



Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas

Graeme Hodgson¹, Jeffrey H. Hager³, Stas Volik¹, Sujatmi Hariono¹, Meredith Wernick¹, Dan Moore¹, Donna G. Albertson^{1,2}, Daniel Pinkel^{1,2}, Colin Collins^{1,2}, Douglas Hanahan³ & Joe W. Gray^{1,2}

Published online: 5 November 2001, DOI: 10.1038/ng771

Carcinomas that develop in the pancreatic islets of transgenic mice expressing the SV40 T-antigen (Tag) under transcriptional control of the rat insulin II promoter (RIP) progress through well-characterized stages that are similar to aspects of human tumor progression, including hyperplastic growth, increased angiogenesis and reduced apoptosis¹. The latter two stages have been associated with recurrent loss of heterozygosity (LOH)² and reduced genome copy number³ on chromosomes 9 (LOH9) and 16 (LOH16), aberrations which we believe contribute to these phenotypes. Earlier analyses localized LOH9 to approximately 3 Mb and LOH16 to approximately 30 Mb (both syntenic with human 3q21–q25) but were limited by low throughput and a lack of informative polymorphic markers. Here we show that comparative genomic hybridization to DNA microarrays (array CGH)^{4–7} overcomes these limitations by allowing efficient, genome-wide analyses of relative genome copy number. The CGH arrays used in these experiments carried BACs distributed at 2–20-MB intervals across the mouse genome and at higher density in regions of interest. Using array CGH, we further narrowed the loci for LOH9 and LOH16 and defined new or previously unappreciated recurrent regions of copy-number decrease on chromosomes 6, 8 and 14 (syntenic with human chromosomes 12p11–p13, 16q24.3 and 13q11–q32, respectively) and regions of copy-number increase on chromosomes 2 and 4 (syntenic to human chromosomes

20q13.2 and 1p32–p36, respectively). Our analyses of human genome sequences syntenic to these regions suggest that *CYP24*, *PFDNA4*, *STMN1*, *CDKN1B*, *PPP2R3* and *FSTL1* are candidate oncogenes or tumor-suppressor genes. We also show that irradiation and genetic background influence the spectrum of aberrations present in these tumors.

We produced arrays for these studies by printing 5' amine-linked degenerate oligonucleotide primer PCR-amplified (DOP-PCR)⁸ BAC DNA onto amine-reactive slides using the clones summarized in Table 1. We carried out array CGH by hybridizing CY3- and CY5-labeled test and reference DNA samples to the arrays along with unlabeled murine Cot-1 DNA, to suppress hybridization of interspersed repeat sequences. We then assessed ratios of CY3:CY5 fluorescence intensity (hereafter called CGH ratios) as previously described⁴. We assessed CGH ratio variability in normal-versus-normal as well as in tumor-versus-normal hybridizations by fitting Gaussian distributions to the histograms of the CGH ratios for each analysis (Fig. 1). We used the standard deviation (s.d.) of the central Gaussian distribution to estimate normal measurement variation within each CGH analysis. The s.d. varied slightly between CGH analyses (0.08±0.03; mean±1 s.d. for the 85 analyses included in this study). We scored a region as having either significantly increased or decreased copy number in the tumor if its CGH ratio was ±3 s.d. higher or lower than the mean of the central distribution.

We carried out a CGH analysis of DNA from a normal DBA/2 (D2) × C57Bl/6 (B6) F1 female and a normal C3He (C3) × B6 F1 female (Fig. 1a). The variation in CGH ratio is 0.04, with one clone on chromosome 7 showing a significantly different CGH ratio. We obtained similar results when the test and reference genome labels were reversed. The ratio for this clone was normal, however, when we compared DNA from a D2B6F1 female with itself. We attribute this variant ratio to genome sequence differences between the mouse strains. We noted and discounted these variants while interpreting CGH profiles of tumor genomes. We also carried

Table 1 • Summary of the genomic clones included in the customized mouse array

Genomic locus	Clone type	Resolution	Library
chromosome 9 (55–56 cM) ^a	BAC	1 clone/0.5 Mb	RPCI-23/CITB-CJ-7B
chromosome 16 (14–34 cM) ^a	BAC	1 clone/1 Mb	RPCI-23/CITB-CJ-7B
chromosome 4 ^b	BAC	1 clone/2 Mb	RPCI-23
chromosome 8 (64–65cM) ^c	BAC	1 clone/200 kb	RPCI-23
chromosome 2 (99–100cM) ^c	BAC	1 clone/200 kb	RPCI-23/CITB-CJ-7B
genome-wide ^d	BAC	~1 clone/20 Mb	CITB-CJ-7B
centromere and telomere of each mouse chromosome ^e	P1	2 clones per chromosome	ref. 23

^aBACs were obtained by screening the RPCI-23 and CITB-CJ-7B mouse BAC libraries by overgo hybridization²³. We derived murine-specific probes from genetically and radiation hybrid-mapped markers and genes known to map within LOH9/16 and the respective syntenic region in humans. We used PCR primers established from the marker sequence to confirm all positive hybridization signals; most clones were FISH-mapped to confirm correct chromosomal localization. Clones not mapped by FISH were usually part of a contig containing a clone that had been previously FISH-mapped. ^bObtained from the mouse BAC-PAC Resource (<http://www.chori.org/bacpac>). ^cClones were identified by overgo hybridization¹¹ using mapped markers as probes, confirmed by PCR, and FISH-mapped. ^dClones were picked from the CITB-CJ-7B library based on previously published addresses²⁵. ^eObtained from a previously established set of clones²⁴.

