# Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays

**Kerstin Lindblad-Toh[1]*, David M. Tanenbaum[2–4]*, Mark J. Daly[1], Ellen Winchester[1], Weng-Onn Lui[5], Anuradha Villapakkam[1], Sasha E. Stanton[2], Catharina Larsson[5], Thomas J. Hudson[1,6], Bruce E. Johnson[2,3], Eric S. Lander[1,7] and Matthew Meyerson[2,4]**

*[1]Whitehead Institute/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, Cambridge, MA 02139. [2]Department of Adult Oncology, Dana-Farber Cancer Institute, Boston, MA. [3]Departments of Medicine and [4]Pathology, Harvard Medical School, Boston, MA. [5]Department of Molecular Medicine, CMM, Karolinska Hospital, Stockholm, S-171 76 Sweden. [6]Montreal Genome Centre, McGill University Health Centre, Montréal, Québec. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02142. *These two authors contributed equally to this work. Corresponding authors (lander@genome.wi.mit.edu) and (matthew_meyerson@dfci.harvard.edu).*

Received 3 May 2000; accepted 26 June 2000

**Human cancers arise by a combination of discrete mutations and chromosomal alterations. Loss of heterozygosity (LOH) of chromosomal regions bearing mutated tumor suppressor genes is a key event in the evolution of epithelial and mesenchymal tumors[1]. Global patterns of LOH can be understood through allelotyping of tumors with polymorphic genetic markers[2]. Simple sequence length polymorphisms (SSLPs, or microsatellites) are reliable genetic markers for studying LOH[3], but only a modest number of SSLPs are used in LOH studies because the genotyping procedure is rather tedious. Here, we report the use of a highly parallel approach to genotype large numbers of single-nucleotide polymorphisms (SNPs) for LOH, in which samples are genotyped for nearly 1,500 loci by performing 24 polymerase chain reactions (PCR), pooling the resulting amplification products and hybridizing the mixture to a high-density oligonucleotide array[4]. We characterize the results of LOH analyses on human small-cell lung cancer (SCLC) and control DNA samples by hybridization. We show that the patterns of LOH are consistent with those obtained by analysis with both SSLPs[5] and comparative genomic hybridization (CGH), whereas amplifications rarely are detected by the SNP array. The results validate the use of SNP array hybridization for tumor studies.**

To test the potential of SNP array hybridization for the detection of genome-wide LOH in human tumors, we measured allelic loss and retention patterns for 17 matched pairs of human SCLC and normal control DNA samples[5] using the GeneChip HuSNP Mapping Assay (HuSNP chip), which contains detectors for 1,494 SNP loci. Each DNA sample is subjected to 24 multiplex PCR reactions (each involving a mixture of T3/T7-tailed primer pairs representing 50–100 loci, followed by 24 multiplex reactions with biotinylated-T3/T7 primers). The resulting products are pooled, hybridized to the SNP array, stained with streptavidin–phycoerythrin, and assayed by fluorescence detection. The principles underlying genotyping with SNP arrays were described in an earlier study[4]. Briefly, the detector for each SNP locus contains four rows of 25-mer oligonucleotides, two of which contain oligonucleotides that perfectly match either SNP allele A or SNP allele B, whereas the other two contain single-base mismatches at various positions. The allelotype at a locus is determined by fluorescence intensity ratios in an automated fashion. The approach dramatically decreases the work involved in assaying 1,500 loci, as well as the amount of DNA required (to a total of only 120 ng DNA, corresponding to ~20,000 diploid human genomes) in comparison to both SSLPs and CGH.

The call rate (the proportion of loci to which genotypes could be assigned) was 80.7% ± 3.0% over all samples, yielding ~1,205 SNPs scored per sample (Table 1). The rate did not differ between normal and tumor samples. Many SNPs performed in a robust fashion,

