

Detection of Gene Amplification by Genomic Hybridization to cDNA Microarrays

Mervi A. Heiskanen, Michael L. Bittner, Yidong Chen, Javed Khan, Karl E. Adler, Jeffrey M. Trent, and Paul S. Meltzer¹

Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [M. A. H., M. L. B., Y. C., J. K., J. M. T., P. S. M.], and DuPont NEN Life Science Products, Inc., Boston, Massachusetts 02118 [K. E. A.]

Abstract

Gene amplification is one of the major mechanisms of oncogene activation in tumorigenesis. To facilitate the identification of genes mapping to amplified regions, we have used a technique based on the hybridization of total genomic DNA to cDNA microarrays. To aid detection of the weak signals generated in this complex hybridization, we have used a tyramide-based technique that allows amplification of a fluorescent signal up to 1000-fold. Dilution experiment suggests that amplifications of 5-fold and higher can be detected by this approach. The technique was validated using cancer cell lines with several known gene amplifications, such as those affecting *MYC*, *MYCN*, *ERBB2*, and *CDK4*. In addition to the detection of the known amplifications, we identified a novel amplified gene, *ZNF133*, in the neuroblastoma cell line NGP. Hybridization of NGP cDNA on an identical array also revealed over expression of *ZNF133*. Parallel analysis of genomic DNA for copy number and cDNA for expression now provides rapid approach to the identification of amplified genes and chromosomal regions in tumor cells.

Introduction

Chromosomal anomalies resulting in gain or loss of genetic material are frequent in tumor cells. These changes may result in alterations in the level of expression of numerous genes. Indeed, gene amplification is one of the most common mechanisms for oncogene activation in solid tumors. It is therefore of considerable interest to develop strategies for identifying amplified genes and determining their expression levels in cancer. Until recently, gene amplification has been detected by DNA-based techniques (PCR or Southern blot) or by molecular cytogenetic techniques (FISH² with gene-specific probes). These techniques are inherently restricted to the analysis of previously known amplified genes. In contrast, genome-wide scanning of amplified chromosomal regions with CGH (1) has become an important technique for the detection of amplified regions in tumor DNA. However, CGH has limited sensitivity and resolution (2). In addition, the identification of the specific target gene within an amplicon defined by CGH remains daunting because of the limited mapping resolution provided by the metaphase chromosomes.

DNA microarray technology offers the possibility to replace the target metaphase chromosomes with arrays of DNA clones on a microscope slide. Arrayed fragments of cloned genomic DNAs have been used for this purpose (3, 4). These CGH microarray techniques allow amplification detection on the resolution level equal to the length of the arrayed DNA clones (~100 kb). Genomic microarrays have been applied to amplicon mapping (4), but the technique can also be used for rapid surveys of known copy number alterations in tumor

samples. In principle, a further increase in resolution can be obtained by using arrayed cDNAs rather than genomic DNA. This approach is particularly attractive because of the availability of thousands of accurately mapped cDNAs. Furthermore, expression analysis can be carried out in parallel on the same microarray slides, enabling a correlation of copy number and gene expression. However, signal intensities in genomic hybridizations are proportional to the length of the target DNA (4). Reproducibly achieving a measurable hybridization signal from total genomic DNA hybridized to targets covering only 0.5–2 kb is difficult and requires a signal detection system with high sensitivity and low background. An approach for CGH on cDNA microarrays was reported recently by Pollack *et al.* (5) using directly labeled fluorochrome probes. We report here an alternative technique that also uses cDNA arrays prepared for expression studies. Our results indicate that this technique is reproducibly capable of detecting gene amplifications of 5-fold or higher. Finally, the suitability of the technique for genome-wide screening of amplified and overexpressed genes was tested by hybridizing both genomic neuroblastoma DNA and mRNA on a microarray containing 1400 genes. In addition to known amplifications on 2p and 12q, a previously unrecognized amplicon on 20p containing a zinc finger gene (*ZNF133*) was identified.

Materials and Methods

Preparation of cDNA Microarrays

cDNA microarrays were prepared as described previously (6). To validate the technique, microarrays containing 14 different cDNAs representing genes known to be amplified in cancer cell lines were printed (*CDK4*, *MDM2*, *OS4*, *OS9*, *MYCN*, *MYC*, *MYCL*, *EGFR*, *AKT2*, *ERBB2*, *AIB1*, *IGFR1*, *CLND*, and *SAS*). For screening of unknown amplified genes, an array containing 1400 cDNAs was used (6).