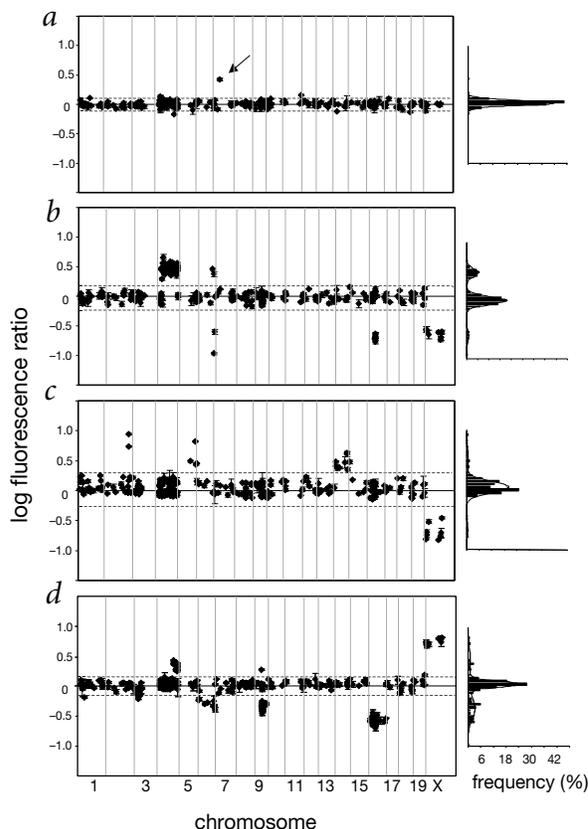
¹Cancer Genetics and Breast Oncology Programs, UCSF Cancer Center, University of California at San Francisco, Box 0808, San Francisco, California 94143-0808, USA.

²Department of Laboratory Medicine, University of California at San Francisco, Box 0808, San Francisco, California 94143-0808, USA.

³Department of Biochemistry and Biophysics, UCSF Diabetes and Comprehensive Cancer Centers, University of California at San Francisco, San Francisco, California 94143-0534, USA. Correspondence should be addressed to J.W.G. (e-mail: jgray@cc.ucsf.edu).



Fig. 1 CGH analyses of genome copy number in murine tissues, tumors and cell lines. The CGH ratio for each BAC array element is plotted as a function of its genome location, with chromosome 1 to the left and X to the right; for each chromosome the order is left to right, centromeric to telomeric. Each ratio is the log mean of the quadruplicate array measurements; the error bars show 1 s.d. Vertical lines indicate chromosome boundaries. Within each chromosome, clones associated with a known marker are ordered based on the consensus map position established by the Y2000 mouse chromosome committee reports (<http://www.informatics.jax.org/ccr/>). Clones that were mapped by fluorescence *in situ* hybridization (FISH) to a specific cytogenetic band but not associated with a known marker were placed on the consensus map to the nearest approximate location by integrating information from the cytogenetic and consensus maps (<http://www.informatics.jax.org/>). Histograms of each set of array CGH ratios are shown to the right of each array CGH analysis. Solid lines superimposed on each histogram show the results of the least-squares/maximum likelihood fit of the sum of three Gaussian distributions. The horizontal dashed lines in each array CGH analysis are drawn 3 s.d. above and below the mean of the central distribution. **a**, Array CGH analysis of normal D2B6F1 versus C3B6F1 DNA samples. The increased fluorescence ratio for the highlighted chromosome 7 clone (arrow) is probably caused by a genomic sequence difference between the C3H and DBA/2 genomes. The fluorescence ratio for this clone significantly decreased upon reverse labeling of these two genomes, and seemed within normal limits when compared with DNA from the same mouse background. **b**, Array CGH analysis of a RIP-Tag tumor-derived cell line H5#1 and normal D2B6F1 DNA. H5#1 has a *Mus musculus castaneus* × B6 F1 genotype and is known to have LOH16 from SSLP analysis. A cluster of clones mapping to LOH16 shows a clear reduction in CGH ratio. Additional single-copy number changes on chromosomes 4 and 6 can be seen in this cell line. Because the test and reference genomes were male and female, respectively, the mean \log^2 CGH ratio of the X-chromosome clones was reduced to -0.62 . This compares favorably to similar sex-mismatched hybridizations to arrays spotted with whole BACs showing a \log_2 converted mean CGH ratio of -0.53 (ref. 4). **c**, Array CGH analysis of DNA from a D2B6F1 RIP-Tag tumor versus normal DNA from a D2B6F1 mouse. Significant gains on chromosomes 2, 5 and 14 are apparent, as is the sex mismatch involving the X chromosome. The segmental increase in copy number for chromosome 2 involved two overlapping BACs harboring *CYP24*, *PFDN4* and *BCAS1* (ref. 11). **d**, Array CGH analysis of DNA from a D2B6F1 RIP-Tag tumor versus normal DNA from a D2B6F1 mouse. The reduced CGH ratio of all chromosome 16 clones indicates the complete loss of one copy of chromosome 16. The X chromosome clones show an increase in CGH ratio due to the sex mismatch of the two genomes. Gains of a portion of chromosome 4 and losses of portions of chromosomes 6 and 9 also are seen in this tumor.



