Identifying Allelic Loss and Homozygous Deletions in Pancreatic Cancer without Matched Normals Using High-Density Single-Nucleotide Polymorphism Arrays

Eric S. Calhoun,1 Tomas Hucl,1 Eike Gallmeier,1 Kristen M. West,3 Dan E. Arking,3 Anirban Maitra,1,2 Christine A. Iacobuzio-Donahue,2 Aravinda Chakravarti,1 Ralph H. Hruban,1,2 and Scott E. Kern1,2

Abstract
Recent advances in oligonucleotide arrays and whole-genome complexity reduction data analysis now permit the evaluation of tens of thousands of single-nucleotide polymorphisms simultaneously for a genome-wide analysis of allelic status. Using these arrays, we created high-resolution allelotype maps of 26 pancreatic cancer cell lines. The areas of heterozygosity implicitly served to reveal regions of allelic loss. The array-derived maps were verified by a panel of 317 microsatellite markers used in a subset of seven samples, showing a 97.1% concordance between heterozygous calls. Three matched tumor/normal pairs were used to estimate the false-negative and potential false-positive rates for identifying loss of heterozygosity: 3.6 regions (average minimal region of loss, 720,228 bp) and 2.3 regions (average heterozygous gap distance, 4,434,994 bp) per genome, respectively. Genomic fractional allelic loss calculations showed that cumulative levels of allelic loss ranged widely from 17.1% to 79.9% of the haploid genome length. Regional increases in “NoCall” frequencies combined with copy number loss estimates were used to identify 41 homozygous deletions (19 first reports), implicating an additional 13 regions disrupted in pancreatic cancer. Unexpectedly, 23 of these occurred in just two lines (BxPc3 and MiaPaCa2), suggesting the existence of at least two subclasses of chromosomal instability (CIN) patterns, distinguished hereby allelic loss and copy number changes (original CIN) and those also highly enriched in the genomic “holes” of homozygous deletions (holey CIN). This study provides previously unavailable high-resolution allelotype and deletion breakpoint maps in widely shared pancreatic cancer cell lines and effectively eliminates the need for matched normal tissue to define informative loci. (Cancer Res 2006; 66(16); 7920-8)

Introduction
A greater efficiency in analyzing loss of heterozygosity (LOH) in cancer genomes is needed. Long-standing problems have hampered efforts to characterize LOH, particularly in tumor types such as adenocarcinoma of the pancreas, which contain as few as 10% neoplastic cells (1). Primary tissues are also inherently limited resources, which eventually deplete through analysis. Access is often limited to select individuals/institutions where patients are seen, limiting the ability of the scientific community to validate published results. By establishing in vitro and/or xenograft cultures of carcinoma cells, many of the problems associated with primary tissues can be overcome. From a genetic perspective, early-passage cell lines represent an almost ideal form of tissue free of contaminating normal cells. With unlimited expandability, as well as portability, cell lines have the potential to be shared freely between investigators and/or established in commercial cell-line banks. Unfortunately, in vitro cultures generally lack a matched normal tissue. In the past, this impaired the ability to discern between hemizygosity and homozygosity as well as between polymorphic (germ line) variants and true somatic mutations.

The efficiency of analyzing allelic loss is also dependent on technological limitations. The analysis of microsatellite markers has been the gold standard for estimating allelic loss because the rate of heterozygosity at each locus is comparably higher than that seen for diallelic single-nucleotide polymorphisms (SNP; refs. 2–4). Such low heterozygosity frequencies have been one reason why analyses based on SNPs have found limited use in genetic studies (5). Technical difficulties in the scalability for microsatellite marker studies, however, limit the number of samples and/or the number of markers used, effectively ensuring a low genomic coverage and poor resolution of LOH breakpoints (6). Recent advances in detecting chromosomal copy number (e.g., bacterial artificial chromosome array–based comparative genomic hybridization and representational oligonucleotide microarray analysis; refs. 7–11) have provided a solution to some of these problems but fail to identify many cases of LOH, as when uniparental disomy or polysomy is present. Additionally, genomic coverage remains somewhat limited by the number of probes arrayed (recently ∼ 5,400 probes per array; ref. 10).

New high-throughput technologies, such as the Affymetrix genotyping arrays which evaluate >100,000 genotypic-variant loci simultaneously, effectively allow for a high-resolution genome-wide analysis of allelic status on large sample sets (3, 12). The reduced informativeness of diallelic markers is overcome by an increase in the number of evaluated loci and subsequently the number of informative calls. For example, Matsuzaki et al. (3) reported a heterozygous frequency rate of 30.41% using the 100K SNP arrays, which would yield nearly 30,000 informative calls in diploid individuals. The quality of the results, however, is dependent on the purity of the starting tissue. Allelic loss studies are dramatically complex and require high-quality starting material.
affected by contaminating normal tissues, a condition common to many scirrhous neoplasms (1, 13), and the effect of contaminating mouse tissues in xenografts has yet to be determined for SNP analysis. Consequently, studies of these tumor systems using the Affymetrix genotyping arrays are limited to established cell line cultures.

