American Journal of Pathology, Vol. 160, No. 1, January 2002

Technical Advance

Single Nucleotide Polymorphism Array Analysis of Flow-Sorted Epithelial Cells from Frozen *Versus* Fixed Tissues for Whole Genome Analysis of Allelic Loss in Breast Cancer

Elizabeth L. Schubert,* Li Hsu,†
Laura A. Cousens,* Jeri Glogovac,* Steve Self,†
Brian J. Reid,*‡ Peter S. Rabinovitch,*§ and
Peggy L. Porter*§

From the Divisions of Human Biology and Public Health Sciences, Program in Cancer Biology, and the Division of Public Health Sciences,† Program in Biostatistics, Fred Hutchinson Cancer Research Center, Seattle, and the Departments of Gastroenterology‡ and Pathology,§ University of Washington, Seattle, Washington*

Analysis of allelic loss in archival tumor specimens is constrained by quality and quantity of tissue and by technical limitations on the number of chromosomal sites that can be efficiently evaluated in conventional analyses using polymorphic microsatellite markers. Newly developed array-based assays have the potential to yield genome-wide data from small amounts of tissue but have not been validated for use with routinely processed specimens. We used the Affymetrix HuSNP assay, composed of 1494 single nucleotide polymorphism sites, to compare allelic loss results obtained from both formalin-fixed and frozen breast tissue samples. Tumor cells were separated from normal epithelia and nonepithelial cells by dissection and bivariate cytokeratin/DNA flow sorting; normal breast cells from the same patient served as constitutive normal. Allele results from the HuSNP array averaged 96% reproducibility between duplicates and were concordant between the fixed and frozen normal samples. We also analyzed DNA from the same samples after whole-genome amplification (primer extension preamplification). Although overall signal intensities were lower, the genotype data from the primer extension preamplification material was concordant with genomic DNA data from the same samples. Results from genomic normal tissue DNA averaged informative single nucleotide polymorphism at 379 (25%) loci genome-wide. Although data points were clustered and some segments of chromosomes were not informative, our data indicated that the Affymetrix HuSNP assay could provide an efficient and valid genome-wide analysis of allelic imbalance in routinely processed and whole genome-amplified pathology specimens. *(Am J Pathol 2002, 160:73–79)*

Loss of heterozygosity (LOH), or allelic loss, is one of the most frequent genetic abnormalities in breast cancer. It may serve as a marker of generalized genomic instability, and when frequently observed in a region, it is considered indirect evidence for the presence of a tumor suppressor gene within that region of loss. In sporadic breast cancer, allelic loss at multiple chromosomal locations has been identified in a range of invasive and preinvasive breast cancers as well as benign and normal breast epithelium adjacent to tumor.[1–3] However, a complete evaluation of LOH in breast cancer has been hampered by the limited number of polymorphic markers available for study; the heterogeneity of breast tissue (mixed nontumor and tumor cells); the lack of sufficient numbers of fresh or frozen samples with associated demographic or clinical data, and the small amount of tissue available from currently diagnosed breast cancers. To address these limitations, we used flow cytometry to select and purify tumor cells from routinely processed tissue blocks, whole genome amplification to increase the amount of DNA available for study, and a microarray assay to assess all chromosomes efficiently and simultaneously.

The newly developed Affymetrix HuSNP array, which contains 1494 single nucleotide polymorphism (SNP)

Supported by NIH/NCI grants RO1 CA78855 (to P. S. R.) and RO1 CA71735 (to P. L. P.).

Accepted for publication September 25, 2001.

Address reprint requests to Dr. Peggy L. Porter, Department of Pathology, Box 357705, University of Washington, 1959 NW Pacific St., Seattle, WA 98195-7705. E-mail: pporter@fhcrc.org.

sites genome-wide and requires only 135 ng of genomic DNA (gDNA) per assay, is a potential platform for evaluating genome-wide genetic analysis of breast tissue. The usefulness of a prototype SNP array and the current HuSNP array for analysis of allelic loss in fresh lung tumors removed at autopsy and fresh biopsies from esophageal cancers, respectively, has been previously described.[4,5] However, the analysis of formalin-fixed, paraffin-embedded pathology specimens by the commercially available HuSNP assay has not been reported.