out an array CGH analysis of a normal female mouse and a male RIP-Tag tumor-derived cell line (H5#1) known to carry LOH16 (Fig. 1b). The CGH and SSLP analyses were completely concordant over this region (data not shown). Other single-copy changes detected in this analysis include a gain of chromosome 4, a gain and loss towards the distal end of chromosome 6, and a 'loss' of the X chromosome (because the compared genomes were sex-mismatched). We also detected single-copy changes in partially purified RIP-Tag tumor cells, including: (i) genome copy number gains

on chromosomes 2, 5 and 14 and the sex mismatch on the X-chromosome (Fig. 1c) and (ii) genome copy number gains on chromosome 4, losses on chromosomes 9 and 16 and the sex-chromosome mismatch (Fig. 1d). The decreases in CGH ratios for chromosomes 6 and 9 were less than those for chromosome 16. We speculate that most of the cells in this tumor had lost chromosome 16, whereas only a fraction had lost chromosomes 6 and 9. This is consistent with earlier observations that LOH16 typically appears at an earlier stage of development than LOH9 in RIP-Tag tumors².

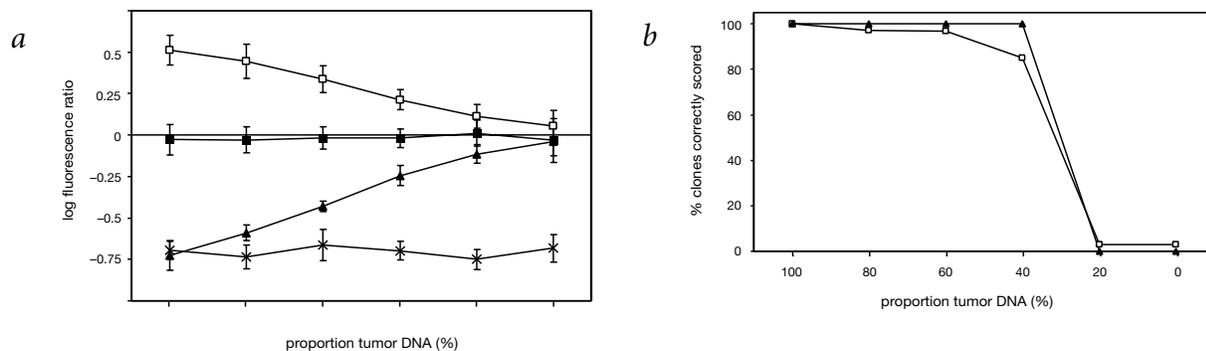
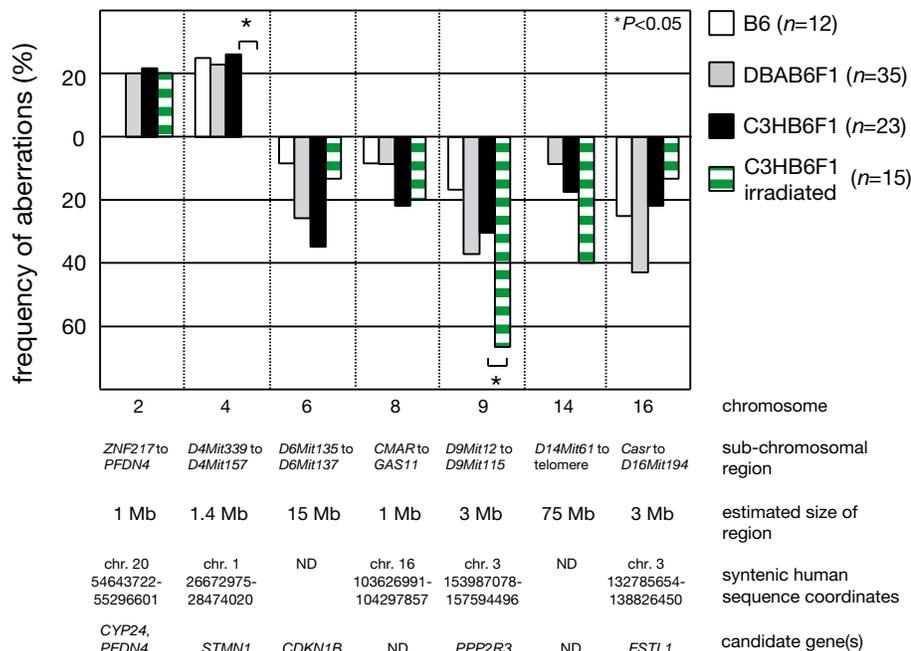


Fig. 2 Influence of normal genomic DNA contamination on detecting single copy-number abnormalities in tumors. The test genomes in this set of experiments are derived from male tumor-cell line DNA (genomic profile shown in Fig. 1b) mixed with normal D2B6F1 male DNA prior to labeling with CY3. The percentage of the tumor genome in the mix varies from 100–0, as shown on the x axis. The reference genome is normal D2B6F1 female DNA. **a**, The mean \log_2 fluorescence ratio (± 1 s.d.) is shown for clones mapping to regions present in this cell line at normal genome copy number (black squares), gained on chromosome 4 (white squares) or deleted on chromosome 16 (black triangles). Clones mapping to the X chromosome are also plotted (black Xs). **b**, Percentage of clones mapping to the chromosome 4 gain (white squares) and chromosome 6 loss (black triangles) that were scored correctly is plotted on the y axis. The unchanged, gained or deleted status of a clone in this cell line was determined as described (Methods) and as illustrated in Fig. 1. The copy-number status of most clones was correctly determined with as little as 40% tumor DNA. At 20% tumor DNA, most clones were erroneously scored as unchanged in copy number.