By combining the analysis of established, commercially available pancreatic cancer cell lines, as well as several early-passage cell lines with the Affymetrix 100K genotyping arrays, we have overcome long-standing problems associated with allelic-loss analysis in pancreatic cancer. We present here the high-resolution allelotype and breakpoint maps of 26 pancreatic cancer cell lines without the use of matched normal tissues.

Materials and Methods

**Cell lines.** The pancreatic cancer lines AsPC1, BxPc3, CAPAN1, CFPAC1, Hs766T, MiaPac2, PL1 (Panc 02.13), PL9 (Panc 08.13), PL5 (Panc 04.03), PL11 (Panc 03.27), Panc-1, and Su86.86 were purchased from the American Type Culture Collection (Manassas, VA). COLO357 was obtained from the European Collection of Animal Cell Cultures (Salisbury, United Kingdom). PK8 and PK9 were kind gifts from Dr. Arata Horii to C.A.I-D. (14). XPA1 (PXA184), XPA2 (PXA196), and XPA3 (PXA154) were established by A.M. and described elsewhere (15). The remaining pancreatic cancer lines, PL3 (Panc 04.14), PL13 (Panc 09.06), PL19 (TS0111), PL20 (TS0120), PL21 (TS0129), PL22 (TS0226), PL23 (TS031), and PL24 (TS065), were kind gifts from Dr. Elizabeth M. Jaffee (Johns Hopkins, Baltimore, MD). Lines were propagated in RPMI 1640 (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum, 2 mmol/L L-glutamine, 100 units/ml penicillin, and 100 µg/ml streptomycin, unless otherwise specified by the supplier, and maintained in a 5% CO2 humidified incubator at 37°C. Genomic DNA was isolated from cell lines using a Qiagen Blood Tissue kit ( Valencia, CA) and used directly in the assays described herein.

**Microsatellite marker analysis.** Microsatellite marker mapping was done in collaboration with the NIH Center for Inherited Disease Research using a total of 386 microsatellite markers (Supplementary File 1; ref. 4). The positions of 317 of these markers (chromosomes 1-22) are currently mapped by the Ensembl database (16) and were used for comparisons made here. The genotype (the identity of the alleles at a particular locus) for each marker was “called” as homozygous/hemizygous (A) or heterozygous (B) based on the resolution of one or two PCR bands, respectively. If the PCR reaction yielded no detectable band, the marker was classified as missing (M). Homozygous/hemizygous and missing calls were considered uninformative whereas heterozygosity indicated the retention of both parental alleles. Data were visualized with Excel (Microsoft, Redmond, WA) by plotting the calls with respect to their genomic position in the May 2004 human genome build.

**SNP allele mapping.** The genotypes of 115,353 SNPs were analyzed using the Affymetrix CentXba and CentHind oligonucleotide arrays hybridized to reduced-complexity genomic DNA as described (3). Briefly, 250 ng of genomic DNA were digested with either XhoI or HindIII adapters were ligated to the digested DNA, and PCR was done to preferentially amplify fragments 250 to 2,000 bp in length. Samples were then fragmented, fluorescently labeled, and hybridized to the arrays according to the protocol of the manufacturer. Genotypes were called by the GeneChip DNA Analysis Software Tool (version 3.0) using a 0.05% error setting (Supplementary Files 2-5). Four potential genotype calls were made: AA, BB, AB, and NoCall. No distinction was made between homozygosity for either allele; each result was considered uninformative and provided only a visual reference to indicate genome coverage. Heterozygous calls indicated the retention of both alleles and were considered informative. "NoCall" data were not used in the analysis of LOH. Genotyping calls were visualized by plotting with respect to their genomic position listed in the May 2004 assembly of the human genome. Any probes of which the position was not known or not annotated by GeneChip DNA Analysis Software Tool were subsequently omitted as were SNPs with a Caucasian heterozygous frequency of zero (3). This left a maximum of 104,502 evaluated SNP loci per sample.

**Critical gap distance for LOH.** The distances between heterozygous calls for each of the 42 Caucasian normal, EBV-transformed, lymphoblastoid cell lines were calculated from previously published allelotypes (3). An interheterozygous call distance of 2.75 million bp (Mbp) represented a value greater than the 99.9th percentile of all gap sizes observed. This distance was used as the cutoff to eliminate smaller regions of contiguous homozygous SNP markers that occur naturally (i.e., by chance) in each cell line. Candidate regions evaluated by at least 40 SNPs were considered as "critical" gaps, indicating instances of LOH. Immediately adjacent critical gaps were combined and represented as one contiguous region of LOH. Instances of regions occurring between critical gaps, or those occurring at the ends of chromosome arms, were combined into one contiguous region at our best judgment. Centromeric regions were included in LOH estimates when critical gaps were located immediately adjacent on both p and q arms; otherwise, each breakpoint estimate was terminated using the position of the last SNP evaluated by the arrays on each respective arm. Excessively homozygous regions (between 2.75 and 8 Mbp and at least 40 SNPs) identified in the 131 Caucasian normals (Supplementary Table S1) were removed from the list of critical gaps as they may represent false-positive events of identity-by-descent. Finally, homozygous regions >8 Mbp identified in the normals were believed to represent true LOH events affecting the immortalized “normal” cell line samples and, as such, were not used to remove critical gaps from the cell line data.