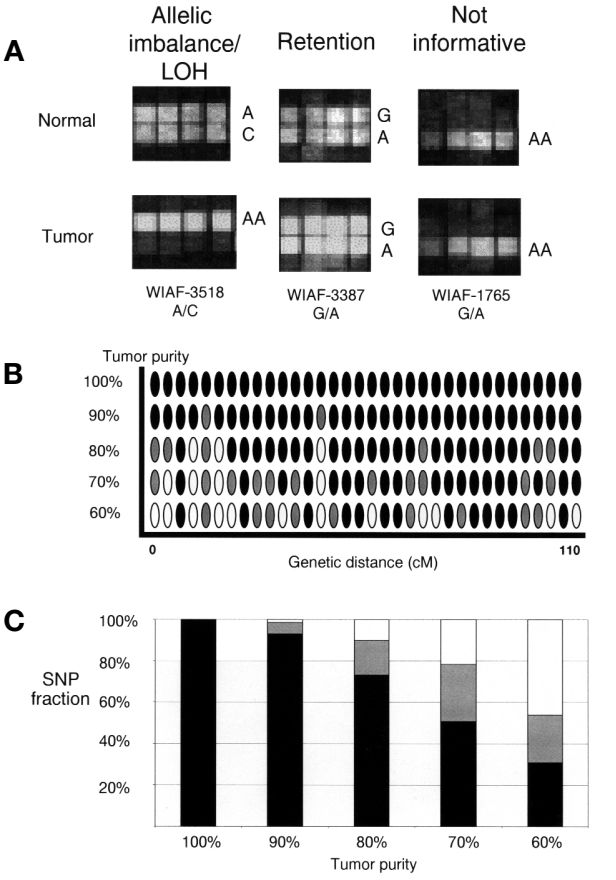


**Figure 1. (A) Representative images of fluorescence intensities for SNP array hybridization to normal and tumor DNA samples. The intensity fraction of allele A/(A+B) was used to determine the genotype based on predetermined values for each SNP (see the algorithms described in Experimental Protocol). For each sample pair, a call of LOH, retention, or uninformative was made. WIAF-3518 is a SNP marker on chromosome 9, and WIAF-3387 and WIAF-1765 are SNPs on chromosome 3. (B) Tumor mixing experiment. SNPs on chromosome 3p showing LOH in a pure tumor cell line DNA (100%) were assayed in mixtures with corresponding normal DNA (90% to 60% tumor content). As the percentage of tumor DNA content drops, SNPs showing LOH progress from LOH to uncertain to false retention. Black denotes LOH, gray uncertain, and white retention of heterozygosity. (C) Effect of tumor DNA content on accuracy of "true LOH" calls, considering all 130 informative samples in the tumor mixing experiment. As in (B), black denotes LOH, gray uncertain, and white retention of heterozygosity. Note that samples with 80% tumor content or more give virtually the same result as a 100% tumor sample.**

**Table 1. HuSNP chip performance**

| | Sample no. | | | | | | | | | | | | | | | | | Mean | s.d. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **All loci**[a] | 1 | 3 | 9 | 10 | 11 | 12 | 13 | 18 | 19 | 20 | 23 | 24 | 27 | 28 | 29 | 31 | 40 | | |
| Normal | 82 | 77 | 82 | 87 | 80 | 85 | 74 | 82 | 81 | 85 | 83 | 80 | 77 | 76 | 79 | 81 | 78 | 81 | 3.5 |
| Tumor 1st replicate | 83 | 79 | 80 | 77 | 78 | 81 | 79 | 85 | 82 | 84 | 80 | 81 | 75 | 82 | 83 | 81 | 81 | 81 | 2.5 |
| Tumor 2nd replicate | 78 | 86 | 79 | 85 | 80 | 83 | 80 | 85 | 83 | 85 | 76 | 82 | 78 | 77 | 77 | 81 | 80 | 81 | 3.2 |
| Both tumors | 52 | 75 | 74 | 73 | 60 | 77 | 71 | 81 | 77 | 79 | 69 | 75 | 69 | 71 | 72 | 77 | 74 | 72 | 7.0 |
| Discordant tumors | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.3 | 0.0 | 0.2 | 0.1 | 0.1 | 0.9 | 0.3 | 0.1 | 0.0 | 0.1 | 0.0 | 0.2 | 0.5 |
| **Mapped loci** | | | | | | | | | | | | | | | | | | | |
| SNPs heterozygote in normal | 299 | 280 | 283 | 337 | 212 | 308 | 266 | 302 | 281 | 295 | 300 | 255 | 242 | 273 | 252 | 275 | 262 | 278 | 29.0 |
| Informative SNPs[b] | 279 | 274 | 273 | 320 | 200 | 294 | 250 | 288 | 259 | 281 | 265 | 239 | 224 | 264 | 237 | 266 | 243 | 262 | 28.5 |
| Fraction not informative (%) | 6.7 | 2.1 | 3.5 | 5.0 | 5.7 | 4.5 | 6.0 | 4.6 | 7.8 | 4.7 | 11.7 | 6.3 | 7.4 | 3.3 | 6.0 | 3.3 | 7.3 | 5.6 | 2.2 |
| SNPs with allelic imbalance[c] | 83 | 56 | 106 | 89 | 82 | 16 | 72 | 81 | 77 | 168 | 52 | 65 | 73 | 57 | 84 | 75 | 90 | 78 | 30.5 |
| SNPs with retention | 196 | 218 | 167 | 231 | 118 | 278 | 178 | 207 | 182 | 113 | 213 | 174 | 151 | 207 | 153 | 191 | 153 | 184 | 41.2 |
| Fraction with allelic imbalance (%) | 29.7 | 20.4 | 38.8 | 27.8 | 41.0 | 5.4 | 28.8 | 28.1 | 29.7 | 59.8 | 19.6 | 27.2 | 32.6 | 21.6 | 35.4 | 28.2 | 37.0 | 30.1 | 11.4 |
| Allelic imbalance with retentions (%)[d] | 2.4 | 3.6 | 2.3 | 5.0 | 3.2 | 0.0 | 0.0 | 9.1 | 13.5 | 9.8 | 14.3 | 0.0 | 12.0 | 5.3 | 7.3 | 5.9 | 3.3 | 5.7 | 4.6 |