Here we discuss the use of the HuSNP to examine allelic imbalance in both frozen and fixed pathology specimens and compare results between the two preservation methods. To purify populations of cells from the tissue for analysis we used bivariate flow cytometry, which allowed us to sort tumor cells for analysis based on positive cytokeratin staining and gDNA content.[6] In addition to gDNA, we also examined the use of a polymerase chain reaction (PCR)-based whole-genome amplification method, primer-extension preamplification (PEP) that increases the amount of template available for analysis ~30-fold[7,8] and compared allelic loss results from the PEP product to results with gDNA. HuSNP allelic loss results were also compared to results from conventional polymorphic microsatellite markers (short tandem repeats or STRs) on chromosomes 11 and 17.

Materials and Methods

Tissue Samples

Tumor and normal tissue from two breast cancer patients were obtained from the University of Washington tissue bank with patient consent and in compliance with the Institutional Review Board. Samples taken at the time of surgery were divided into two portions and each portion was processed routinely either by freezing in OCT media or formalin fixation followed by paraffin embedding. No gross difference was apparent between the portions selected for either preservation method. The presence of tumor in each block was confirmed microscopically. A formalin-fixed tissue block from each of the cases was tested for estrogen receptor, progesterone receptor, c-erbB2 oncogene protein, and p53 tumor suppressor gene protein by immunohistochemistry as previously described.[9]

Flow Cytometry

Flow cytometry was performed on the frozen and fixed samples to purify tumor cells from normal epithelia and nonepithelial cells. Hematoxylin and eosin (H&E)-stained slides from both frozen and paraffin-embedded tumor sections were examined to confirm that the samples contained tumor epithelium. Similarly, H&E slides taken from the normal block confirmed that the sample contained no tumor.

From each frozen breast tissue sample, 20 to 50 50-μm sections were cut and placed into phosphate-buffered saline containing 1% bovine serum albumin (PBA).[10]

The samples were mechanically disaggregated and washed in PBA. The resulting cell suspensions were fixed in 0.5% electron microscopy grade formaldehyde and permeabilized in 0.1% triton/PBA before staining. From formalin-fixed tissue blocks, flow cytometry preparation was performed as described.[6] Briefly, 1 to 20 60-μm sections were cut from normal and tumor tissue blocks; regions of tumor in each section were dissected from surrounding tissue with a scalpel blade. All sections were deparaffinized, rehydrated, and digested in collagenase before a brief pepsin digestion.

Cell suspensions from both the frozen and fixed samples were stained with 4,6-diamidino-2-phenylindole and R-phyco-erythrin labeled AE1/AE3 (Roche, Indianapolis, IN), which recognizes a wide variety of acidic and basic cytokeratins. A parallel sample of cells was stained with R-PE-labeled isotype-matched mouse Ig (R-PE labeled IgG1; DAKO, Carpinteria, CA) and used as a negative control. Before sorting, all samples were forced through a 25-gauge needle10 times to ensure a single cell suspension.

Cytokeratin-positive tumor cells were sorted by bivariate analysis with 488 nm and UV excitation on a Becton Dickinson (Mountain View, CA) FACS Vantage. R-PE, cytokeratin-positive populations were sorted based on their 4,6-diamidino-2-phenylindole-fluorescent DNA content, expressed as DNA index (DI = mean aneuploid G_1 fluorescence/mean diploid G_1 fluorescence). Cells from the normal blocks were processed and stained similarly to the tumor samples. The DNA from all cells in the normal blocks was used as the constitutive normal for comparison with the tumor cell DNA.

Preparation of DNA Samples

DNA was extracted from frozen cells using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN), following the manufacturer's suggestions with the addition of 1 μl of 20 mg/ml of Proteinase K to the cell lysis buffer, followed by incubation at 50°C for 1 to 16 hours. DNA was extracted from fixed cells using a simple Proteinase K digestion method previously described.[11] Extracted DNA samples were quantified using the Picogreen dsDNA Quantitation Kit (Molecular Probes, Eugene, OR) on the Cytofluor II Fluorescence Multiwell Plate Reader (PerSeptive Biosystem Inc., Framingham, MA).

Whole genome amplification using the primer extension protocol (PEP) was performed as described.[8] For each gDNA sample, six individual PEP reactions, each using 7 ng of gDNA as template, were performed and the PEP material pooled.[12] PEP material was used directly in the array protocol without purification or alteration of concentration.