Genomic Hybridization, Signal Detection, and Amplification

Total genomic DNA was labeled with biotin by nick translation for 2.5 h at 15°C. The fragment size of the labeled probe was between 400–2000 bp. Unincorporated nucleotides were removed using Micro Bio-Spin 6 Chromatography Columns (Bio-Rad, Hercules, CA). Hybridization mixture (10 μ l) was composed of 1 μ g of biotinylated probe, 40 μ g of Cot-1 DNA (Life Technologies, Inc., Rockville, MD) and 8 μ g of poly dA in 3X SSC (0.15 M NaCl, 0.015 M sodium citrate)/0.01% SDS. After denaturation, the hybridization mixture was added on the slide and hybridized in a hybridization chamber at 65°C over night. Slides were washed in 0.5X SSC/0.01% SDS, 0.06X SSC/0.01% SDS, and 0.06X SSC at room temperature for 5 min each. Hybridization signals were developed using tyramide reagents (Renaissance TSA-indirect ISH; DuPont NEN Life Science Products). Slides were blocked using 10% goat serum in TN blocking buffer [0.1 M Tris (pH 7.6), 0.15 M NaCl]. Hybridization was detected by first incubating the slide with streptavidin conjugated with horseradish peroxidase (1:100 in TN-10% goat serum), followed by signal amplification with biotinyl tyramide (1:50 in reaction buffer with 1% blocking reagent). Biotinyl tyramide was detected by streptavidin conjugated with Cy3 (1:500 in TN-10% goat serum). Between and after incubations, slides were washed with TNT buffer [0.1 M Tris-HCl (pH 7.5),

Received 9/22/99; accepted 1/3/00.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ To whom requests for reprints should be addressed, Cancer Genetics Branch, National Human Genome Research Institute, NIH, 49 Convent Drive, MSC 4470, Bethesda, MD 20892.

² The abbreviations used are: FISH, fluorescence *in situ* hybridization; CGH, comparative genomic hybridization.

0.15 M NaCl, and 0.05% Tween 20] 3×1 min. All of the incubations and washings were done at room temperature.

cDNA Hybridization for Gene Expression Analysis

Hybridization of NGP cDNA on a microarray containing 1400 genes has been described in detail previously (6).

Image Analysis and Outlier Detection

Amplification Intensity. A gray scale fluorescent image for each microarray slide was obtained from a confocal scanning microscope. DNA target segmentation and signal detection methods were then used to determine the actual target regions, average signal intensities, and local background intensities (7). The background subtracted average signal intensity was reported as the hybridization intensity.

Fold Increase. An iterative amplification intensity outlier detection algorithm was then applied as follows:

(a) Assuming there were N cDNA targets presented in microarray slides, we first sorted all intensities I_k in ascending order, $I_1 > I_2 > \dots > I_N$. We then partitioned intensities into two groups (I_1, \dots, I_{N-m}) and (I_{N-m+1}, \dots, I_N). Initially, we chose $m = N/2$.

(b) The discordance test (for a single outlier in a normal sample with m and s unknown) was performed for the first amplification intensity from the second group:

$$\frac{\mu_1 - I_{N-m+1}}{\sqrt{(\sigma_1^2 + \sigma_b^2)}}$$

where

$$\mu_1 = \frac{\sum_{k=1}^{N-m} I_k}{N-m}$$

and

$$\sigma_1 = \frac{\sum_{k=1}^{N-m} (I_k - \mu_1)^2}{N-m}$$

and σ_b is the SD of local background at the same location of I_{N-m+1} . The test statistic T can be converted to Student's t test with $N-2$ degrees of freedom [Barnett and Lewis (8)]. As an example, the critical value for $n = 40$ and $\alpha = 1\%$ discordance, t must be >3.24 .

3) If the I_{N-m+1} was not an outlier, let $m \leftarrow m+1$, and then repeat step 2 until the first outlier intensity was obtained.

If at least one outlier was obtained, the amplification intensities were partitioned into two groups: (a) negative targets that exhibit no significant amplification intensities; and (b) positive targets of which intensities were statistically different from those from the negative group. The fold increase r_k was then calculated for all genes in the positive target group (I_{N-m+1}, \dots, I_N) by $r_k = I_k/\mu_1$ for $k = N-m+1, \dots, N$.

P Value. To further assess the significance of each reported positive amplification intensity, P from the aforementioned discordance t test statistic (8) can be calculated.