[a]Numbers are percentage calls.
[b]SNPs heterozygous in the normal and scored in one or both tumor replicates.
[c]Allelic imbalance: primarily LOH, but possibly amplification.
[d]Occurrences of retention within a region of allelic imbalance (sum of LLRLL/sum of LLLLL). Could be a result of small interstitial deletions, sample impurity, chromosomal gain, or chip error.

whereas a subset performed poorly: 45% of the SNPs were assigned a genotype in at least 98% of samples, whereas 13% of the SNPs failed in the majority of samples. Moreover, the consistency of calls was high between samples analyzed in duplicate (reproducibility = 99.8% ± 0.5%). The genotyping accuracy of the chip calls was estimated at 95.4% on the basis of validation of random SNPs (n = 216 genotypes) in normal and tumor samples by gel-based length multiplex single-base extension (LM-SBE)[6].

For each tumor-normal pair, each SNP locus was scored as "not informative" (if the normal sample was homozygous at the locus), "Allelic imbalance/LOH" (which refers to allelic imbalance due to either true loss or extreme amplification of an allele), "retention" (of heterozygosity), or "uncertain" (if the result fell in an intermediate range or had missing data). Examples are shown in Figure 1A.

Analysis of tumors for LOH can be confounded by the presence of contaminating DNA from normal surrounding stromal tissue[7]. We performed mixing experiments, to assess the level of tumor sample purity required for accurate allelotyping with SNP arrays. DNA from a pure tumor cell line was mixed with matched normal DNA, to obtain mixtures consisting of 60, 70, 80, 90, and 100% tumor cell line DNA.

For SNP loci showing loss of heterozygosity in the pure tumor cell line, we studied how the allelotype scoring changed to "uncertain" and then "retention of heterozygosity" with increasing amounts of contaminating normal DNA (Fig. 1B, C). Samples with 90% tumor purity gave essentially identical results to those with 100%, whereas decreasing the purity to 80% resulted in an increase in "uncertain" calls and a few false positive "retentions". Accuracy drops off steeply for purity of 70% or less, because the lost allele composes 15% or more of the contaminated sample. To estimate the purity of the SCLC samples studied here, we performed DNA ploidy analysis on nuclei from 9 of the 17 tumor samples. The median proportion of tumor cells was estimated at 92% for 7 samples having aneuploid populations, with 2 samples estimated below 90%. Although tumor purity is dependent on tumor type, a purity of 80% can often be achieved using gross dissection or microdissection.

The average proportion of informative markers (heterozygous in the normal sample) out of callable markers was 31%, which agrees well with the heterozygosity rate for the SNPs used on the array. This is a considerably lower heterozygosity rate than for SSLPs (typically 70%) but could be readily increased to closer to 50% by selecting SNPs with higher heterozygosity (which are rapidly becoming available). Given the variation in SNP allele frequencies among diverse human populations, larger numbers of SNPs will eventually be required for an even representation in all populations.

Of the informative loci (including both mapped and not yet mapped markers), it was possible to call loss or retention in 94% of cases (with the remaining 6% being uncertain). This yielded a total of 354 such SNPs, corresponding to ~1 per 8.5 Mb. By comparison, the highest density analysis of whole-genome LOH reported for SCLC, is at an average density of 1 informative marker per 34 Mb[5].

At present, 1,081 of the SNPs have been assigned to specific chromosomal locations in the human genome by virtue of radiation hybrid (RH) mapping. (The rest should soon be localized with the rapidly advancing sequencing of the human genome.) For these loci, an allelotype map was generated for each SCLC tumor sample. A statistical algorithm was used to calculate the likely windows of LOH for each chromosome (see Experimental protocol). Maps of chromosomes 3 and 20 are shown in Figure 2.

The highest frequency of LOH was seen on chromosome arm 3p (100% of samples), the loss of which is a hallmark of SCLC[5,8–13].This was followed by chromosomes 17p (94%), 13q (82%), 5q (65%), and 16q (59%). Chromosome arms 4q, 15q, and 22q all showed LOH in 45–50% of samples. Several other chromosomes, such as chromosome 20, revealed LOH over a large region in a few samples. The results are in excellent agreement with previously reported SSLP and CGH data for SCLC[9–13], including a previous SSLP-based analysis of the same samples[5]. A comparison of the frequency of tumors with LOH for each chromosome arm using either SNP or SSLP markers is shown in Figure 3. We examined a number of instances of apparent conflict between the SSLP- and SNP-based analysis by repeating the analysis and found that discordances were slightly more often due to errors in SSLP rather than in SNP genotyping. SNP genotyping thus appears to be at least as accurate as the SSLP approach.