STR Protocol

Twenty-four polymorphic repeat loci (STRs) on chromosomes 11 and 17 were amplified using fluorescent prim-

ers with PEP template. Chromosomes 11 and 17 were selected for allelic loss comparison between arrays and conventional repeat markers because both contain sites that are frequently lost in breast cancer.[2,3] Markers were selected from those commercially available from Research Genetics (*www.resgen.com*) to obtain a survey of sites that corresponded as closely as possible to the HuSNP sites along the chromosomes. The physical locations of the markers in Mb are listed in Figure 1, B and C, as given by National Center for Biotechnology Information in July 2001 (*www.ncbi.nlm.nih.gov*). Primers for chromosome 11 were (11ptel) D11S1397, D11S2368, D11S2001, D11S1918, D11S1395, (cen), D11S4076, D11S1394, D11S4151, D11S2360, (11qtel). Those for chromosome 17 were (17ptel), D17S919, D17S1298, D17S1537, TP53, D17S786, D17S1541, D17S974, D17S975, (cen), D17S1293, D17S1158, D17S1294, D17S1185, D17S1305, D17S1290, D17S1288, (17qtel). PCR reactions were performed using standard protocols with PEP material as a template. PCR reaction products were multiplexed and then purified using Microcon-100 columns, after which the DNA was resuspended in sterile water. Reactions were run on an ABI 377 and analyzed using ABI Prism Gene Scan software.

To evaluate allelic loss for each marker[12,13] the peak height of the first allele was divided by the peak height of the second allele to obtain the allelic ratio (AR). Samples were deemed informative at a locus if the AR for the normal tissue sample was sufficiently close to 1 (defined operationally as $0.7 < AR < 1/0.7$). For informative loci, an index Q was computed as the AR of the tumor tissue sample divided by the AR of the normal tissue sample.[14] A locus was scored as having LOH if the Q value was sufficiently far from 1 (defined operationally as either $Q < 0.3$ or $Q > 1/0.3$). A locus was scored as retaining heterozygosity if the Q value was sufficiently close to 1 (defined operationally as $0.7 < Q < 1/0.7$).

HuSNP Protocol

The Affymetrix HuSNP protocol was performed according to manufacturer's instructions and as described.[4] Each individual gDNA sample from both cases was analyzed twice in completely separate reactions, to yield data from a total of 18 HuSNP arrays for the two cases (five samples from case 1 in duplicate plus four samples from case 2 in duplicate). Similarly, the PEP material from each sample was analyzed by HuSNP in duplicate (18 HuSNP arrays). Data analysis using the Affymetrix Genechip software resulted in genotype calls that were used in the statistical analysis. The genetic map used in the analysis came from Affymetrix, release date June 2001.

Statistical Methods

To quantify the reproducibility of the HuSNP chips, the reliability measure was calculated. The reproducibility for making a consistent genotype call was defined as the number of SNPs with the same genotype calls from both replicates divided by the total number of SNPs for which both replicates yielded signal calls. The reproducibility for making no-signal calls was also calculated, and was defined as the number of SNPs for which both replicates yielded no-signal calls divided by the total number of SNPs for which at least one replicate yielded a no-signal call.

Similarly, concordance of genotype and no-signal calls were measured between frozen and fixed tissue samples as well as between gDNA and PEP samples. Because each sample was analyzed in duplicate, there were a total of four possible comparisons between each set of fixed and frozen samples. The concordance measure was calculated by the ratio of the average over the four comparisons of the number of SNPs with same genotype calls from both samples and the average over the four comparisons of the number of SNPs for which both samples yielded signal calls. The concordance measure for no-signal calls was calculated similarly.

The informativity and allelic loss of the SNPs was examined for both cases. We defined a SNP as informative when one normal tissue replicate of the SNP was heterozygous (AB) and the other replicate was either heterozygous or had no signal. We defined a SNP site as having allelic loss when that SNP was informative in the normal tissue, one tumor tissue replicate of the SNP was hemizygous or homozygous (AA or BB), and the other replicate was either hemizygous, homozygous, or no signal.

All statistical analyses were performed using SPLUS statistical software (S-PLUS Reference Manual, version 3:2; Statistical Sciences I, Seattle, Washington).