**Precision of critical gap breakpoints.** Matched normal tissue (not immortalized cell lines) was available for three of the early-passage carcinoma cell lines (PL9, PN139; PL11, PN192; PL13, PN147) and were used to test the method for estimating critical gaps. SNP allelotypes for each normal were determined as described above and compared with the respective carcinoma line allelotype maps. Gaps containing at least three informative hemizygous markers were used to identify a region of true LOH, and the positions of the outermost hemizygous markers established the minimal region of loss (MRL) breakpoints. Regions identified as lost using the matched normals, but not identified as a critical gap, were used to estimate the false-negative rate. Critical gaps containing less than three informative markers, as determined from the matched normal genotypes, were used to estimate the false-positive rate. Critical gaps showing true loss were compared with the MRL estimates to gauge the precision of our allelic loss breakpoint positions.

**Estimation of total allelic loss.** The extent of total allelic loss was estimated by calculating genomic fractional allelic loss (GFAL; ref. 4), with the following modifications. An expected Caucasian frequency of total heterozygous calls for the Affymetrix CentXba and CentHind oligonucleotide arrays (0.3041; ref. 3) was used as an estimate of the total number of informative markers. GFAL was calculated by dividing the total observed heterozygous frequencies by 0.3041, subtracting from 1, and converting to a percent.

**Statistical analysis.** The Spearman rank correlation coefficient was used to evaluate whether the degree of allelic loss (i.e., increasing levels of GFAL) correlated with the number of years because the cell line was isolated and established in culture. The year of establishment was obtained from published records (17–23) or from the respective suppliers and ranked with the oldest as last (ranked 24th, Table 1). GFAL scores were similarly ranked in descending order with cell lines having the highest degree of allelic loss last (ranked 24th). Cell lines with the same year of establishment were assigned an average of the ranks for that year. A high level of significance \( p < 0.01; \text{critical region } r > 0.485 (n = 29) \) was used to evaluate the test statistic. The critical value as well as the method for calculation of the test statistic was obtained from published sources (24). The microsatellite unstable lines, PL3 and PL5, were not included in this analysis as they showed no deviation of GFAL from the 42 published Caucasian controls (Table 1 and data not shown; ref. 3).

**Copy number analysis.** Single-point analysis of copy number estimates for each SNP was done by the Copy Number Analysis Tool (version 2.2;
Supplementary Files S2-S5; ref. 25). Variations of copy number estimates were reduced by applying a positional moving average algorithm where each SNP copy number value was averaged with a maximum of 10 immediately adjacent (5 telomeric and 5 centromeric) SNP copy number estimates; only adjacent SNPs positioned within 750kb of the central SNP were included in the average. Data from the CentXba and CentHind chips were analyzed separately and plotted individually to compare copy number variations between arrays.

Homozygous deletions. Dramatic drops in copy number values below 1.0, evidenced by both arrays, were indicative of a true homozygous deletion. With decreasing regional size (and/or corresponding copy number averages closer to 1.0), the ability to distinguish between homozygous deletion and LOH was reduced. To compensate, we calculated a 20-SNP moving average of the frequency of NoCall genotypes across the genome for each array. Agreement between increasing NoCall frequency and decreases in copy number averages was used to identify additional candidate homozygous deletions. Representative candidate regions, as well as several previously identified, were validated using PCR with at least two consecutive nonoverlapping primer sets. Regional chromosomal gains and amplifications were not specifically evaluated by this study. Raw data and positional moving average (PMA) estimates of copy number are included in Supplementary data.

Results and Discussion

Genotype determination of pancreatic cancer cell lines. Several studies have reported the usefulness of the Affymetrix 10K arrays in determining allelic loss when using matched normal tissues (26, 27). Unfortunately, even at this SNP density, evaluating allelic loss without a matched normal would yield only ~3,000 heterozygous informative calls. With the introduction of the 100K SNP arrays, genomic coverage improves 10-fold and would yield a similar 10-fold increase in heterozygous calls (i.e., ~30,000), suggesting it may be possible to do allelic studies without matched normal tissues.