Fig. 3 Frequency of occurrence of regions of aberration observed in 85 RIP-Tag tumors. Only aberrations occurring in greater than 20% of tumors in at least one genetic background are shown. Aberrations occurring in the minimally defined regions indicated at the bottom of the figure were used to calculate frequencies. The sizes of these regions are indicated. Asterisk indicates significant P values less than 0.05: chromosome 4, $P=0.031$; chromosome 9, $P=0.0278$. Note that the arrays used for analysis of all B6 tumors and 6 of 35 D2B6F1 tumors did not contain BAC elements for the region of recurrent copy-number increase on chromosome 2. This precluded analysis of chromosome 2 gains in B6 tumors. The frequency of chromosome 2 gains in D2B6F1 tumors has been adjusted to account for the missing data. Flanking markers and syntenic human sequence coordinates are based on the human genome assembly at <http://genome.ucsc.edu> (October 2000 freeze); candidate genes are listed for each minimally defined region in the lower part of the figure.



We assessed the influence of normal or unaffected cells on our ability to detect single-copy number changes by mixing normal genomic DNA with DNA from cell line H5#1 prior to labeling. We detected single-copy gains and losses reliably in samples in which as little as 40% of the DNA was from the tumor (Fig. 2). This sensitivity allows detection of copy-number abnormalities

(CNAs) in tumors containing substantial stromal components that cannot be removed, as well as detection of CNAs that are present in only a fraction of the cells in the tumor.

We observed recurrent changes in 70 Rip-Tag tumors (Fig. 3). These tumors developed in B6 (12 tumors), D2B6F1 (35 tumors) or C3B6F1 (23 tumors) mice. Recurrent deletions occurred on chromosomes 6, 8, 9, 14 and 16 and recurrent copy-number increases occurred on chromosomes 2 and 4. Significantly, we narrowed LOH16 to approximately 3 Mb (26 cR) between *Casr* and *D16Mit194* by identifying two narrow, overlapping interstitial deletions (Fig. 4). Similarly, we localized the gain on chromosome 4 to approximately 1.5 Mb between markers *D4Mit339* and *D4Mit157*. We localized the region of gain on chromosome 2 to less than 1 Mb by identifying a tumor in which the copy-number increase involved only two overlapping clones from a six-clone contig (Fig. 1c). Notably, these data suggest that genetic background may affect the spectrum of recurrent aberrations. For example, the frequency of deletion on chromosome 6 was higher in C3B6F1 (8/23) than in B6 mice (1/12; Fig. 3). Although the difference is not quite statistically significant ($P=0.089$), a trend is suggested.

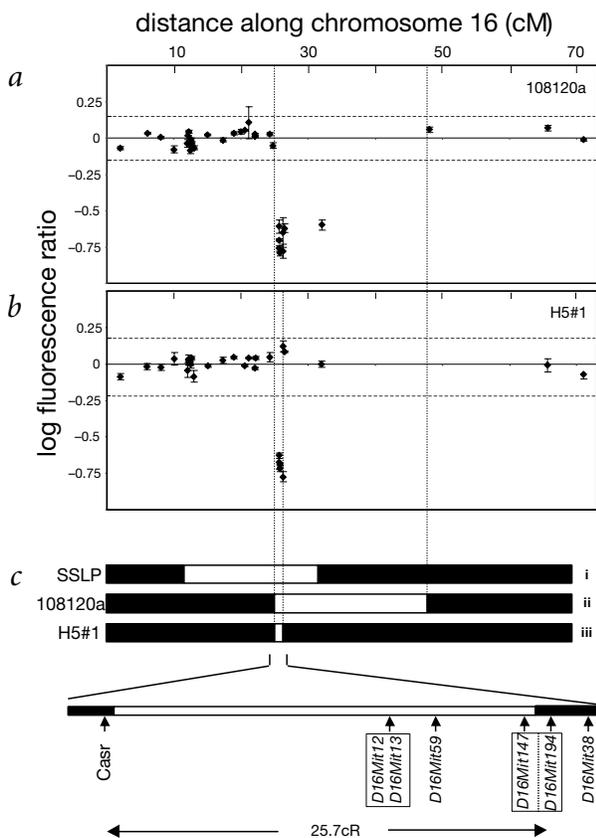


Fig. 4 Identification of chromosome 16 interstitial deletions by array CGH delimits LOH16. **a**, Normalized CGH ratios of chromosome 16 clones in a primary RIP-Tag tumor. **b**, Normalized CGH ratios of chromosome 16 clones in a RIP-Tag tumor-derived cell line H5#1. Array CGH data are presented as described in Fig. 1. The proximal breakpoints in the tumor and cell line occur between the same markers (*Casr* at 26.3 cM and *D16Mit12* at 27.6 cM), whereas the distal breakpoints differ. **c**, Schematic representation of the smallest common region of deletion (white) on chromosome 16. Vertical dashed lines link the breakpoint-defining clones in **a** and **b** to the schematic representation in **c**. Subsections represent: (i) LOH16 as defined by SLP analyses between markers *D16Mit1* (14 cM) and *D16Mit14* (34 cM) (ii) the deleted region as defined by the tumor shown in **a** and (iii) the deleted region as defined by the cell line shown in **b**. A blowup of the smallest common deleted region (white) shows the markers mapping within this locus and their relative genetic positions. Boxes indicate that the enclosed markers have not been resolved genetically. The dashed line between markers *D16Mit147* and *D16Mit194* indicates that the proximal breakpoint in cell line H5#1 occurs between these two markers. RH mapping data places flanking markers *Casr* and *D16Mit194* a distance of 25.7 cR apart, corresponding to an estimated physical distance of 2.9 Mb.

