate late imbalances +13q, +20q, 17p−, 5q−, +1q, 8p−, and 18q−, appearing in the given order, and then followed by even later imbalances (Fig. 6). Results from the PCA of the imbalances, the temporal analysis, and the individual correlations revealed 7q− to be an early change followed by 5q−, 6q−, 8p−, and –16, as moderately late imbalances, and then by imbalances in common with the 1p− pathway. The PCA also indicated a third route of karyotypic evolution, which was initiated by 17p− with 18q− as an important secondary event. However, the nature of this pathway is unclear because the proportion of low-complexity karyotypes was very low, and, furthermore, 17p− was not identified as a classifying imbalance. On the other hand, it may suggest that this subtype acquires an increased CIN at early stages and thus progresses more rapidly to highly complex tumors. Several of the chromosomal regions found to be frequent secondary targets for loss correspond to loci for tumor suppressor genes known to be mutated in colorectal cancer, *e.g.*, *APC* in 5q21-2 and *TP53* in 17p12.

The two major pathways showed significant differences with regard to the adenoma-carcinoma transition. In the +7/+12p tumors, the only late imbalance that occurred at a significantly higher frequency in the carcinomas and may thus be linked to the carcinoma develop-

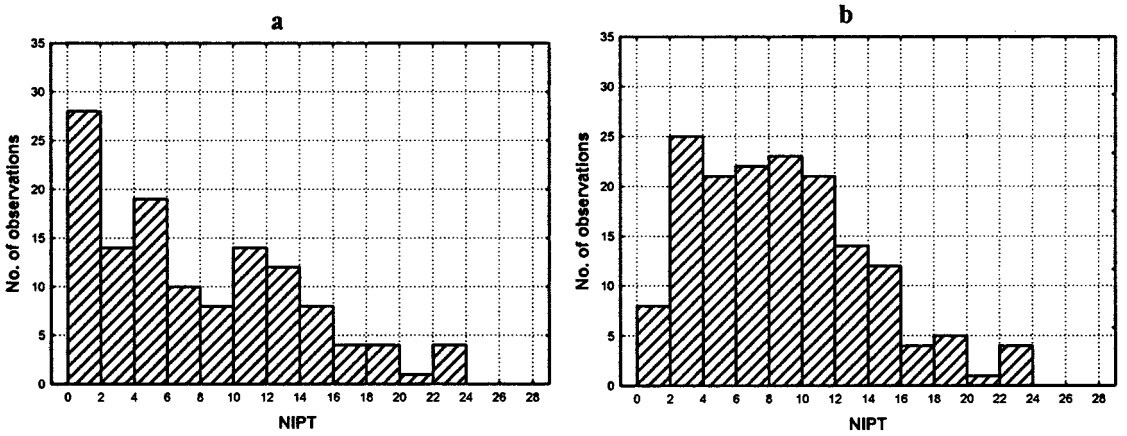

Fig. 5. The NIPT distribution of (*a*) the pooled 1p− and 7q− cases and (*b*) the 17p− cases.

Table 4 *The frequency of the imbalances in +7/+12p and 1p−/7q− containing tumors*

Imbalances[a]	+7/+12p (N = 218)	1p−/7q− (N = 126)	+7/+12p adenoma	+7/+12p carcinoma	χ^2 test[b]	1p−/7q− adenoma	1p−/7q− carcinoma	χ^2 test[b]
+1q	6	12	11	9		0	16	0.043
+2	13	13	22	14		5	12	
+3	13	10	24	13		0	10	
+5p	11	10	16	12		9	10	
+6p	14	11	26	15		5	13	
+7	90	29	96	87		18	34	
+8q	23	18	22	26		5	22	
+9	15	6	23	14		0	6	
+11	11	9	18	13		0	10	
+12p	22	6	13	28	0.014	0	8	
+13q	39	37	51	36		27	38	
+14q	7	6	16	7		9	5	
+16	15	7	26	17		5	8	
+17q	8	11	11	11		5	13	
+19	10	8	20	10		0	9	
+20q	26	25	38	26		0	29	0.004
+21q	6	7	13	7		0	3	
+X	13	10	14	18		0	15	
1p−	14	81	13	18		95	8	
3p−	6	17	17	9		5	19	
3q−	4	14	17	5	0.004	5	15	
5q−	14	28	12	18		9	30	0.044
6q−	6	17	12	9		5	22	
7q−	4	29	6	6		9	33	0.028
8p−	21	34	18	29		0	40	0.004
9p−	7	15	20	10		0	16	0.043
9q−	8	17	22	10	0.025	5	17	
−10	13	27	22	16		5	28	0.020
11q−	5	13	20	4	0.000	9	13	
12q−	11	16	14	16		0	15	
13p−	11	21	16	16		5	27	0.025
14p−	14	29	18	18		9	31	0.035
−15	11	25	18	15		9	26	
−16	7	17	14	9		5	16	
17p−	26	52	15	35	0.004	18	57	0.001
18q−	31	56	31	36		36	60	0.042
−19	8	18	19	11		5	20	
−20	12	13	19	16		0	15	
−21	14	25	16	20		9	29	
−22	15	31	19	20		18	30	
−X	7	16	14	9		5	15	