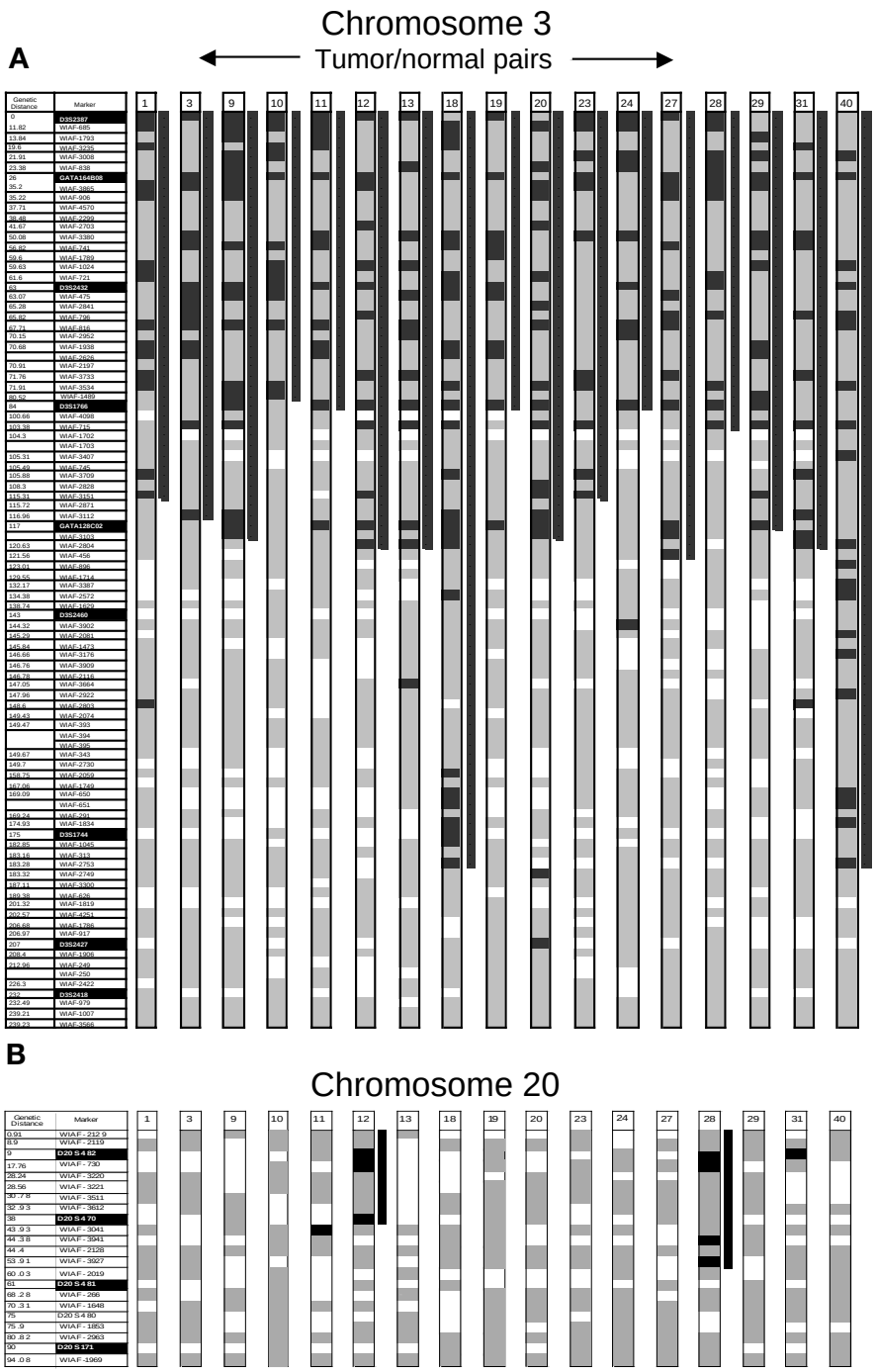


**Figure 2. Maps of chromosome 3 (A) and 20 (B) with estimated windows of LOH[19]. "Genetic distance" is actual genetic distance for SSLPs and inferred genetic distance for SNPs, based on RH map position. Markers in black are SSLPs. Each marker/sample pair is annotated as follows: LOH (black), retention (white), and uninformative (gray). Estimated windows of LOH (calculated as described in the Experimental Protocol) are marked by black bars to the right of each sample. Chromosome 3p showed LOH in all tumors with a few windows extending to the q-arm. This high degree of allelic loss on 3p agrees with previous results. In contrast, on chromosome 20 only two samples show windows of LOH.**

Some regions of the maps show loci with retention interspersed among loci with loss. An example is found near the centromere of chromosome 3p (Fig. 2A). Such data points could represent true interstitial losses, errors in allelotyping, or—most likely—errors in the fine-structure order of the RH map. The issue of map error should be resolved with the rapidly advancing sequencing of the human genome.

We studied whether the allelic imbalance detected in the SCLC tumors represented true allelic losses or, alternatively, increases in the copy number of one chromosome, by performing CGH[14]. Copy number losses identified by CGH, detected most frequently on chromosomes 3p, 4q, and 13q, corresponded well with the allelic imbalance detected by SNP array hybridization (Fig. 3); the fact that the loss is seen by CGH indicates true haploidy in the region rather than reduction to homozygosity by virtue of somatic recombination. Increases in copy number were seen most frequently on chromosomes 19p (75%), 19q (67%), 1p (62%), 17q (56%), and 14q (50%). These results agree with previous CGH experiments[11–13]. The chromosomal gains seen by CGH did not typically correspond to regions of allelic imbalance in the SNP analysis. This is consistent with the results of the mixing experiments above, which indicate that a modest increase in copy number (such as trisomy for a region) would not give rise to allelic imbalance in the SNP assay. Allelic imbalance in the SNP assay should thus usually indicate true loss of heterozygosity, except in the case of extreme amplification of one allele.