Results

Subject and Tumor Characteristics

The results of the pathology review and immunohistochemical assays from the two cases used in this study are shown in Table 1. The patients' ages were similar at their respective times of diagnosis. Flow cytometric analysis revealed multiple aneuploid cell populations in the tumor from case 1. One cell population from the fixed tumor and one from the frozen tumor had very similar DIs (1.49 and 1.43, respectively), whereas an additional cell population seen only in the fixed portion of the tumor had a distinct DI of 1.82. Case 2 had a single aneuploid tumor

Table 1. Patient and Tumor Characteristics of the Two Breast Cancer Cases

	Case 1	Case 2
Histologic type	Lobular	Ductal
Age at diagnosis	48	46
AJCC stage	IIIA	IIA
ER protein status	Negative	Positive
PR protein status	Negative	Positive
c-erb-2 protein status	Negative	Positive
P53 protein status	Positive	Negative
DNA index	Multiple aneuploid	Aneuploid
DNA index—frozen tumor	1.43	1.79
DNA index—fixed tumor	1.49 and 1.82	1.76

ER, estrogen receptor; PR, progesterone receptor.

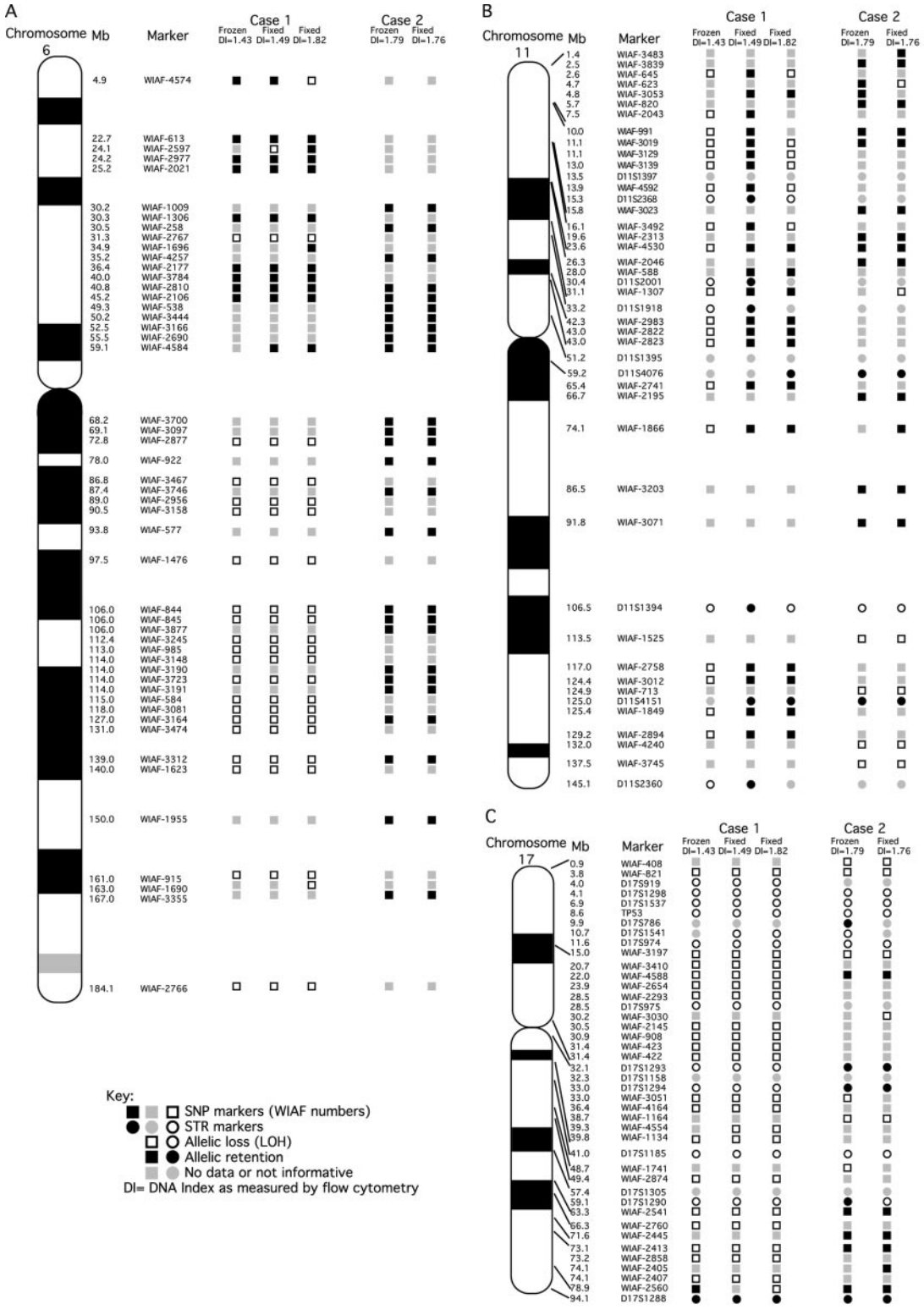

Figure 1. Graphical representation of HuSNP and STR allelic loss data for chromosomes 6 (**A**), 11 (**B**), and 17 (**C**) from both patients. SNP data are depicted as **squares** and STR data are indicated by **circles**. Shown are the data from all five cell populations isolated by bivariate flow cytometry in the two patients' tumors. SNP loci that were uninformative in both the frozen and fixed material from a patient were removed from this diagram. **White squares** and **circles** indicate allelic loss, **black square** and **circles** indicate retention, and **gray squares** and **circles** indicate lack of informativity or data (in the STR markers) or no signal calls (in the HuSNP). Markers are arranged on the diagram to represent their physical distribution on the chromosome.

Table 2. Reproducibility of HuSNP Calls for Each Sample Type*

	Frozen normal		Frozen tumor		Fixed normal		Fixed tumor	
	Definite calls†	No signal calls	Definite calls	No signal calls	Definite calls	No signal calls	Definite calls	No signal calls
gDNA	98%	72%	96%	65%	98%	75%	95%	75%
PEP	97%	75%	95%	67%	95%	70%	93%	79%

*The average percent of definite (AA, AB, and BB) and no signal HuSNP genotype calls that were identical between independent, duplicate analyses of each sample type.