To test this proposal, we evaluated an initial panel of seven pancreatic cancer cell lines (AsPc1, BxPc3, COLO357, Hs766T, MiaPaCa2, Panc-1, and Su86.86) using the Affymetrix CentXba and CentHind oligonucleotide arrays. A total of 104,502 loci were evaluated per cell line, yielding informative heterozygous calls ranging from 18.8% (17,049 of 90,822) to 6.1% (5,326 of 87,005). The data were plotted with respect to their physical positions, indicating several areas of chromosomal loss (Fig. 1 and Supplementary Figs. S1-S9). Validation of the experimental method and data was accomplished using a large panel of microsatellite markers. In total, 2,001 microsatellite calls were obtained, yielding 729 informative heterozygous calls. Heterozygous microsatellite frequencies ranged from 23.3% (65 of 279) to 47.2% (125 of 265) for each cell line (Supplementary Table S2).

We found a high concordance between SNP and microsatellite calls. Of the 729 microsatellite heterozygous calls, 708 (97.1%) were in areas also characterized by the SNP arrays to contain two alleles. The discordant calls were distributed among 13 separate chromosomes in six of these seven cell lines and involved 16 individual critical gap regions (Supplementary Table S2). We felt several possible explanations existed for these discrepancies,

<table>
<thead>
<tr>
<th>Cell lines</th>
<th>Heterozygous genotyping calls</th>
<th>All genotyping calls</th>
<th>Heterozygous calls (%)</th>
<th>GFAL (%)*</th>
<th>Year established (rank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiaPaCa2</td>
<td>5,326</td>
<td>87,005</td>
<td>6.12</td>
<td>79.87</td>
<td>1975 (2.5)</td>
</tr>
<tr>
<td>XPA1</td>
<td>7,865</td>
<td>87,664</td>
<td>8.97</td>
<td>70.50</td>
<td>2003 (20.5)</td>
</tr>
<tr>
<td>PL9</td>
<td>8,465</td>
<td>85,485</td>
<td>9.90</td>
<td>67.44</td>
<td>1995 (13.5)</td>
</tr>
<tr>
<td>PL21</td>
<td>10,312</td>
<td>93,263</td>
<td>11.06</td>
<td>63.64</td>
<td>2003 (20.5)</td>
</tr>
<tr>
<td>COL0357</td>
<td>11,480</td>
<td>102,273</td>
<td>11.22</td>
<td>63.09</td>
<td>1978 (5)</td>
</tr>
<tr>
<td>Hs766T</td>
<td>11,128</td>
<td>90,135</td>
<td>12.35</td>
<td>59.40</td>
<td>1973 (1)</td>
</tr>
<tr>
<td>Panc-1</td>
<td>11,700</td>
<td>90,208</td>
<td>12.97</td>
<td>57.35</td>
<td>1975 (2.5)</td>
</tr>
<tr>
<td>XPA2</td>
<td>12,914</td>
<td>92,140</td>
<td>14.02</td>
<td>53.91</td>
<td>2003 (20.5)</td>
</tr>
<tr>
<td>CAPAN1</td>
<td>13,603</td>
<td>93,014</td>
<td>14.62</td>
<td>51.91</td>
<td>1977 (4)</td>
</tr>
<tr>
<td>PL24</td>
<td>14,785</td>
<td>99,289</td>
<td>14.89</td>
<td>51.03</td>
<td>2002 (16.5)</td>
</tr>
<tr>
<td>BxPc3</td>
<td>13,934</td>
<td>91,565</td>
<td>15.22</td>
<td>49.96</td>
<td>1982 (7)</td>
</tr>
<tr>
<td>PK8</td>
<td>14,826</td>
<td>92,706</td>
<td>15.99</td>
<td>47.41</td>
<td>1986 (9)</td>
</tr>
<tr>
<td>PL1</td>
<td>15,545</td>
<td>89,336</td>
<td>17.40</td>
<td>42.78</td>
<td>1995 (13.5)</td>
</tr>
<tr>
<td>PL11</td>
<td>14,735</td>
<td>84,411</td>
<td>17.46</td>
<td>42.60</td>
<td>1995 (13.5)</td>
</tr>
<tr>
<td>PK9</td>
<td>16,697</td>
<td>93,299</td>
<td>17.90</td>
<td>41.15</td>
<td>1986 (9)</td>
</tr>
<tr>
<td>Su86.86</td>
<td>15,541</td>
<td>86,285</td>
<td>18.01</td>
<td>40.77</td>
<td>1986 (9)</td>
</tr>
<tr>
<td>PL13</td>
<td>13,661</td>
<td>75,324</td>
<td>18.14</td>
<td>40.36</td>
<td>1995 (13.5)</td>
</tr>
<tr>
<td>AnPc1</td>
<td>17,049</td>
<td>90,822</td>
<td>18.77</td>
<td>38.27</td>
<td>1981 (6)</td>
</tr>
<tr>
<td>CFPA1</td>
<td>17,938</td>
<td>93,501</td>
<td>19.18</td>
<td>36.91</td>
<td>1990 (11)</td>
</tr>
<tr>
<td>PL23</td>
<td>19,065</td>
<td>95,290</td>
<td>20.01</td>
<td>34.21</td>
<td>2002 (16.5)</td>
</tr>
<tr>
<td>PL22</td>
<td>19,055</td>
<td>93,899</td>
<td>20.29</td>
<td>33.27</td>
<td>2004 (24)</td>
</tr>
<tr>
<td>PL19</td>
<td>24,594</td>
<td>98,647</td>
<td>24.93</td>
<td>18.01</td>
<td>2003 (20.5)</td>
</tr>
<tr>
<td>PL20</td>
<td>24,582</td>
<td>98,101</td>
<td>25.06</td>
<td>17.60</td>
<td>2003 (20.5)</td>
</tr>
<tr>
<td>XPA3</td>
<td>23,357</td>
<td>92,661</td>
<td>25.21</td>
<td>17.11</td>
<td>2003 (20.5)</td>
</tr>
<tr>
<td>PL5</td>
<td>28,742</td>
<td>97,882</td>
<td>29.36</td>
<td>3.44</td>
<td>1975 (2.5)</td>
</tr>
<tr>
<td>PL3</td>
<td>30,085</td>
<td>100,023</td>
<td>30.08</td>
<td>1.09</td>
<td>2003 (20.5)</td>
</tr>
</tbody>
</table>