In summary, these experiments demonstrate that SNP array hybridization is an accurate and efficient method for evaluating genome-wide tumor LOH. The approach allows for a substantial increase in the number of loci studied and a decrease in the amount of sample required, raising the possibility of analyzing smaller tumor samples than currently possible. With the increasing number of SNPs available, it will be possible to probe the entire genome, as well as specific regions, at much higher resolution than has been routinely performed so far.

## Experimental protocol

**DNA samples.** Tumor and normal lung tissue specimens were collected from SCLC patients at autopsy. Specimens were dissected to remove adjacent tissue, and DNA was isolated as described[10]. Primary DNA stock solutions were diluted into Tris–EDTA (TE). Cell line DNA was prepared using a Genomic Maxiprep Kit (Qiagen, Valencia, CA). Working solutions were diluted to 4 ng/μl in water and processed as below.

**HuSNP assay: PCR amplification.** Normal DNA was assayed once and tumor samples in duplicate according to the HuSNP protocol supplied by Affymetrix, Inc. For each sample, 24 pools of primer pairs (50–100 loci/pool at 50 nM each) were mixed with 5 ng of DNA, 5 mM $MgCl_2$, 0.5 mM dNTPs, 1.25 U Amplitaq Gold (PE Biosystems, Foster City, CA), and the supplied buffer in 12.5 μl per pool. Samples were denatured for 5 min at 95°C, followed by 30 cycles of 95°C for 30 s, 52°C + 0.2°C/cycle for 55 s, and 72°C for 30 s; 5 cycles of 95°C for 30 s, 58°C for 55 s, and 72°C for 30 s; and a final extension of 72°C for 7 min. A 1:1,000 dilution of each pool was made by adding 1 μl of the amplification product to 999 μl of $ddH_2O$. Subsequently, 2.5 μl of the 1:1,000 dilution were transferred to a new plate and amplified with 0.8 μM biotinylated-T7 and 0.8 μM biotinylated-T3 primers, 4 mM $MgCl_2$, 0.4 mM dNTPs, 2.5 U Taq, and the supplied buffer in 25 μl for 8 min at 95°C, followed by 40 cycles of 95°C for 30 s, 55°C for 90 s, and 72°C for 30 s, and a final extension of 72°C for 7 min. Then, 1.5 μl from each pool was tested for successful amplification on a 3% agarose gel. For each sample, the remainder of each of

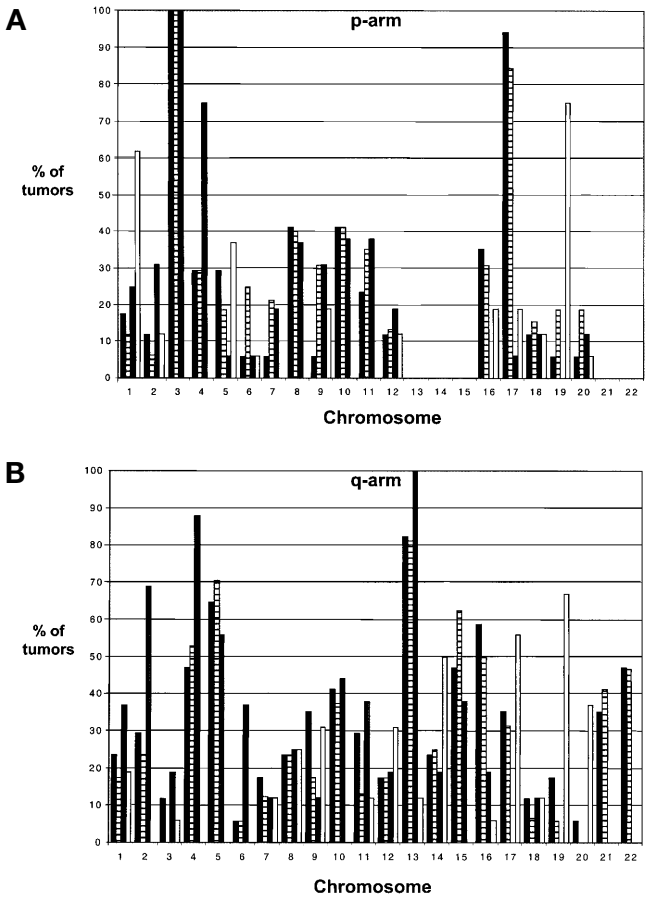


**Figure 3. Fraction of tumors showing allelic imbalance/LOH using either SNP (black), SSLP (striped), or CGH (gray) or gain using CGH (white) on each p (A) and q (B) arm of the chromosomes. Note that the level of allelic imbalance/LOH corresponds well between SNPs and SSLPs as well as with the loss seen by CGH, whereas the gain seen by CGH is essentially not reflected by allelic imbalance/LOH of SNPs.**

the 24 pools was mixed and loaded on a Microcon-10 spin column (Amicon Bioseparations, Bedford, MA). Samples were concentrated by spinning the column for 20 min at 13,000 *g* at room temperature (RT) and eluted by inverting the column and centrifuging for 3 min at 3,000 *g*. Volumes were adjusted to 60 µl.