In Situ Hybridization

A BAC clone specific for *ZNF133* (169o05) was screened from a human BAC library (Research Genetics, Inc. Huntsville, AL) using primers specific for marker WI-18789.³ The probe was labeled with Spectrum Orange by random priming (BioPrime DNA Labeling System; Life Technologies, Inc.). FISH-based copy number determination for the *ZNF133* region included a fluorescein-labeled satellite probe for chromosome 20 (Oncor).

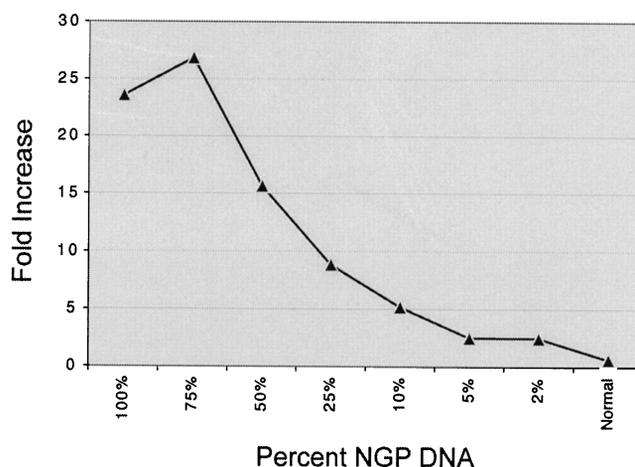


Fig. 1. Dilution of NGP neuroblastoma cell line DNA with normal DNA. The X-axis shows the different dilutions of tumor DNA, from 100% of NGP (representing ~ 100 -fold amplification of *MYCN*) down to 2% of NGP DNA in the labeling reaction. The *MYCN* ratio on the Y-axis was calculated by dividing the *MYCN* signal intensity with the average intensity of all of the nonpositive spots on the array. Four hybridizations of the same normal DNA used for NGP dilution were included in the experiment. The average ratio of the *MYCN* spot to the other spots on the array in four normal control hybridizations was 0.66 (SD, 0.70).

Results

We initially tested tyramide signal detection on microarrays containing 14 genes amplified in various cancers by hybridizing genomic DNA from a neuroblastoma with a 100-fold amplification of the *MYCN* oncogene (9). *MYCN* was specifically identified as an outlier in this hybridization. To determine the sensitivity of the procedure, tumor DNA was diluted with normal DNA prior to labeling. Seven dilutions ranging from 100 to 2% of NGP DNA were hybridized to cDNA microarrays (Fig. 1). Four normal control hybridizations were also included in the experiment. The *MYCN* signal intensities were normalized to the average signal intensity of all of the nonamplified spots on the array. The results revealed a decrease in the *MYCN* cDNA signal intensity with dilution of the NGP DNA. In the hybridization containing 2% NGP DNA, the *MYCN* signal intensity was increased 2.5-fold relative to the nonamplified spots, indicating that amplifications of ~ 5 -fold can be detected. In this experiment, the signal intensities were not directly proportional to the gene copy number. The lower signal intensity of the undiluted NGP hybridization compared with the first dilution (75% of NGP) is likely explainable by the increased self annealing of the probe, which limits the hybridization of the probe to the target cDNA.

We then extended this technique to other cancer cell lines with known gene amplifications (Refs. 10–12; Table 1; Fig. 2). Amplified genes were identified according to the distribution of signal intensities, and a fold change in signal intensity was calculated by dividing the intensity of any particular spot by the average intensity of all of the nonamplified spots on the array. By these criteria, 11 of 13 known gene amplifications were detected. The amplification level of identified genes ranged from 5-fold (*ERBB2*) up to 100-fold (*MYCN*). Only two previously recognized gene amplifications were not detected by this method were *OS9* in NGP and *MYC* in BT474. Although the NGP *OS9* signal intensity was 2-fold above the average, this result did not meet our criteria for statistical significance. The level of *MYC* amplification in BT474 has been determined to be only 3.5-fold by interphase FISH.⁴ However, *ERBB2* amplification was detected in BT474 (6-fold⁴) as well as in SKBR3 (5-fold⁴). These results correlate well

³ Internet address: <http://www.genome.wi.mit.edu>.

⁴ A. Kallioniemi, personal communication.