*GFAL was calculated using the Caucasian heterozygous genotype frequency of 30.41%.

*Spearman Rank Correlation test statistic, rs = 0.405, indicates no strong correlation between GFAL and the year of establishment at α = 0.01.
including one or more of the following: the ambiguity (i.e., precision; see below) of assigning the critical gap breakpoints as well as the choice of method to combine adjacent critical gaps into contiguous regions (affects seven regions); the presence of a mosaic population of cells in culture with different genomic constitutions (inspection of SNP densities in five regions were visibly decreased from normal but not consistent with complete loss); large regions of possible identity-by-descent that could lead to an incorrect classification as a critical gap (the size of seven critical gaps were <8 Mbp); and the possible incorrect assessment of allelic status by mis-calls in the microsatellite marker analysis.

As expected, the ability to use the microsatellite marker data to determine areas of LOH proved difficult due to the low number of

**Figure 1.** Multidimensional genotypic analysis of CAPAN1 chromosome 2 and BxPc3 chromosome 9. A, copy number moving averages for the CentHind (pink) and CentXba (dark blue) chips. B, individual heterozygous genotyping calls are plotted above the grouped uninformative homozygous and NoCall genotypes (black). C, NoCall genotypes (green, CentHind; blue, CentXba) are inversely plotted with respect to increasing frequency. Agreement between valleys in the copy number and NoCall genotype plots commonly indicates the presence of a homozygous deletion. Yellow guidelines are for reference purposes only and indicate one and four copies of the haploid genome. Three homozygous deletions are illustrated (arrows) including one previously unreported deletion at 9 Mbp in BxPc3.
informative calls. Using the SNP arrays, a 120-fold increase in heterozygous calls was obtained over the microsatellite analysis (86,158 versus 729). Using 2.86 billion bp (Chr1-22; ref. 16) as the combined length of the human genome and after subtracting centromeric and telomeric regions not evaluated with the SNP arrays (3), the average heterozygous intermarker distance between these methods varied from nearly 25.5 Mbp (microsatellite analysis) to 216 kbp (SNP arrays). Additionally, plotting of the genotype calls of the SNP arrays revealed clear demarcations of chromosomal loss even without the use of normals, which otherwise would increase the number of informative markers (Fig. 1 and Supplementary Figs. S1-S9).

Considering these results, the ease of the assay, and the time and tissue savings shown, we expanded our analysis to an additional panel of 19 pancreatic cancer lines evaluated with the 100K arrays. In total, 2.72 million loci were evaluated, yielding 2.39 million genotype calls with more than 415K heterozygous calls (Supplementary Files 2 and 3). The frequencies of informative heterozygous calls ranged from 30.1% (PL3) to 6.1% (MiaPaCa2; Table 1). The average interheterozygous marker distance was calculated to be 166 kbp.