**Hybridization, washing, and staining.** For this step, 5–30 µl of the sample (depending on the intensity of the chip lot) was diluted in 3 M tetramethylammonium chloride (TMACl), 2 nM control oligonucleotide B1 (supplied by Affymetrix, Inc.), 5× Denhardt's solution, 100 µg/ml herring sperm DNA, 5 mM EDTA pH 8.0, 10 mM Tris pH 7.8, and 0.01% Tween 20 in a volume of 135 µl, and denatured for 10 min at 95°C. After 2 min on ice, the samples were loaded into HuSNP chips and hybridized for 16 h at 44°C and 40 r.p.m. Each chip was washed and stained on the Affymetrix fluidics station as follows. Chips were washed for two cycles of two mixes with 6× SSPET (Bio Whittaker, Walkersville, MD) (6× SSPE (sodium chloride, sodium phosphate, sodium EDTA) + 0.01% Triton X-100) at 25°C, and for six cycles of five mixes with 4× SSPET (4× SSPE + 0.01% Triton X-100) at 35°C. Chips were stained for 30 min at 25°C with 50 µg/ml streptavidin–phycoerythrin and 0.25 mg/ml biotinylated-anti-streptavidin antibody in 6× SSPE, 1× Denhardt's solution, and 0.01% Tween 20 in a volume of 500 µl. The chip was filled with 6× SSPET following six washes of four mixes with 6× SSPET at 25°C.

**Scanning and genotype generation.** Subsequent to hybridization, washing, and staining, the HuSNP probe arrays were scanned using the HP GeneArray Scanner according to the HuSNP Mapping Assay Manual (Affymetrix P/N 700308). Genotype calls were made automatically from the collected hybridization signal intensities by the Affymetrix GeneChip 3.1 software. Each allele (A or B) of a SNP is represented by 4 or 5 complementary probes with different locations of the SNP base position within the 20 nucleotide probes. Each of these probes, in turn, is paired with a probe of the same sequence except for a central mismatch at or near the SNP position. These mismatch probes help to factor cross-hybridization out of the data analysis. A series of data quality and pattern recognition metrics must be passed in order for the software to make a genotype call. The pattern recognition component relies on the relative allele signal previously determined for each SNP[4] and is further described in the HuSNP Mapping Assay Technical Note, available from Affymetrix (P/N 700318).

**SNP analysis and assignment of LOH**. The call rate was calculated as the number of SNPs on each chip that had calls AA, BB, or AB out of the total 1494 SNPs. Any calls falling between clusters were placed in the "no signal" category. For each sample, each SNP was then scored as loss of heterozygosity, retention of heterozygosity, or uninformative/uncertain by comparing the calls for the normal control and the tumor duplicates. Any SNP showing both alleles (AB) in the normal and only one allele (AA or BB) in at least one of the tumor duplicates was scored as loss. Any SNP showing both alleles (AB) in the normal and either both alleles (AB) in both tumor duplicates, or both alleles (AB) in one and "no signal" in the other tumor duplicate, was scored as retention. SNPs that showed only one allele (AA or BB) or "no signal" in the normal, or both alleles (AB) in the normal and "no signal" in both tumor duplicates, were scored as uninformative/uncertain. The number of informative SNPs was defined as the number of SNPs scored as either loss or retention. The reproducibility of the chip was calculated by comparing the duplicate experiments of each tumor sample: (number of matching informative calls)/(number of matching + discordant informative calls). The fraction of LOH was calculated as the fraction of calls with loss compared to the total number of calls with either loss or retention for each sample.

**SNP mapping**. Of the 1,494 SNPs tiled on the chip, 1,149 SNPs have been assigned a map position on the TNG RH panel[15] and 1,016 on the Genebridge4 RH mapping panel[16]. We used the map from the TNG panel, except that we eliminated 68 loci whose position disagreed between the two maps (by virtue of being assigned to different chromosomes or to positions separated by more than 50 cM on the same chromosome). The remaining 1,081 SNPs mapped on the TNG panel were used for all further analysis.

**Estimation of regions of LOH.** A map was created for each sample by placing the 1,081 mapped SNPs in their reported position (see above). An approximate region of LOH was then defined by ignoring all uninformative loci and proceeding as follows. We scanned each chromosome, starting at the tip of the p-arm, to identify the first occurrence of two consecutive markers showing loss. We then initiated a running tally, denoted δ with an initial value of 2 (indicating two consecutive markers showing loss). The tally was incremented by +1 for each successive locus showing loss and by –1 for each successive locus showing retention. The region of LOH was defined as the region corresponding to the largest value of δ. If the informative SNP closest to either end of the chromosome was considered part of a window, the remaining uninformative SNPs at that end of the chromosome were also considered part of the window. The procedure is intended to be robust against occasional errors in genotype or map position, and aims to provide only a rough estimate of the critical region. The analysis was performed both with and without SSLP data for all samples.