†AA, AB, or BB calls.

cell population distinguishable in both the fixed (DI = 1.76) and frozen (DI = 1.79) tissue samples. All three tumor cell populations from case 1 and both from case 2 were tested independently and included in the subsequent array analysis using both gDNA and PEP material from these cases.

HuSNP Analysis

For each SNP site on the chip, genotype results from the Affymetrix Genechip software were reported as definite calls (AA, AB, BB), no signal, or an intermediate call (AB_A or AB_B). The Genechip software does not score allele copy number but instead always indicates two alleles (AA and BB). The HuSNP chip contains 1494 individual SNP sites, however our experience was similar to that of a previous report,[4] in that more than 100 of the 1494 sites on the chip consistently failed, yielding most of the no-signal calls. Intermediate calls were rare, seen in ~1% of sites in each assay. The reproducibility statistics for definite calls and no-signal calls between duplicate assays using the same gDNA or PEP sample are shown in Table 2. Table 3, A and B, show the concordance of definite and no-signal results between the fixed and frozen gDNA and PEP samples from each case (Table 3A) as well as the concordance between the gDNA and PEP results (Table 3B). A graphical representation of concordance between fixed and frozen gDNA samples on chromosomes 6, 11, and 17 are presented in Figure 1.

The allelic loss results for each chromosome and gDNA sample source are shown in Table 4. PEP results were similar to the gDNA, as indicated by the genotype concordances shown in Table 3B. Informativity varied slightly between the fixed and frozen samples and resulted in some differences in allelic loss results between the paired samples as shown in Table 4. This variance was primarily because of no-signal calls at a particular SNP site in one sample type or the other and not to actual differences in calls between the fixed and frozen samples. The exception to this is in case 1 that contained more than one aneuploid population of tumor cells (Table 1). Allelic losses were reproducible in the duplicate analyses and generally concordant with adjacent sites in large regions along chromosomes, as is shown visually for chromosomes 6, 11, and 17 in Figure 1.