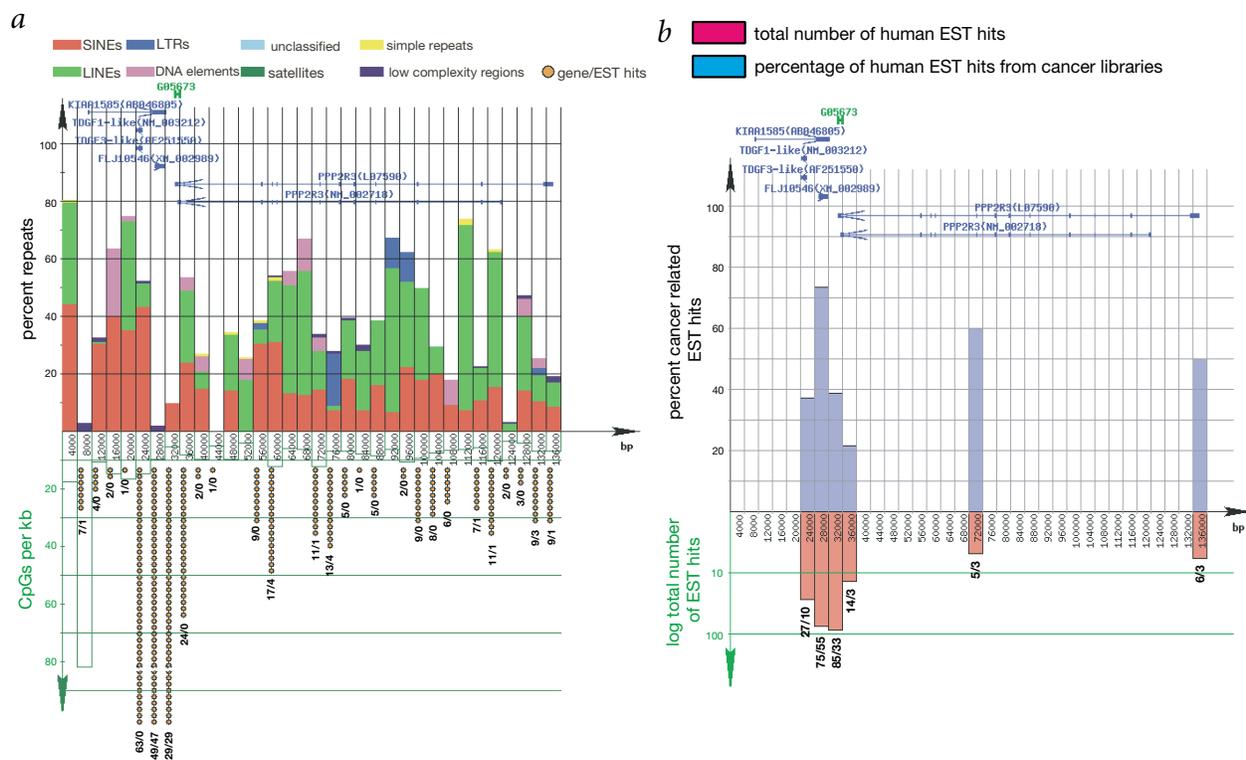


Fig. 5 Genome Cryptographer annotation of human genomic sequence syntenic to the region of chromosome 9 loss encoding the apoptotic regulator *PPP2R3*. **a**, Genes and real or predicted exons are plotted in the upper part of the figure according to their genomic location. The abundance of EST hits are plotted in the lower part of the figure; values represent the total number of EST hits per number of mouse EST hits. The repeat sequence contents long (LINEs) and short interspersed nuclear elements (SINEs), long terminal repeats (LTRs) and others are displayed in the upper portion of the figure. Features of the DNA sequence are shown in 4-kb windows. **b**, Percentage of human EST hits that originated in cDNA libraries created from tumor material; values in the lower portion of the figure represent the total number of human EST hits/number of EST hits from tumor-derived libraries. Genes and real or predicted exons are plotted as in **a**.

We also tested whether irradiation of tumor-precursor cells would increase the frequency of informative copy-number abnormalities. We analyzed 15 additional tumors from C3B6F1 mice that received 600 rads of gamma irradiation at 6 wk, when the mice had early hyperplastic, pre-angiogenic lesions but no solid tumors. Irradiation changed both the frequency and the spectrum of recurrent aberrations. The frequency of deletions involving LOH9 increased significantly in the irradiated mice, whereas the frequency of deletions involving LOH16 did not (Fig. 3). In addition, the gain of chromosome 4 occurred significantly less frequently in irradiated mice, whereas loss of chromosome 14 occurred more frequently. This suggests that environmental effects can alter the biology and genetics of tumor development, but complicates efforts to use irradiation to facilitate positional localization of cancer genes.