**Sample mixing experiment.** Working stock solutions (4 ng/µl) of pure normal (NCI-BL128) and SCLC (NIC-H128) cell line DNA were mixed to produce samples of 90, 80, 70, and 60% tumor DNA content. These dilutions, along with 100% tumor DNA and 100% normal cell line DNA were processed as above on the HuSNP chip. All samples were assayed in duplicate including the normal DNA. The calls from the normal duplicates were used to make one file of composite calls (included were SNPs with the same call on both chips, or a call on one and a "no signal" on the other). The composite call file was then compared to tumor duplicates as above. The 130 SNPs showing LOH in the 100% tumor sample were then checked for calls of LOH, uncertain or retention in the mixed samples.

**Comparison of SNP and SSLP data.** A comparison of the fraction of samples showing LOH using SNP and SSLP data was performed. For the SNP data, a sample was considered to show LOH if it showed a window of LOH on a given chromosome arm, using the algorithm defined above. For SSLP data, chromosome arms were scored as LOH if they showed one (p-arm) or two consecutive informative (q-arm) SSLPs with LOH (Fig. 3A). The p-arms were defined as a region from the beginning of a chromosome up to a defined point on the genetic map, whereas the q-arms were defined as a defined genetic point going to the end of that chromosome. These genetic points are defined as follows (p-end in centimorgans/q start in centimorgans) as follows: chromosome 1 (155.14/172.82), 2 (98.79/123.09), 3 (117/144.32), 4 (54.82 / 88.26), 5 (47.05/71.59), 6 (51.04/79.56), 7 (55.89/76.47),

8 (60.11/79.1), 9 (19.9/44.77), 10 (45.56/76.37), 11 (48.25/74.78), 12 (18.22/35.59), 13 (acrocentric), 14 (acrocentric), 15 (acrocentric), 16 (43.32/58.74), 17 (29.15/61.54), 18 (32.16/50.45), 19 (47.96/75.2), 20 (43.93/60.03), 21 (acrocentric), and 22 (acrocentric). Different definitions were used because there were substantially fewer SSLP loci than SNP loci.

**Validation of SSLPs and SNPs.** We performed PCR amplification using SSLP primers labeled with the fluorescent labels FAM, HEX, and TET, and capillary gel electrophoretic analysis as described[5]. The LM-SBE was performed as described[6], on seven SNPs randomly chosen from the HuSNP assay. Any uncertain SBE calls were removed before the SBE genotypes were compared to the HuSNP genotypes.

**Validation by RH mapping.** Seventeen SNPs with dubious placement (a retention flanked by two SNPs with LOH on each side (LLRLL) or the alternative (RRLRR)) were chosen and validated by RH mapping. Primers for the SNP-containing fragments were designed to yield products of 100–300 bp. For each reaction, 1 µl (25 ng) of Genebridge 4 RH screening panel DNA (Research Genetics, Huntsville, AL) was amplified according to standard protocols. Calls of positive ("1"), negative ("0"), or inconclusive ("2") were submitted to the Whitehead Institute/MIT Center for Genome Research WWW-based radiation hybrid mapping server for placement[17].

**Comparative genomic hybridization (CGH).** With a few modifications, CGH was performed essentially as described[14]. Hybridization of the slides was performed at 37°C for 48 h, and in the evaluation of heterochromatic regions, the short arm of the acrocentric chromosomes and chromosomes X and Y were not included.

**DNA ploidy analysis.** Ploidy analysis was performed as described[18]. Fluorescence at 488 nm was analyzed using the FACScan flow cytometer (Becton-Dickinson, Franklin Lakes, NJ) and the associated Cellquest software, in the Molecular Cytogenetics Reference Laboratory at the Dana-Farber Cancer Institute. It must be noted that the samples used for flow cytometry do not necessarily arise from the same lesion as the DNA used in the allelotyping study, but do originate from the same patient at the same time point.

## Acknowledgments

*We thank T. Webster, N. Nouri, N. Patil, and T. Ryder at Affymetrix, Inc., for valuable assistance with the HuSNP chip. This research was supported in part by a research contract from Bristol-Myers Squibb, Millennium Pharmaceuticals, Inc., Affymetrix (E.S.L. and T.J.H.), and grants from the Pew Foundation (M.M.), the American Lung Association (M.M.), and the National Institutes of Health (HG01806). K.L. is the recipient of a scholarship from the Swedish Institute (4478/1998). D.M.T. is the recipient of a Harvard Medical School Pathology Fellowship (5T34CA72320-03). L.W.O and C.L. were supported by the Swedish Cancer Foundation and by the Torsten and Ragnar Söderberg Foundations. T.J.H. is a recipient of a clinician-scientist award from the Medical Research Council of Canada.*