Comparison of HuSNP and STR Analyses

Of the 24 STR markers analyzed on chromosomes 11 and 17, there were a total of 69 informative sites between all five tumor populations identified in the two patients (see Figure 1, B and C, for details). Of these 69 sites, 60 showed correlation with data from adjacent HuSNP markers. However, at the nine STR sites that do not correlate

Table 3A. The Concordance‡ of Definite and No Signal Calls between the Duplicate Analyses of Frozen and Fixed Samples from Each of Two Breast Cancer Cases

	Case 1				Case 2			
	Normal tissue		Tumor tissue		Normal tissue		Tumor tissue	
	Definite calls*	No signal calls	Definite calls	No signal calls	Definite calls	No signal calls	Definite calls	No signal calls
gDNA	1162 (97%)‡	185 (64%)	1071 (92%)	199 (61%)	1150 (93%)	182 (71%)	1019 (95%)	189 (45%)
PEP	878 (92%)	543 (36%)	766 (87%)	288 (47%)	1131 (93%)	185 (68%)	919 (95%)	196 (37%)

Table 3B. Concordance† of gDNA and PEP DNA Data for Each Sample Type in Both Cases

	Frozen normal		Frozen tumor		Fixed normal		Fixed tumor	
	Definite calls*	No signal calls	Definite calls	No signal calls	Definite calls	No signal calls	Definite calls	No signal calls
Case 1	1004 (95%)	211 (49%)	1006 (92%)	219 (54%)	909 (91%)	188 (38%)	1200 (98%)	207 (78%)
Case 2	1183 (98%)	162 (57%)	1166 (97%)	167 (58%)	1166 (97%)	167 (58%)	825 (93%)	263 (43%)

*AA, AB, or BB calls.
†The concordance of genotype calls between frozen and fixed material from each of the cases. The numbers are the number of sites average over replicates that gave concordant genotypes out of the 1494 sites on the HuSNP array.
‡Percentages are the average percent of concordant sites out of those with that type of call. See text for a description of how concordance was calculated.

Table 4. Number of Informative SNPs and LOH by Chromosome, Identified in the Fixed and Frozen Samples from Two Cases of Breast Cancer

	Case 1			Case 2	
		Fixed tissue			
Chromosome	Frozen tissue LOH†/informative* (DI = 1.43)	LOH/informative (DI = 1.49)	LOH/informative (DI = 1.82)	Frozen tissue LOH/informative (DI = 1.79)	Fixed tissue LOH/informative (DI = 1.76)
1	1/27	0/26	2/26	0/32	0/28
2	8/26	10/25	9/25	1/29	1/23
3	12/26	11/28	13/28	0/36	0/25
4	11/16	3/16	12/16	0/22	0/15
5	1/17	1/17	1/17	0/14	0/10
6	20/33	19/35	21/35	0/27	1/26
7	1/16	1/16	2/16	0/21	0/19
8	0/33	0/33	1/33	18/36	15/27
9	5/22	5/22	4/22	0/17	0/15
10	7/10	8/12	9/12	0/20	1/15
11	17/20	0/21	6/21	5/23	4/22
12	2/18	1/18	3/18	0/6	0/6
13	0/14	0/11	1/11	0/8	0/5
14	0/9	0/9	0/9	1/13	1/10
15	0/16	0/17	1/17	0/16	0/14
16	0/11	0/12	1/12	6/11	6/8
17	19/20	18/21	21/21	6/15	6/13
18	2/13	3/13	4/13	7/10	4/7
19	0/16	0/17	1/17	7/17	1/11
20	0/5	0/6	0/6	0/9	0/6
21	0/6	0/6	0/6	0/8	0/5
22	0/6	0/8	1/8	5/5	5/5
X	0/5	0/6	5/6	0/7	0/7
Unmapped	0/2	1/3	1/3	0/4	0/2
Total	106/387 (27%)	81/398 (20%)	119/398 (30%)	56/406 (14%)	45/324 (14%)

*SNP sites where the normal sample genotype call was AB (see Materials and Methods text).
†SNP sites which were informative in the normal and had either AA or BB genotype calls in the tumor sample (see Materials and Methods text).

with adjacent HuSNP markers, it is difficult to determine whether the apparent discordance is because of technical limitations or if the STR marker is recognizing a small region with a different allelic loss pattern than the adjacent regions scored by SNP. At four of the nine sites (D11S1394 and D17S1288 in case 1, and D17S1294 in case 2), there was at least a 5-Mb distance between the STR and SNP markers, which may be the reason for the discrepancy.