We used the draft human sequence^{9,10} assembled at <http://genome.ucsc.edu> (October 2000 freeze) to identify genes in regions syntenic to the recurrent copy-number abnormalities (CNAs) in the murine pancreatic islet carcinomas. We used a Genome Cryptographer¹¹ to display human genes and ESTs mapping to these regions in addition to *in silico* information about the tumor and normal tissue expression patterns. Genes mapping to the large regions of loss on chromosomes 6 and 14 were identified on the Y2000 mouse chromosome committee maps (<http://www.informatics.jax.org/ccr/>). Fig. 5 shows the Genome Cryptographer annotation of the human region syntenic to the narrowly defined region of loss on mouse chromosome 9. Detailed analyses of regions of abnormality on chromosomes 2, 4, 8, 9 and 16 can be seen at http://shark.ucsf.edu:8080/~graeme/Nat_Gen_2001.html.

These analyses suggested candidates for several regions of recurrent abnormality. The region of increased copy number on chromosome 2 is syntenic to a region of amplification in human breast cancer containing the genes *CYP24* and *PFDN4* (ref. 12). Amplification of *CYP24* has been implicated in loss of vitamin D-mediated differentiation¹². *PFDN4* is a subunit of the heterohexameric chaperone protein prefoldin family¹³ and may regulate transcription or function as a cofactor in cell-cycle regulation¹⁴. The gene *STMN1* (op18, stathmin) maps to the region of copy-number increase on chromosome 4. This gene encodes a cytosolic phosphoprotein thought to regulate cell division through its effects on microtubule dynamics and is overexpressed in a variety of human malignancies¹⁵. The region of chromosome 6 loss contains the tumor suppressor *CDKN1B* (p27^{Kip1}), encoding a cell cycle-dependent kinase inhibitor frequently downregulated in human tumors¹⁶ that mediates tumor suppression in mice in a haplo-insufficient manner¹⁷. The region of chromosome 9 loss associated with downregulation of apoptosis in RIP-Tag tumors includes *PPP2R3*. This gene encodes a regulatory subunit of protein phosphatase 2A, a serine/threonine phosphatase known to regulate the phosphorylation status of apoptotic regulators BCL2 (ref. 18) and BAD (ref. 19). The region of loss on chromosome 16 contains *FSTL1* (follistatin-like 1). *FSTL1* was identified in a screen for TGFβ1-induced genes²⁰ and has significant homology to *SPARC*, encoding a multifunctional matricellular glycoprotein that inhibits the mitogenic effects of VEGF²¹. This anti-angiogenic effect is consistent with our hypothesis that LOH16 carries a tumor-suppressor gene that negatively regulates tumor angiogenesis. We found no obvious candidate genes for the chromosome 8 deletion.



We have described a useful array CGH procedure for the analysis of genome copy-number changes in murine tumors that develop in transgenic mice and demonstrated its use in the localization of genes associated with tumorigenesis. The measurement precision of array CGH is sufficiently high that even single copy-number changes occurring in a fraction (>40%) of the cells of a tumor can be detected and mapped. Array CGH is particularly useful for the detection of genomic gains and losses in mouse cancer models, as it obviates the need for slow and expensive out-crossing of tumor-prone mice to other genetic backgrounds to produce the polymorphisms required for SSLP analysis. In addition, it avoids the changes in phenotype that out-crossing may cause by introducing modifier genes.

Methods

Tumor genomic DNA isolation. We isolated end-stage tumors from RIP-Tag mice at 14–15 wk and purified oncogene-expressing transformed β -cells from whole tumors as described²². We extracted genomic DNA using the Wizard DNA Purification Kit (Promega), further purified it with a phenol/chloroform extraction and quantitated it using a fluorometer. Phenol/chloroform extraction of the resulting DNA increased measurement precision significantly in some experiments, presumably by removing proteins that interfered with DNA labeling and hybridization.

Array preparation. We isolated BAC DNA using Plasmid Maxi Kits (QIAGEN) following the manufacturer's recommendations and then further purified it by phenol/chloroform extraction. The identities of the BAC clones are listed at http://shark.ucsf.edu:8080/~graeme/Nat_Gen_2001.html. Clones are available from Research Genetics (<http://www.resgen.com>). We carried out DOP-PCR amplification in 96-well format. We included a water control in each 96-well PCR to ensure that contaminating DNA was not present. We included 1 μ l of BAC DNA (concentrations ranged from 50 ng/ μ l to 500 ng/ μ l) in a final reaction volume of 100 μ l. Each reaction mix contained 3 mM MgCl₂, 5 U *Taq* polymerase (Life Technologies), 200 μ M dNTPs, 1 \times PCR buffer and 1.5 μ M 5' amine-modified DOP primer (5'-CCGACTCGAGNNNNNNATGTGG-3' as described)⁸. The primer contained a 5' amine group that covalently linked the PCR products to the amine-reactive slide surface. A 3-min, 94 °C denaturation step was followed by 25 cycles of 94 °C for 30 s, a 37–72 °C linear ramp for 10 min, and 72 °C for 1 min, with a final 7-min extension at 72 °C. PCR reactions (2 μ l) checked on 1% agarose gels typically showed DNA fragments ranging in size from 0.2 kb to 5 kb. We observed distinct bands, however, suggesting that some sequences within each BAC were preferentially amplified. The concentration of DNA following PCR was typically around 100–120 ng/ μ l. We precipitated samples with ethanol and resuspended them overnight in 10 μ l of printing buffer (150 mM sodium-phosphate, pH 8.5). Samples were then transferred to 864-well round-bottom polypropylene plates (Whatman) and printed in quadruplicate onto 3D-Link activated slides (Surmodics) using a custom DNA arraying device developed at UCSF Cancer Center (D. Pinkel, personal communication). We printed replicate spots adjacent to each other on the array. The center-to-center spacing of the spots on the arrays was 150–175 μ m, depending on the print. Two arrays were printed per slide, each covering a 6 mm \times 9 mm area. Slides were treated and stored after printing according to the manufacturer's recommendations.