[a] For abbreviations, see Table 1.
[b] The χ^2-test was used to investigate whether the frequency of the given imbalance differed between adenomas and carcinomas, respectively. Only Ps < 0.05 are given.

ment was 17p−. In the 1p−/7q− pathway, at least six late imbalances (8p−, −10, 13p−, 14p−, 17p−, and +20q) occurred at significantly higher frequencies in the carcinomas. Furthermore, in contrast to +7, 1p− showed a low TO value in the adenomas and a high value in the carcinomas, and hence 1p− is present in low-complexity karyotypes in the adenomas but in highly complex karyotypes in the carcinomas. This indicates that the transition from adenoma to carcinoma is more closely linked to karyotypic evolution in the 1p−/7q− pathway than in the +7/+12p pathway. Particularly intriguing was the behavior of 8p− and +20q. Loss of 8p and gain of 20q showed high frequencies in both +7/+12p adenomas and carcinomas, but the frequencies increased dramatically in 1p−/7q− carcinomas compared with adenomas of the same cytogenetic subtype. This indicates that these two imbalances may have different genetic outcomes in the two different cytogenetic pathways. A similar difference was not observed for 17p−, which was seen at significantly higher frequencies in carcinomas of both subtypes.

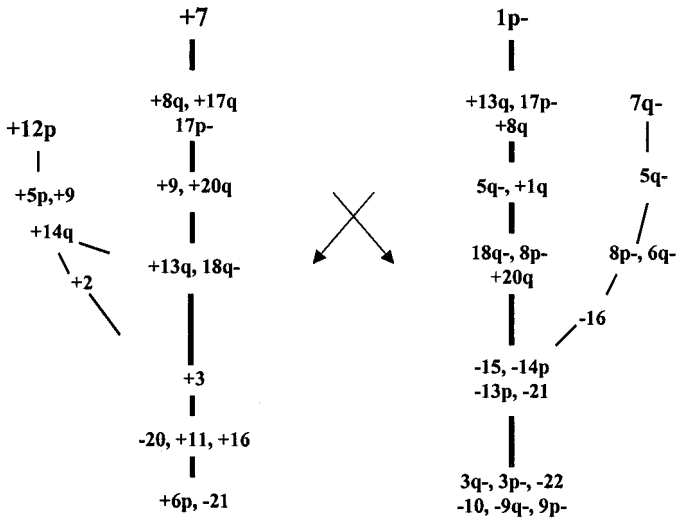

Fig. 6. A summary of the karyotypic pathways in colorectal tumors. The pathway scheme was compiled from the PCA, the temporal analysis, the individual correlations between imbalances, and the frequencies given in Tables 2 and 4. The putative 17p− pathway is not included.

In conclusion, the accumulated cytogenetic data on colorectal adenomas and carcinomas reveal a complex pattern of chromosomal alterations and show that specific imbalances occur in different contexts and combinations. Despite this, we were able to establish a temporal order of the karyotypic events that was consistent with the available histopathological data. Furthermore, two major cytogenetic subtypes were revealed, one hyperdiploid and one pseudodiploid. To disclose potential karyotypic pathways, we performed several statistical analyses, which resulted in the identification of two major pathways. Even though we were able to describe these in some detail, they merely represent a schematic outline of all steps taking place. To be more precise, the disparate data imply that complex karyotypes may not easily be reduced to the presence or absence of specific

imbalances but rather seem to be an expression of intricate relations among tumor-promoting changes. In this sense, the present investigation may only have touched upon the complex biological processes resulting in cellular transformation.