Table 1 Gene amplifications detected by genomic hybridization of cancer cell line DNA on cDNA microarrays

Cell line	Nonpositive mean intensity	Gene	Amplified gene intensity	Fold change	P^a
NGP	6861.00	<i>MYCN</i>	42138.00	6.14	1.4×10^{-12}
		<i>SAS</i>	23302.00	3.39	
		<i>CDK4</i>	19192.00	2.79	
		<i>MDM2</i>	20509.00	2.99	
		<i>OS4</i>	36832.00	5.37	
HL60	1097.20	<i>MYC</i>	30675.00	28.20	0.00
SKBR3	347.00	<i>MYC</i>	2801.00	8.07	0.00
		<i>ERBB2</i>	9305.00	26.80	0.00
		<i>ERBB2</i>	4981.00	5.60	0.00
BT474	879.00	<i>ERBB2</i>	4981.00	5.60	0.00
COLO 320	1206.00	<i>MYC</i>	42076.00	34.90	0.00
UACC1598	783.00	<i>MYC</i>	41648.00	53.20	0.00

^a P was calculated from the discordance T test (see "Materials and Methods").

with the NGP dilution experiment (Fig. 1), which suggested that the sensitivity of detection is ~ 5 -fold.

To determine whether previously unknown amplification could be identified this technique, we hybridized NGP genomic DNA to a 1400 element cDNA microarray. In addition to the known amplified genes, a novel gene amplification was identified in this experiment (Fig. 3). The signal intensity of an expressed sequence tag representing *ZNF133* was observed to be significantly increased (8-fold over the mean). *ZNF133* has been mapped previously to 20p11.2 by *in situ* hybridization (13). CGH of NGP on metaphase chromosomes also indicated the amplification of this chromosomal region (data not shown). To verify the amplification, a *ZNF133* BAC clone was hybridized on NGP cells *in situ*. The signal copy number in interphase nuclei was heterogeneous, ranging from 2 copies up to 22 copies of *ZNF133* per cell. The average copy number calculated from 50 cells was 6.6. In contrast, two copies of the chromosome 20 centromere reference probe were detected in all nuclei.

In two-color hybridization of NGP cDNA (red) relative to normal fibroblast cDNA (green), the red:green ratio of the *ZNF133* was 3.4, indicating a significant increase in expression of this amplified gene relative to the reference probe (6). This demonstrates the ability to perform both expression and copy number analysis on the same set of microarrays.

Discussion

Compared with the previously published CGH microarray techniques where genomic DNA is hybridized on arrayed genomic clones

(3, 4), hybridization on cDNA arrays offers significant advantages. The most important is the possibility to directly identify amplified genes rather than amplified genomic regions. Another advantage is the ability to do expression studies on the same arrays using the standard cDNA microarray approach. Finally, thousands of mapped cDNAs are readily available, which facilitates mapping and identification of new cancer genes.

The complexity of the probe and the small sizes of the arrayed target cDNAs (0.5–2 kb) place high demands on the sensitivity of the system. Using tyramide-based signal amplification (14), it is possible to enhance fluorescent signals up to 1000-fold. Deposition of biotin tyramides has been applied previously for the amplification of *in situ* hybridization signals (15). We show here that the peroxidase-mediated deposition of biotin tyramide can also be applied on high sensitivity detection of gene amplification on cDNA microarrays to detect gene amplifications of 5-fold or greater. Although we have tested the addition of a second tyramide reagent to provide two-color CGH, there was excessive cross-talk between tyramide reagents under the conditions necessary for genomic hybridization. Nonetheless, the tyramide method described here consistently generates significant signal intensities necessary for a screening technique for gene amplification.

In the hybridization of NGP DNA to a 1400-element cDNA microarray, we identified the known 2p and 12q amplicons as well as a novel 20p amplicon containing a zinc finger gene *ZNF133*. This gene belongs to the family of Kruppel-related zinc finger genes that have been connected with transcriptional repression (13). Amplification