Hybridization to microarrays. We labeled 500 ng of test (tumor) and reference genomic DNA with CY3 and CY5 (Amersham), respectively as described⁵. Labeled DNA samples were purified using Qiaquick PCR purification columns (Qiagen), co-precipitated with 50 μ g of mouse Cot-1 DNA (Life Technologies) and resuspended in 20 μ l hybridization buffer (50% formamide, 10% dextran sulfate, 2 \times SSC, 2% SDS, 200 μ g yeast tRNA). We heated the hybridization mix at 75 °C for 10 min, and then carried out a 60-min incubation at 37 °C to allow pre-annealing of Cot-1. We processed Surmodics slides for hybridization following the manufacturer's recommendations; this included a 2-min denaturation step in boiling water. We placed a rubber cement dam around each array using a 5-ml syringe fitted with a 20- μ l pipette tip and allowed the rubber cement to dry for approximately 45 min on a 37 °C slide warmer. We added pre-annealed hybridization mix to the array, gently swirled it to cover the array area and placed the slide in a plastic

slide holder (pre-warmed to 37 °C) containing 250 μ l of wash buffer (50% formamide, 2 \times SSC, pH 7.0) to prevent evaporation. The hybridization chamber was placed at 37 °C on a slowly rocking table to move the hybridization mix over the array area for 48–72 h. Signal intensities increased notably when the incubation was increased from 24 h to 48 h; however, we saw no further benefit from extending it to 72 h. After hybridization, we washed the slides for 15 min at 48 °C in 50% formamide, 2 \times SSC, pH 7.0 and then for an additional 30 min at 48 °C in 2 \times SSC, 0.1% SDS. We removed the rubber cement dam and washed the slide for an additional 10 min in PN buffer (0.1 M sodium phosphate buffer, 0.1% Nonidet P-40, pH 8.0) at room temperature. We briefly dipped the slides in 2 \times SSC and mounted them with DAPI counter stain (0.5 μ M DAPI, 0.1 \times PBS, 90% glycerol) for imaging. Most of the data presented are based on the results of a single hybridization. Our experiences with repeatedly measuring the same DNA sample, however, show that the results were highly reproducible. Fig. 2, for example, demonstrates reliable and reproducible detection of genomic aberrations even in the presence of substantial admixed normal genomic DNA. In addition, limited analyses using the hybridization protocol described here allowed detection of single genome-copy number changes using arrays from Spectral Genomics (<http://www.spectralgenomics.com>).

Image and data analyses. We collected 16-bit TIF images using a CCD camera through CY3, CY5 and DAPI filters and analyzed them using custom software as described⁴. We excluded from further analysis spots that were contaminated with fluorescent debris and those that failed to print. These spots could in general be recognized using criteria known to affect reliable ratio measurements, including low reference-hybridization intensity, low DAPI intensity, low correlation of the CY3-versus-CY5 pixel intensities and low pixel number for the segmented spot. We normalized data to the median raw CY3:CY5 ratio and converted numbers to the log₂ domain to weight gains and losses equally. We calculated the mean and s.d. of the quadruplicate spots for the normalized log₂ ratios. Following this analysis, we eliminated clones if the s.d. of the four replicate spots exceeded 0.2 or if the ratio measurement was based on a single spot. We assessed the effect of reducing the number of replicate spots by reanalyzing data from three different tumors, with two of the four replicates removed, carrying out separate analyses in which the first, middle or last two spots were removed. We saw negligible differences between the data sets with respect to gains and deletions, s.d. of the central Gaussian distribution and percent of clones eliminated by the removal of bad spots. This suggests that reducing the number of replicate spots per array has minimal impact on ratio measurements. A summary of the data is available at http://shark.ucsf.edu:8080/~graeme/Nat_Gen_2001.html.

Statistical analysis. We used the data from within an experiment to determine the thresholds above or below which CGH ratios were scored as increased or decreased, as the variation observed within an experiment differed slightly between experiments. We accomplished this by using a hybrid least-squares/maximum likelihood method to fit a mixture of three Gaussian distributions to a histogram of log₂ ratios from each CGH array analysis. The fitted distribution was

$$f(y) = \sum_{i=1}^3 p_i (2\pi\sigma_i)^{-1/2} e^{-\frac{1}{2} \left(\frac{y-\mu_i}{\sigma_i} \right)^2}$$

where (p_i, μ_i, σ_i) were the relative proportion, mean and s.d. of the i th distribution. We obtained the fit by first plotting the histogram of log₂ ratios and visually selecting initial estimates for means and s.d. for each of the three components of the combined data: (i) a 'normal' component centered at 0 (ii) a 'loss' component centered at a mean less than zero and (iii) a 'gain' component centered at a mean greater than zero. The fitting program provided visual feedback in the form of a smooth fit line superimposed on the observed data histogram, so that the user could see the effect of changing the initial estimates for the means, s.d. and component proportions. More than 95% of the data sets could be fitted accurately using initial estimates that assumed the 'normal' component contained 85% of the data points and a σ of 0.1, and that the 'lost' and 'gained' components each contained 7.5% of the data points and a σ of 0.2. After we obtained initial estimates, we minimized the sum of squared deviations of the fit from the observed histogram using



the Solver in Microsoft Excel. We then used the new estimates to maximize the likelihood of the individual observations. This two-step method was more reliable than applying maximum likelihood directly to the data. In addition, the visual feedback provided by the histogram and fit line plot allowed rapid assessment of the quality of the fit. The Gaussian distributions fitted to the parts of the histogram representing gains and losses usually did not influence the parameters of the Gaussian distribution fitted to the 'normal' part of the histogram, except when the CGH ratios for the gains and losses were not well separated from normal (an unusual event). We used the distribution fitted to the part of the histogram representing genomically 'normal' values to determine 3σ upper and lower thresholds for determining gains and losses. CGH analyses in which σ exceeded 0.15 were considered unreliable and were eliminated from further analyses (~6% of the data sets).