Discussion

This study examined the feasibility of using array technology, specifically the commercially available Affymetrix HuSNP array, for genome-wide allelic loss analysis of both fixed and frozen breast pathology specimens. Archival pathology specimens are a valuable resource for the genetic analysis of tumors. However, the limited quantity and quality of DNA available is a serious limitation for genetic analysis of such specimens. The quality of DNA obtained from pathology specimens is compromised by routine preservation methods that were neither designed for, nor are optimal for, DNA preservation. Formalin, the most commonly used fixative for pathology tissue specimens, has been shown to reduce the size of PCR segments that may be amplified from a sample.[15] In our experience as well as in reports from the literature, DNA extracted from paraffin-embedded tissues most reliably yields PCR results in small amplicons, often under 200 nucleotides.[15,16] Tissues frozen in OCT media for frozen section diagnosis suffer less of a direct insult to DNA quality but are still subject to handling and storage exposures that may result in DNA fragmentation. Ideally, a genomic analysis technique for pathology specimens would maximize the data obtained from nanogram quantities of low-molecular weight DNA. Our study sought to validate array technology such as that used in the HuSNP array for use with such specimens.

In this study, samples were analyzed in duplicate to generate reliability statistics for each type of sample, and genotype data were compared between fixed and frozen samples to examine the data concordance between sample types. The HuSNP array yielded genotype results that were reliable and concordant for both fixed and frozen tumor and normal breast pathology specimens. Importantly, the DNA fragmentation that occurs with formalin fixation does not seem to affect HuSNP results, presumably because the assay relies on PCR amplicons that are shorter than 100 nucleotides in length.

In addition to analysis of genomic DNA extracted from these specimens, we also examined the data obtained from whole genome amplified material (PEP) generated from our specimens, and found similar reliability for either genomic DNA and PEP genotypes when analyzed by HuSNP. The concordance was similarly high for both genomic and PEP DNA, although slightly lower for the PEP material. This lower concordance was primarily because of an increase in no-signal genotype calls seen in the PEP material *versus* the genomic DNA and an indica-

tor of the potential data not obtained with amplified DNA. However, in cases in which sample is limited, using the HuSNP assay on PEP material may be an acceptable approach to genome-wide analysis. In cases in which the original sample is extremely limited, the use of whole genome amplification may make analysis possible.

We also used the data generated from the two cases to examine allelic loss in the cell populations isolated from the tumors by bivariate flow cytometry. There was an average of 379 informative SNP sites throughout the genome from all of the gDNA HuSNP assays. This is very similar to the expected distribution of heterozygosity as defined using biallelic SNP markers[17,18] and in previously reported HuSNP data.[4] As has been previously reported,[4] allelic loss data obtained from the HuSNP agreed well with data obtained by the more standard method of microsatellite (STR) analysis.

Although the fixed and frozen samples from case 2 yielded highly concordant HuSNP results on all chromosomes, the cell populations with close DI (1.43 and 1.49, respectively) identified in case 1 exhibited substantial differences in LOH on chromosomes 4 and 11. Because the differences were confined to these two chromosomes, the data were not likely to be the result of a general cross-contamination, but rather reflected a biological difference between these cell populations. The second population with a DI of 1.82, identified in the formalin-fixed tissue block was also distinct and exhibited a slightly higher frequency of allelic loss throughout the genome. A diversity of cell populations is common within advanced breast tumors[19,20] and may reflect the development of distinct genotypic clones with different behavior potential. Flow cytometric analysis can initially define cell populations with DNA content differences that can be further resolved by genomic analysis, yielding important information about tumor composition that would otherwise be obscure.

The gDNA HuSNP analysis of the two cases included in this study yielded more genome-wide data than could be obtained with a similar amount of DNA by other means, such as microsatellite marker analysis. However, it is still a low-density map, with an average of one SNP site per 8.5 Mb in the genome. Another limitation of the current HuSNP array is that many SNP sites included in the assay are clustered, so that many regions of the genome are well represented whereas others are underrepresented. Given that array assays for genome-wide analyses are continuing to be developed, we expect that the next generation of genetic marker arrays using similar technology as the HuSNP will provide more uniform and higher density coverage of the genome. The data from this study indicate that future array technologies will be suitable for use with DNA obtained from routinely processed pathology specimens.