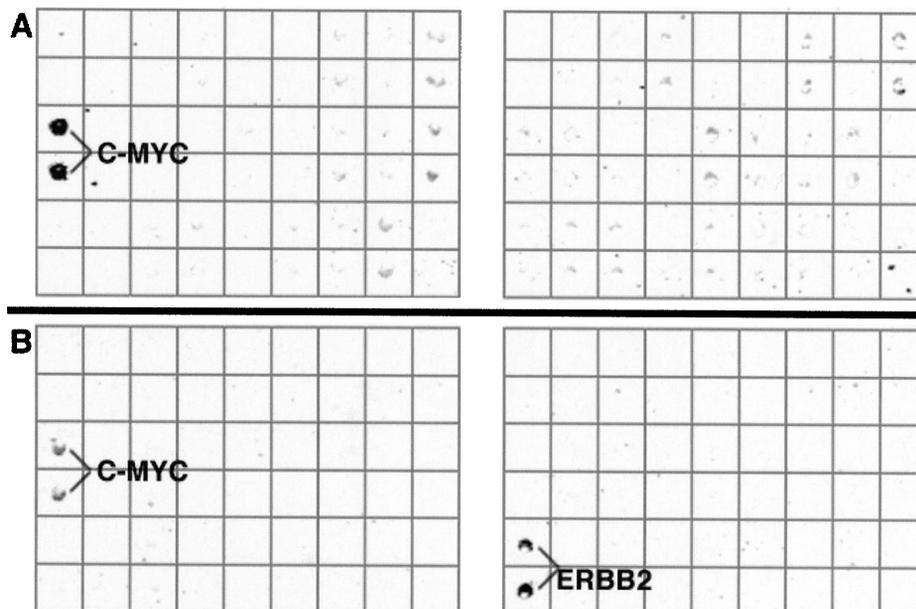


Fig. 2. Two examples of the genomic hybridization on cDNA arrays. A, high-level amplification of *MYC* in colon cancer cell line Colo320. Signal intensity of *MYC* is 16-fold higher than the nonamplified signal intensity in this hybridization. B, hybridization of breast cancer cell line SKBR3. In this hybridization, the signal intensity ratios were 9-fold for *MYC* and 22-fold for *ERBB2*. The level of *ERBB2* amplification has been determined to be 6-fold and *MYC* amplification 32-fold by interphase FISH.⁴ The grid indicates the pattern of cDNAs printed on the microarray.

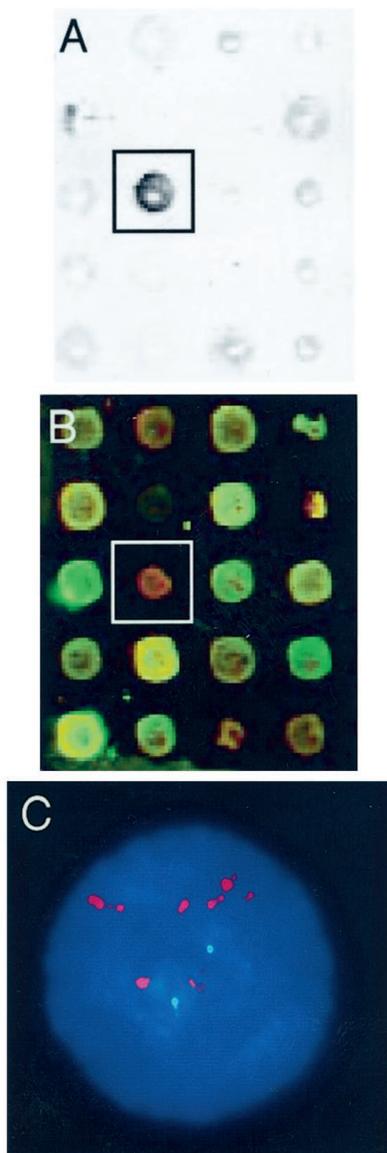


Fig. 3. Amplification and overexpression of zinc finger gene *ZNF133* in NGP neuroblastoma cell line. **A**, section of a 1400-element cDNA microarray containing *ZNF133* (boxed) after hybridization of NGP total genomic DNA and tyramide detection. The intense *ZNF133* signal (8-fold) suggests amplification of *ZNF133*. **B**, corresponding section of an identical microarray after cDNA hybridization with NGP cDNA (pseudocolor red) and normal fibroblast DNA (pseudocolor green). The red:green ratio of 3.37 for *ZNF133* (boxed) shows that the amplified copies of *ZNF133* are highly overexpressed. **C**, interphase FISH on NGP confirming amplification of *ZNF133* (red) relative to a chromosome 20 α satellite (green).