1. Weinberg, R.A. Tumor suppressor genes. *Science* 254, 1138-1146 (1991).
2. Vogelstein, B. et al. Allelotype of colorectal carcinomas. *Science* 244, 207–211 (1989).
3. Weissenbach, J. et al. A second-generation linkage map of the human genome. *Nature* 359, 794–801 (1992).
4. Wang, D.G. et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082 (1998).
5. Stanton, S.E., Shin, S.W., Johnson, B.E. & Meyerson, M. Recurrent allelic deletions of chromosome arms 15q and 16q in human small cell lung carcinomas. *Genes Chromosom. Cancer* 27, 323–331 (2000).
6. Lindblad-Toh, K. et al. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* 24, 381–386 (2000).
7. Hahn, S.A. et al. Allelotype of pancreatic adenocarcinoma using xenograft enrichment. *Cancer Res.* 55, 4670–4675 (1995).
8. Whang-Peng, J. et al. A nonrandom chromosomal abnormality, del 3p(14-23), in human small cell lung cancer (SCLC). *Cancer Genet. Cytogenet*. 6, 119–134 (1982).
9. Naylor, S.L., Johnson, B.E., Minna, J.D. & Sakaguchi, A.Y. Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer. *Nature* 329, 451–454 (1987).
10. Johnson, B.E. et al. myc family DNA amplification in small cell lung cancer patients' tumors and corresponding cell lines. *Cancer Res*. 48, 5163–5166 (1988).
11. Ried, T. et al. Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization. *Cancer Res.* 54, 1801–1806 (1994).
12. Levin, N.A., Brzoska, P.M., Warnock, M.L., Gray, J.W. & Christman, M.F. Identification of novel regions of altered DNA copy number in small cell lung tumors. *Genes Chromosom. Cancer* 13, 175–185 (1995).
13. Petersen, I. et al. Small-cell lung cancer is characterized by a high incidence of deletions on chromosomes 3p, 4q, 5q, 10q, 13q and 17p. *Br. J. Cancer* 75, 79–86 (1997).
14. Farnebo, F. et al. Alternative genetic pathways in parathyroid tumorigenesis. *J. Clin. Endocrinol. Metab.* **84**, 3775–3780 (1999).
15. Stanford Human Genome Center RH Server. http://www-shgc.stanford.edu/-RH/index.html
16. Walter, M., Spillett, D., Thomas, P., Weissenbach, J. & Goodfellow, P. A method for constructing radiation hybrid maps of whole genomes. *Nat. Genet.* 7, 22–28 (1994).
17. RHmapper. http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl
18. Brodeur, G.M., Williams, D.L., Look, A.T., Bowman, W.P. & Kalwinsky, D.K. Near-haploid acute lymphoblastic leukemia: a unique subgroup with a poor prognosis? *Blood* 58, 14–19 (1981).
19. SCLC-SNP-LOH site. http//www-genome.wi.mit.edu/SNP/human/LOH/

# Simultaneous stochastic sensing of divalent metal ions

**Orit Braha[1]*, Li-Qun Gu[1], Li Zhou[2], Xiaofeng Lu[1], Stephen Cheley[1], and Hagan Bayley[1,2]**

*[1]Department of Medical Biochemistry & Genetics, The Texas A&M University System Health Science Center, College Station, TX 77843-1114, and [2]Department of Chemistry, Texas A&M University, College Station, TX 77843-3225. *Corresponding author (obraha@medicine.tamu.edu)*

Received 13 March 2000; accepted 6 July 2000

**Stochastic sensing is an emerging analytical technique that relies upon single-molecule detection. Transmembrane pores, into which binding sites for analytes have been placed by genetic engineering, have been developed as stochastic sensing elements[1–3]. Reversible occupation of an engineered binding site modulates the ionic current passing through a pore in a transmembrane potential and thereby provides both the concentration of an analyte and, through a characteristic signature, its identity[1–3]. Here, we show that the concentrations of two or more divalent metal ions in solution can be determined simultaneously with a single sensor element. Further, the sensor element can be permanently calibrated without a detailed understanding of the kinetics of interaction of the metal ions with the engineered pore.**

By using engineered channel-forming peptides and proteins, analytes can be detected by processes that occur naturally, including ligand gating[4], channel block[1–3], and selective ion permeation[5]. In favorable circumstances, the analysis can be done at the single-molecule level. Accordingly, we have used the pore formed by staphylococcal α-hemolysin (αHL)[6] as a stochastic sensing element. αHL is a water-soluble 293-amino acid polypeptide that assembles[7;8] to form heptameric pores in membranes[9]. Molecules transported by αHL move through a 100 Å long channel centered on the molecular sevenfold axis[6]. As expected, the relatively narrow diameter of the 14-stranded transmembrane β-barrel compared with the rest of the channel[6] determines the rate of transport of ions through the pore. Therefore, to make sensor elements from αHL, we have built binding sites in the lumen of the barrel so that analytes are registered by their partial block of a transmembrane ionic current[1–3].

The αHL mutant 4H contains histidine residues at positions 123, 125, 133, and 135 (ref. 1). 4H can be combined with the wild-type αHL subunit (WT) to form pores that respond to submicromolar concentrations of divalent metal ions, M(II). Here, we use the heteromer $WT_6 4H_1$ (ref. 1), which contains a single M(II)-binding site,