Acknowledgments

We thank Stephanie Stafford for assistance with data management and analysis, Carlo Maley for assistance in genetic map correlation, Patty Galipeau and Michael Barrett for help in the planning of the experiments included in this study, Cassandra Neal and Jeff Delrow of the Fred Hutchinson Cancer Research Center DNA Array Facility for assistance in the hybridization and analysis of the HuSNP arrays, and Cintia De Barros for assistance with the flow sorting.

References

1. Deng G, Lu Y, Zlotnikov G, Thor AD, Smith HS: Loss of heterozygosity in normal tissue adjacent to breast carcinomas. Science 1996, 274: 2059–2057
2. Devilee P, Cornelisse CJ: Somatic genetic changes in human breast cancer. Biochim Biophys Acta 1994, 1198:113–130
3. Osborne R, Hamshere M: A genome-wide map showing common regions of loss of heterozygosity/allelic imbalance in breast cancer. Cancer Res 2000, 60:3706–3712
4. Lindblad-Toh K, Tanenbaum D, Daly M, Winchester E, Lui W, Villapakkam A, Stanton S, Larsson C, Hudson T, Johnson B, Lander E, Meyerson M: Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. Nat Biotechnol 2000, 18:1001–1005
5. Mei R, Galipeau P, Prass C, Berno A, Ghandour G, Patil N, Wolff R, Chee M, Reid B, Lockhart DJ: Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. Genome Res 2000, 10:1126–1137
6. Glogovac J, Porter P, Banker D, Rabinovich P: Cytokeratin labeling of breast cancer cells extracted from paraffin embedded tissue for bivariate flow cytometric analysis. Cytometry 1996, 24:260–267
7. Barrett MT, Reid BJ, Joslyn G: Genotypic analysis of multiple loci in somatic cells by whole genome amplification. Nucleic Acids Res 1995, 23:3488–3492
8. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N: Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci USA 1992, 89:5847–5851
9. Porter PL, El-Bastawissi AY, Mandelson MT, Lin MG, Khalid N, Watney EA, Cousens L, White D, Taplin S, White E: Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 1999, 91:2020–2028
10. Bauer KD, Jacobberger JW: Analysis of intracellular proteins. Methods Cell Biol 1994, 41:351–376
11. Frank T, Svoboda-Newman S, Hsi E: Comparison of methods for extracting DNA from formalin-fixed paraffin sections for nonisotopic PCR. Diagn Mol Pathol 1996, 5:220–224
12. Paulson T, Galipeau P, Reid B: Loss of heterozygosity analysis using whole genome amplification, cell sorting and fluorescence-based PCR. Genome Res 1999, 9:482–491
13. Galipeau P, Cowan D, Sanchez C, Barrett M, Emond M, Levine D, Rabinovitch P, Reid B: 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. Proc Natl Acad Sci USA 1996, 93:7081–7084
14. Hahn M, Matzen S, Serth J, Pingoud A: Semiautomated quantitative detection of loss of heterozygosity in the tumor suppressor gene p53. BioTechniques 1995, 18:1040–1047
15. Greer C, Lund J, Manos M: PCR amplification from paraffin-embedded tissues: recommendations on fixatives for long-term storage and prospective studies. PCR Methods and Applications 1991, 1:46–50
16. Ben-Ezra J, Johnson D, Rossi J, Cook N, Wu A: Effect of fixation on the amplification of nucleic acids from paraffin-embedded material by the polymerase chain reaction. J Histochem Cytochem 1991, 39:351–354
17. Kwok P-Y, Deng Q, Zakeri H, Taylor S, Nickerson D: Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics 1996, 31:123–126
18. Kruglyak L: The use of a genetic map of biallelic markers in linkage studies. Nat Genet 1997, 17:21–24
19. Bonsing B, Corver W, Fleuren G, Cleton-Jansen, A-M, Devilee P, Cornelisse C: Allelotype analysis of flow-sorted breast cancer cells demonstrates genetically related diploid and aneuploid subpopulations in primary tumors and lymph node metastasis. Genes Chromosom Cancer 2000, 28:173–183
20. Fujii H, Marsh C, Cairns P, Sidransky D, Gabrielson E: Genetic divergence in the clonal evolution of breast cancer. Cancer Res 1996, 56:1493–1497