REFERENCES

1. Pisani, P., Parkin, M. D., Bray, F., and Ferlay, J. Estimates of the worldwide mortality from 25 cancers in 1990. Int. J. Cancer, *83:* 18–29, 1999.
2. Houlston, R. S., and Tomlinson, I. P. M. Genetic prognostic markers in colorectal cancer. J. Clin. Pathol. Mol. Pathol., *50:* 281–288, 1997.
3. Fearon, E., and Vogelstein, B. A genetic model for colorectal tumorigenesis. Cell, *61:* 759–767, 1990.
4. Lengauer, C., Kinzler, K. W., and Vogelstein, B. Genetic instability in colorectal cancers. Nature (Lond.), *386:* 623–627, 1997.
5. Thibodeau, S. N., Bren, G., and Dchaid, D. Microsatellite instability in cancer of the proximal colon. Science (Wash. DC), *260:* 816–819, 1993.
6. Bardi, G., Sukhikh, T., Pandis, N., Fenger, C., Kronborg, O., and Heim, S. Karyotypic characterization of colorectal adenocarcinomas. Genes Chromosomes Cancer, *12:* 97–109, 1995.
7. Heim, S., and Mitelman, F. Cancer Cytogenetics. New York: Wiley-Liss, 1995.
8. Ried, T., Knutzen, R., Steinbeck, R., Blegen, H., Schrock, E., Heselmayer, K., du Manoir, S., and Auer, G. Comparative genomic hybridization reveals a specific pattern of gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer, *15:* 234–245, 1996.
9. Meijer, G. A., Hermsen, M. A., Baak, J. P., van Diest, P. J., Meuwissen, S. G., Belien, J. A., Hoovers, J. M., Joenje, H., Snijders, P. J., and Walboomers, J. M. Progression from adenoma to carcinoma is associated with non-random chromosomal gains as detected by comparative genomic hybridisation. J. Clin. Pathol., *51:* 901–909, 1998.
10. DeAngelis, P. M., Calusen, O. P. F., Schjolberg, A., and Stokke, T. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour ploidy, genotypes and phenotypes. Br. J. Cancer, *80:* 526–535, 1999.
11. Aragane, H., Sakamura, C., Nakanishi, M., Yasuoka, R., Fujita, Y., Tanguchi, H., Hagiwara, A., Yamagushi, T., Abe, T., Inazawa, J., and Yamagishi, H. Chromosomal aberrations in colorectal cancers and liver metastases analyzed by comparative genomic hybridization. Int. J. Cancer, *94:* 623–629, 2001.
12. Zhou. W., Goodman, S. N., Galizia, G., Lieto, E., Ferraraccio, F., Pignatelli, C., Purdie, C. A., Piris, J., Morris, R., Harrison, D. J, Pay, P. B., Cuilford, A., Romans, K. E., Montgomery, E. A., Choti, M. A., Kinzler, K. W., and Vogelstein, B. Counting alleles to predict recurrence of early-stage colorectal cancers. Lancet, *359:* 219–225, 2002.
13. Höglund, M., Gisselsson, D., Mandahl, N., Johansson, B., Merterns, F., Mitelman, F., and Säll, T. Multivariate analyses of genomic imbalances in solid tumors reveal distinct and converging pathways of karyotypic evolution. Genes Chromosomes Cancer, *31:* 156–171, 2001.
14. Höglund, M., Gisselsson, D., Säll, T., and Mitelman, F. Coping with complexity: multivariate analysis of tumor karyotypes. Cancer Genet. Cytogenet., *135:* 103–109, 2002.
15. Mooney, C. Z., and Duval, R. D. Bootstrapping: a Nonparametric Approach to Statistical Inference. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-095. Thousand Oaks, CA: Sage, 1993.
16. Höglund, M., Gisselsson, D., Hansen, G. B., Säll, T., and Mitelman, F. Multivariate analysis of chromosomal imbalances in breast cancer delineates cytogenetic pathways and reveals complex relationships among imbalances. Cancer Res., *62:* 2675–2680, 2002.
17. Affifi, A., and Azen, S. (eds.). Statistical Analysis: A Computor Oriented Approach, pp. 318–324. New York: Academic Press, 1979.
18. Höglund, M., Säll, T., Heim, S., Mitelman, F., Mandahl, N., and Fadl-Elmula, I. Identification of cytogenetic subgroups and karyotypic pathways in transitional cell carcinoma. Cancer Res., *61:* 8241–8246, 2001.
19. Clausen, S. Applied Correspondence Analysis: An Introduction. Sage University Papers Series on Quantitative Applications in the Social Sciences, Series No. 07-121. Thousand Oaks, CA: Sage, 1988.

Dissecting Karyotypic Patterns in Colorectal Tumors: Two Distinct but Overlapping Pathways in the Adenoma-Carcinoma Transition

Mattias Höglund, David Gisselsson, Gunnar B. Hansen, et al.

Cancer Res 2002;62:5939-5946.