**Regions of LOH determined by the critical gap distance.**

Data were separated by chromosome and plotted with respect to physical position. This enabled a visual comparison of overlapping regions of LOH. Large regions of homozygosity encompassing entire chromosomes or chromosomal arms were clearly evident as were smaller areas spanning <10 Mbp (Fig. 1 and Supplementary Figs. S1-S9). Lacking a matched normal control, the distinction between identity-by-descent and true areas of homozygosity accomplished through allelic loss became more difficult with a decreasing size of the region of interest. Therefore we used the published record of 42 Caucasian normals (3) to determine the average distance between heterozygous calls on the 100K SNP arrays. We assigned a critical gap distance of >2.75 Mbp (evaluated with at least 40 SNP probes) as a cutoff for filtering out the areas of chance homozygosity; this size represents greater than the 99.9th percentile of all gaps observed (data not shown). Regions >8.0 Mbp identified in the controls were not filtered out of the candidate regions identified in the carcinoma cell lines as many of these homozygous regions may indeed represent true areas of LOH, a phenomenon known to occur in clones of lymphoblasts transformed by EBV (28, 29). Regions of excessive homozygosity between 2.75 and 8 Mbp appearing within the control population were more difficult to evaluate. Several possible mechanisms could account for the observation of excessively homozygous regions in the control population. Foremost, chip design artifacts, such as decreasing SNP-probe densities as well as stretches of polymorphisms with very low minor allele frequencies in Caucasian populations (chosen to minimize gaps in genomic coverage), may lead to such phenomena. These regions may represent areas of true LOH or simply represent inherently stable chromosomal regions having little genomic variation in the population.

To gain a better perspective of the ability of the arrays to evaluate the above Caucasian population, we catalogued all gaps >2.75 Mbp (containing at least 40 evaluated SNPs) occurring in these 42 Caucasian normals, as well as a second panel of 89 HapMap CEPH normals previously published (30). In total, 178 instances of homozygosity >2.75 Mbp but <8 Mbp occurred in the 131 normal controls (Supplementary Table S1). More than half of those (n = 90) were clustered (three or more instances per cluster) within 14 common regions across the genome, suggesting the chip density or SNP heterozygous frequencies are inadequate to accurately assess the allelic status in those areas. Taken together, the 178 instances of homozygosity would represent a theoretical false-positive rate of 1.37 regions per genome analyzed. This rate decreases to 0.68 regions per genome after elimination of common clusters.

Using all of the catalogued homozygous regions from the 131 normals as a filter, we eliminated any common critical gaps occurring in the carcinoma cell lines as candidate regions of LOH as these were likely to be false positives. The remaining (qualifying) areas of LOH were then plotted (Fig. 2). The breakpoint positions of each region (Supplementary Table S3) were taken from the SNP genomic positions of the first and last heterozygous genotype call (or the terminal chromosomal SNP evaluated by the arrays for each chromosomal arm) that defined a critical gap. Chromosome arms with the greatest frequency of allelic loss (Fig. 3) were confirmed on 18q (95.8%, MADH4), 17p (91.7%, TP53), 9p (87.5%, CDKN2A), 8p (79.2%), 3p (75.0%), 6q (75.0%), and 12q (70.8%; refs. 4, 10, 31, 32). Importantly, detailed maps of allelic loss are provided for these widely shared cell lines allowing for the rapid and efficient evaluation of candidate genes for somatic mutations (Fig. 2; Supplementary Table S3).

**Critical gap distance closely estimates the MRL.** Despite the increased resolution of the arrays, the actual precision of our critical gap estimates to the MRL remained unknown. Matched normal tissues, required to define a MRL, were available for three of the early-passage cell lines (PL9, PL11, and PL13) and subsequently assayed using the SNP arrays to address this issue. Of the critical gaps previously identified for these cell lines (Supplementary Table S4), 45 of 52 (86.5%) regions were confirmed as true losses, exemplified by at least three hemizygous informative markers. Of the regions that were not confirmed, one showed loss of two informative markers and may indeed represent an area of LOH. The remaining six regions all failed to contain any informative markers despite an average of 71 SNPs evaluated per region. Of the 45 confirmed regions, the median difference between critical gap and MRL breakpoints was 288,424 bp.

The use of matched normal tissues allowed for a separate estimate of our false-positive and false-negative rates. The false-positive rate, represented by the frequency of critical gaps failing to have any informative markers (i.e., no heterozygous genotypes in the matched normal tissue), was ~2.3 per genome (N = 7). The median size of these critical gaps was 3,478,995 bp. We identified 11 instances of false negatives (i.e., regions containing at least three hemizygous markers within a single heterozygous gap), indicating a projected rate of 3.6 regions per genome. The median interheterozygous call distance for these regions was 970,860 bp with a median MRL distance of 493,645 bp. Four of these occurred on chromosome 8 of PL9 and may be indicative of a mosaic chromosome or chromosomal region undergoing change in culture. Other cell lines (e.g., AsPc1 chromosome 1q, PK9 chromosome 13, and PL21 chromosome 3p) show similar genotype patterns (Supplementary Figs. S1-S9) and may therefore underestimate LOH in these regions as aberrant heterozygous calls will interfere with the calculation of critical gaps. Single-cell dilution before determining the allelic status seems to minimize this problem (data not shown) but may introduce a clonal variability artifact to the data set, necessitating the analysis of multiple clones for each line. Each cell line was therefore only analyzed in aggregate.
**Total allelic losses estimated by GFAL.** GFAL, an estimate of the fractional total of all chromosomal losses, is ideally calculated by dividing the number of markers indicating allelic loss by the total number of informative markers. Such a calculation, however, would require a matched normal tissue to determine the informativeness of each marker. Instead, we used an average heterozygosity rate of 30.41% as a standard expected proportion for informative markers. GFAL scores were calculated to be 1.09% and 3.44% for the chromosomal stable cell lines PL3 and PL5, respectively, and varied from 6.47% to −2.40% in the Caucasian normals (data not shown). The SD of heterozygous frequencies in the controls was 0.54%, suggesting that the calculated GFAL in the tumor cell lines may incorrectly estimate total allelic loss by ±3.58% (95% confidence). In the lines having chromosome instability, we found GFAL to vary from 79.87% (MiaPaCa2) to 17.11% (XPA3) of the genome length. Increasing levels of GFAL did not correlate with the year of establishment \( r_s = 0.405 \) (critical region \( r_s > 0.485 \) \( n = 24 \)). Such a result, however, is not unexpected as the total number of cell doublings cannot be accounted for in the commercially available lines.