and overexpression of *ZNF133* have not been reported previously, but the amplification of this chromosomal region has been detected by CGH in several different types of malignancies. In a CGH study of 58 primary gastric cancers, 20p gain was detected in 38% of cases (16). In chondrosarcomas, gain of 20p was observed in 31% of the analyzed tumors (17). 20p amplification has also been reported in ductal carcinoma *in situ* of the breast, in bladder tumors, as well as in osteosarcoma, ovarian cancer, adenocarcinoma of gastroesophageal junction, squamous cell carcinoma, small cell lung cancer, and non-small cell lung cancer (18). Although it is impossible to delineate the size and genetic composition of the 20p amplicon from the small microarray used in this study, larger arrays will provide an amplicon map at

higher density. This information should prove extremely useful for focusing efforts to identify amplification target genes. With sufficient cDNA density, it should prove possible to map core regions of amplification in multiple tumor specimens. This information can then be further correlated with expression patterns determined across the amplified region. When integrated with the rapidly emerging human genome sequence, this approach should greatly accelerate the discovery of genes amplified during tumor progression.

Acknowledgments

We thank Kimberly Gayton, William Giasi, Gerald Gooden, and John Lueders for excellent technical assistance.

References

- Kallioniemi, A., Kallioniemi, O-P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* (Washington DC), 258: 818–821, 1992.
- Piper, J., Rutovitz, D., Sudar, D., Kallioniemi, A., Kallioniemi, O-P., Waldman, F. M., Gray, J. W., and Pinkel, D. Computer image analysis of comparative genomic hybridization. *Cytometry*, 19: 10–26, 1995.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, 20: 399–407, 1997.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, 2: 207–211, 1998.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy number changes using cDNA microarrays. *Nat. Genet.*, 23: 41–46, 1999.
- Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., and Meltzer, P. S. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.*, 58: 5009–5013, 1998.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Biomed. Optics*, 2: 364–374, 1997.
- Barnett, V., and Lewis, T., *Outliers in Statistical Data*, Ed. 3, p. 216. New York: John Wiley and Sons, 1994.
- Schwab, M., Varmus, H. E., Bishop, J. M., Grzeschik, K-H., Naylor, S. L., Sakaguchi, A. Y., Brodeur, B., and Trent, J. Chromosome localization in normal human cells and neuroblastomas of a gene related to *C-MYC*. *Nature* (Lond.), 308: 288–308, 1984.
- Dalla-Favera, R., Wong-Staal, F., and Gallo, R. C. Oncogene amplification in promyelocytic leukemia cell line HL-60 and in primary leukaemic cells of the same patient. *Nature* (Lond.), 299: 61–63, 1982.
- Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E., and Bishop, J. M. Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (*C-MYC*) in malignant neuroendocrine cells from a human colon carcinoma. *Proc. Natl. Acad. Sci. USA*, 80: 1707–1711, 1983.
- Elkahloun, A., Bittner, M., Hoskins, K., Gemmil, R., and Meltzer, P. Molecular cytogenetic characterization and physical mapping of 12q13–15 amplification in human cancers. *Genes Chromosomes Cancer*, 17: 205–214, 1996.
- Tommerup, N., and Vissing, H. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics*, 27: 259–264, 1995.
- Bobrow, M. N., Harris, T. D., Shaughnessy, K. J., and Litt, G. J. Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays. *J. Immunol. Methods*, 125: 279–285, 1989.
- Raap, A. K., van de Corput, M. P. C., Vervenne, E. A. W., Gijlswijk, R. P. M., Tanke, H. J., and Wiegant, J. Ultrasensitive FISH using peroxidase-mediated deposition of biotin of fluorochrome tyramides. *Hum. Mol. Genet.*, 4: 529–534, 1995.
- Sakakura, C., Mori, T., Sakabe, T., Ariyama, Y., Shinomiya, T., Date, K., Hagiwara, A., Yamaguchi, T., Takahashi, T., Nakamura, Y., Abe, T., and Inazawa, J. Gains, losses, and amplifications of genomic materials in primary gastric cancers analyzed by comparative genomic hybridization. *Genes Chromosomes Cancer*, 24: 299–305, 1999.
- Larramendy, M. L., Tarkkanen, M., Valle, J., Kivioja, A. H., Ervasti, H., Karaharju, E., Salmivalli, T., Elomaa, I., and Knuutila, S. Gains, losses, and amplifications of DNA sequences evaluated by comparative genomic hybridization in chondrosarcomas. *Am. J. Pathol.*, 150: 685–691, 1997.
- Knuutila, S., Bjorkqvist, A-M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., Hemmer, S., Wasenius, V-M., Vidgren, V., and Zhu, Y. DNA copy number amplifications in human neoplasms—a review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107–1123, 1998.