DNA sequence analysis. The regions of human sequence syntenic to recurrent CNAs on mouse chromosomes 2, 4, 9 and 16 were defined by BLAST searches of the htgs database of GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>) using mouse sequences (markers, genes or BAC ends) mapping to the boundaries of CNAs. We used Genome Cryptographer (<http://kinase.ucsf.edu/gc/>) to collect genome sequence information from multiple databases and visually display it in analysis intervals of constant width along the genome. Information displayed in the analyses for this paper included genes, EST clusters and fractions of ESTs originating in tumor tissues.

Acknowledgments

This study was supported by USPHS grants CA78601 and R37-CA45234 and by the UCSF Comprehensive Cancer Center Core Grant (CA82103).

Received 22 August; accepted 16 October 2001.

- Dietrich, W.F. *et al.* Genome-wide search for loss of heterozygosity in transgenic mouse tumors reveals candidate tumor suppressor genes on chromosomes 9 and 16. *Proc. Natl Acad. Sci. USA* **91**, 9451–9455 (1994).
- Parangi, S., Dietrich, W., Christofori, G., Lander, E.S. & Hanahan, D. Tumor suppressor loci on mouse chromosomes 9 and 16 are lost at distinct stages of tumorigenesis in a transgenic model of islet cell carcinoma. *Cancer Res.* **55**, 6071–6076 (1995).
- Shi, Y.P. *et al.* DNA copy number changes associated with characteristic LOH in islet cell carcinomas of transgenic mice. *Genes Chromosomes Cancer* **19**, 104–111 (1997).
- Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.* **20**, 207–211 (1998).
- Pollack, J.R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).
- Solinas-Toldo, S. *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**, 399–407 (1997).
- Heiskanen, M.A. *et al.* Detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer Res.* **60**, 799–802 (2000).
- Telenius, H. *et al.* Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725 (1992).
- Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Collins, C. *et al.* Comprehensive genome sequence analysis of a breast cancer amplicon. *Genome Res.* **11**, 1034–1042 (2001).
- Albertson, D.G. *et al.* Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nature Genet.* **25**, 144–146 (2000).
- Vainberg, I.E. *et al.* Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**, 863–873 (1998).
- Iijima, M., Kano, Y., Nohno, T. & Namba, M. Cloning of cDNA with possible transcription factor activity at the G1-S phase transition in human fibroblast cell lines. *Acta Med. Okayama* **50**, 73–77 (1996).
- Curmi, P.A. *et al.* Overexpression of stathmin in breast carcinomas points out to highly proliferative tumours. *Br. J. Cancer* **82**, 142–150 (2000).
- Philipp-Staheli, J., Payne, S.R. & Kemp, C.J. p27(Kip1): regulation and function of a haploinsufficient tumor suppressor and its misregulation in cancer. *Exp. Cell Res.* **264**, 148–168 (2001).
- Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M. & Kemp, C.J. The murine gene *p27Kip1* is haplo-insufficient for tumour suppression. *Nature* **396**, 177–180 (1998).
- Deng, X., Ito, T., Carr, B., Mumby, M. & May, W.S. Jr. Reversible phosphorylation of Bcl2 following interleukin 3 or bryostatin 1 is mediated by direct interaction with protein phosphatase 2A. *J. Biol. Chem.* **273**, 34157–34163 (1998).
- Chiang, C.W. *et al.* Protein phosphatase 2A activates the proapoptotic function of BAD in interleukin-3-dependent lymphoid cells by a mechanism requiring 14-3-3 dissociation. *Blood* **97**, 1289–1297 (2001).
- Shibanuma, M., Mashimo, J., Mita, A., Kuroki, T. & Nose, K. Cloning from a mouse osteoblastic cell line of a set of transforming-growth-factor- β 1-regulated genes, one of which seems to encode a follistatin-related polypeptide. *Eur. J. Biochem.* **217**, 13–19 (1993).
- Kupprion, C., Motamed, K. & Sage, E.H. SPARC (BM-40, osteonectin) inhibits the mitogenic effect of vascular endothelial growth factor on microvascular endothelial cells. *J. Biol. Chem.* **273**, 29635–29640 (1998).
- Efrat, S. *et al.* Beta-cell lines derived from transgenic mice expressing a hybrid insulin gene-oncogene. *Proc. Natl Acad. Sci. USA* **85**, 9037–9041 (1988).
- Han, C.S. *et al.* Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. *Genome Res.* **10**, 714–721 (2000).
- Shi, Y.P. *et al.* FISH probes for mouse chromosome identification. *Genomics* **45**, 42–47 (1997).
- Korenberg, J.R. *et al.* Mouse molecular cytogenetic resource: 157 BACs link the chromosomal and genetic maps. *Genome Res.* **9**, 514–523 (1999).