**Copy number determination and identification of homozygous deletions.** Using raw intensity data from each array, the Copy Number Analysis Tool estimates ploidy for each SNP locus by comparing the distribution of intensities obtained from a reference set of normal DNAs (25). Ideally, this comparison would produce a...
clear and consistent copy number estimate between each of the arrays at a particular locus (or between consecutive SNPs on the same array). In practice, however, copy number estimates often deviate significantly, leading some investigators to determine ploidy values from more direct and less variable techniques such as array-based comparative genomic hybridization (33, 34). To help reduce the variation in copy number estimates, we applied a PMA smoothing algorithm to each of the arrays and plotted the PMAs separately (Fig. 1 and Supplementary Figs. S1-S9). Agreement between the CentHind and CentXba PMA results was used to help reduce our uncertainty of regional copy number estimates.

Visual inspection of the PMA plots revealed complex genomic patterns. For example, PMAs of copy number suggest that CAPAN1 is likely a near-triploid cell line, given that the mean copy number estimates are often noninteger values (i.e., between 3 and 2 or between 2 and 1). On chromosome 2 of CAPAN1, variations in copy number as well as a reduction to homozygosity (i.e., true LOH) occurred in localized regions and did not involve the entire chromosome (Fig. 1). Reductions to homozygosity did not always coincide with copy number reduction (i.e., between ~118 and 141 Mb copy number PMAs remained at ~2.5 despite a reduction to homozygosity) and may have, in part, involved locus-restricted (interstitial) recombinational events (i.e., gene conversion). By no means was this a unique event restricted to just chromosome 2 of CAPAN1; examples are present among other chromosomes in most of the cell lines reported here (Supplementary Figs. S1-S9). Most importantly, the ability to identify regions of LOH that lack copy number reduction illustrates why evaluation of genotypes alone and associated critical gaps is crucial for LOH determination. Such regions are produced by gene conversion, mitotic recombination, and chromosomal nondisjunction combined with chromosome reduplication (35).

Copy number PMAs also proved beneficial in the identification of homozygously deleted regions. Simultaneous decreases in copy number PMAs (<1.0) from both the CentHind and CentXba arrays routinely coincided with the presence of a homozygous deletion (Fig. 1 and Supplementary Figs. S1-S9). We combined copy number data with genotyping NoCall frequencies (homozygous deletions have an inherent increase in NoCall genotypes) to help identify additional candidate regions that were partially masked by the PMA smoothing algorithm. In total, 41 homozygous deletions were identified, of which 22 were previously reported (36-43). The remaining 19 homozygous deletions represent first reports implicating an additional 13 regions (11 separate chromosomes) not previously associated with pancreatic carcinogenesis (Table 2). Unexpectedly, the total number of homozygous deletions seemed to be bimodal: of 41 homozygous deletions, 23 occurred in just two cell lines (15 in BxPc3 and 8 in MiaPaCa2). Based on GFL scores alone, such a high rate of homozygous deletions would not have been predicted for BxPc3. This suggests that the processes leading to chromosomal instability (CIN) having only few homozygous deletions (original CIN) may be distinguishable from an occasional superimposed phenotype having high numbers of these genomic holes (holey CIN).

Until recently, it was laborious and technically challenging to determine the allelic status within cancer tissues that lacked a matched normal control. Early advances in chip-based genotyping technology, evaluating nearly 10,000 loci simultaneously, improved the feasibility of studies evaluating large sample sets but still lacked the resolution to detail allelic loss breakpoints in cancer without matched normal controls. Continual improvements in array design and whole-genome complexity reduction methodologies have now made this possible, enabling the pursuit of large-scale studies that further our understanding of the genomic changes occurring in cancer. It is important to note, however, that the concept of using overlapping regions of LOH to pinpoint the location of novel tumor suppressor genes is a commonly held misconception that we addressed in the past (44). We are currently unaware of any somatically mutated tumor suppressor gene cloned directly from the boundary maps of overlapping LOH. In contrast, such boundary mapping can be successful for germ line mutations (e.g., MEN1). The mutational frequency of any gene rarely approaches the rate of LOH in tumors, with possible exceptions, such as p16 in pancreatic cancer and p53 in colorectal cancer (45). LOH maps permit one to more efficiently evaluate candidate tumor suppressor genes. Most
recently, we have used LOH maps of pancreatic cancer xenografts (4) as a tool to uncover mutations in FBXW7, FANCC, and FANCG (46, 47). Using LOH maps to triage samples (i.e., sequence only samples with LOH in the genomic region of interest) provides researchers with the ability to maximize the potential and efficiency for finding somatic mutations.

LOH represents a key structural foundation for proposed cancer genome projects, a required component for identification of gene conversion and similar chromosome recombinational events [as shown in Fig. 1 and as proposed by Cavenee et al. (35)], and a virtually required accompaniment of somatic (as opposed to polymorphic germ line) homozygous deletions, among other uses. The facile extension of LOH analysis to the most common and widely shared cancer samples (i.e., the commercially available cell lines lacking matched normals) is likely to prove essential for many anticipated cancer studies.

Acknowledgments

Received 2/23/2006; revised 5/4/2006; accepted 6/20/2006.

Grant support: NIH grants CA92624, CA92117, and CA21765; the American Lebanese Syrian Associated Charities; and an unrestricted grant from the Hopkins Capital Group, LLC.

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.

We thank Drs. Elizabeth M. Jaffee and Arata Hori for the kind gifts of cell lines, Dr. Alison P. Klein for statistical advice, Carl S. Kashuk for help in the evaluation of the HapMap data set, and Dr. Steven C. Cunningham for critically reading the manuscript.

Table 2. Summary of homozygous deletions identified by SNP arrays

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position (Mb)</th>
<th>Cell line</th>
<th>Lowest copy number* value</th>
<th>Lowest copy number moving average</th>
<th>Confirmed</th>
<th>Initial finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>141.7</td>
<td>Hs766T</td>
<td>0.040</td>
<td>0.248</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>213.2</td>
<td>BxPc3</td>
<td>0.200</td>
<td>0.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>60.3</td>
<td>AsPc1</td>
<td>0.110</td>
<td>0.381</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.2</td>
<td>MiaPaCa2</td>
<td>0.210</td>
<td>0.436</td>
<td>Yes (2)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>92.2</td>
<td>BxPc3</td>
<td>0.110</td>
<td>0.506</td>
<td>Yes (7)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>92.9</td>
<td>BxPc3</td>
<td>0.090</td>
<td>0.533</td>
<td>Yes (12)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>102.1</td>
<td>XPA1</td>
<td>0.310</td>
<td>0.806</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>157.0</td>
<td>MiaPaCa2</td>
<td>0.110</td>
<td>0.364</td>
<td>Yes (9)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>15.5</td>
<td>MiaPaCa2</td>
<td>0.100</td>
<td>0.516</td>
<td>Yes (9)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>114.4</td>
<td>MiaPaCa2</td>
<td>0.100</td>
<td>0.456</td>
<td>Yes (9)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>18.4</td>
<td>BxPC3</td>
<td>0.090</td>
<td>0.348</td>
<td>Yes (5)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>76.0</td>
<td>PL20</td>
<td>0.090</td>
<td>0.342</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>11.9</td>
<td>AsPc1</td>
<td>0.340</td>
<td>0.720</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>46.8</td>
<td>Hs766T</td>
<td>0.250</td>
<td>0.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>49.6</td>
<td>MiaPaCa2</td>
<td>0.120</td>
<td>0.446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>15.0</td>
<td>XPA3</td>
<td>0.100</td>
<td>0.621</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>35.5</td>
<td>PK9</td>
<td>0.080</td>
<td>0.345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>41.9</td>
<td>PK9</td>
<td>0.140</td>
<td>0.447</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>43.0</td>
<td>BxPc3</td>
<td>0.110</td>
<td>0.435</td>
<td>Yes (12)</td>
<td></td>
</tr>
<tr>
<td>141.4</td>
<td>BxPc3</td>
<td>0.110</td>
<td>0.410</td>
<td>Yes (5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Estimated genomic copy number.

1The number of consecutive loci assayed by PCR that confirm the presence of a homozygous deletion are indicated.
References

Identifying Allelic Loss and Homozygous Deletions in Pancreatic Cancer without Matched Normals Using High-Density Single-Nucleotide Polymorphism Arrays

Eric S. Calhoun, Tomas Hucl, Eike Gallmeier, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/66/16/7920

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2006/08/17/66.16.7920.DC1

Cited articles
This article cites 47 articles, 16 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/16/7920.full#ref-list-1

Citing articles
This article has been cited by 10 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/66/16/7920.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/66/16/7920. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.