

# Genomic and Expression Analysis of the 12p11-p12 Amplicon Using EST Arrays Identifies Two Novel Amplified and Overexpressed Genes<sup>1</sup>

Véronique Bourdon, Félix Naef, Pulivarthi H. Rao, Victor Reuter, Samuel C. Mok, George J. Bosl, Sanjay Koul, Vundaralli V. S. Murty, Raju S. Kucherlapati, and R. S. K. Chaganti<sup>2</sup>

Cell Biology Program [V. B., P. H. R., R. S. K. C.], Department of Medicine [G. J. B., R. S. K. C.], and Department of Pathology [V. R.], Memorial Sloan-Kettering Cancer Center, New York, New York 10021; Center for Studies in Physics and Biology, Rockefeller University, New York, New York 10021 [F. N.]; Laboratory of Gynecologic Oncology, Brigham and Women's Hospital, Boston, Massachusetts 02115 [S. C. M.]; College of Physicians and Surgeons, Columbia University, New York, New York 10032 [S. K., V. V. S. M.]; and Harvard Partners Center for Genetics and Genomics, Boston, Massachusetts 02155 [R. S. K.]

## ABSTRACT

We performed parallel array comparative genomic hybridization and array expression analysis of the 12p11-p12 amplicon in human testicular seminomas and an ovarian carcinoma cell line using an expressed sequence tags (ESTs) array spotted with 8254 ESTs. The data were normalized using a robust statistical modeling and the significance inferred from the local SD. We identified two ESTs within the chromosomal amplicon that were amplified and overexpressed in  $\geq 75$ –100% of analyzed tumors with the 12p11-p12 amplicon. These sequences, belonging to coding regions of two novel genes designated here as *GCT1* and *GCT2*, were broadly expressed in a panel of human tissues, including testis and ovary. *GCT1* and *GCT2* were overexpressed in 92 and 71%, respectively, of a panel of seminomas tested. Combined array comparative genomic hybridization and array expression analysis is a valid approach for gene discovery in large chromosomal amplicons.

## INTRODUCTION

Chromosomal amplification is a common mechanism by which genes achieve overexpression in tumors. Identifying amplified sequences thus provides a starting point for discovering genes involved in cancer biology (1, 2). In recent years, chromosomal CGH<sup>3</sup>, where test (tumor) and reference (normal) gDNAs are cohybridized to normal metaphase chromosomes, has been used to identify DNA copy number variations across the genome at a mapping resolution of  $\sim 20$  Mb (3). In contrast, cohybridization of test and reference gDNAs to arrays spotted with genomic clones (genomic arrays) has been successfully used to identify DNA copy number variation of shorter ( $\sim 100$  Kb) genomic sequences (4–8). EST array CGH, where test gDNAs alone (9) or test and reference gDNAs together are hybridized, can detect genomic amplification of expressed sequences (10, 11). By hybridizing test and reference cDNAs to the same EST array in parallel, the variation in expression of the amplified genes can simultaneously be analyzed. Thus, powerful high throughput techniques are now available to rapidly screen tumors for gene amplification and expression leading to gene discovery and new insights into mechanisms of carcinogenesis.

Human GCTs are a heterogeneous group of neoplasms that present in the testis, ovary, and at extragonadal sites. Cytogenetic analysis of TGCTs and cell lines derived from them has shown that overrepresentation of 12p, either as one or more copies of i(12p) or as tandem

amplification of 12p, occurs in virtually all tumors and comprises the hallmark of this tumor type (12). 12p amplification has also been observed in a broad spectrum of human neoplasms, including ovarian tumors (13)<sup>4</sup>, indicating that one or more amplification targets of relevance to tumorigenesis *per se* reside on this chromosomal arm. In GCTs and other tumors, in addition to the amplification of the entire 12p, a subregional amplification of 12p11-p12 was identified, which has been suggested to be associated with tumor progression (14). Several of the known genes mapped to 12p have been shown to be in increased copy number at the genomic level in TGCTs; however, their possible overexpression and role in tumorigenesis has not been fully established (15–19). Here, we report parallel array CGH and array expression analysis of the 12p11-p12 amplicon using an array spotted with 8254 ESTs. We normalized the array data using a robust statistical modeling (20). The significance of differences in the copy number and level of expression was inferred from the empirically determined local SD of the log ratios. We identified two amplified and overexpressed ESTs mapped to 12p11-p12. These ESTs were in the coding region of novel genes of unknown function, designated here as *GCT1* and *GCT2*. This study shows that combined array CGH and array expression analysis, in conjunction with the data analysis algorithm used here, is a valid approach to the molecular dissection of chromosomal amplification in cancer.

## MATERIALS AND METHODS

**Tumor Specimens, Cell Lines, and Chromosomal CGH.** The normal testis and TGCT biopsies studied were ascertained by the Department of Pathology and the Laboratory of Cancer Genetics at the Memorial Sloan-Kettering Cancer Center. They comprised 24 seminomas and 37 nonseminomas. All of the 12 TGCT cell lines used in these studies (CL-268A, CL-154A, CL-169A, CL-169A, CL-318C, CL218A, CL-354B, CL-577M-Lu, CL-577M-F, CL-2102E-P, CL-2102ER, Tera1, and Tera-2) were of nonseminoma derivation and were either publicly available (CL-577M-Lu, CL-577M-F, CL-2102E-P, CL-2102ER, Tera1, and Tera-2) or were derived at the Laboratory of Cancer Genetics, Memorial Sloan-Kettering Cancer Center from biopsy samples. The OCCL cell line was provided by Dr. Samuel C. Mok. This cell line was developed from a 71-year-old patient with stage III grade 2 ovarian papillary serous adenocarcinoma. Chromosomal CGH was performed as described previously (21).

**RNA Preparation.** RNA was isolated from normal testis, tumor tissues, and cell lines using RNeasy mini or midi kits and following manufacturer-recommended protocols (Qiagen, Valencia, CA). Contaminating DNA was removed by digestion with RNase-free DNase (Life Technologies, Inc., Grand Island, NY) followed by cleaning using an RNeasy cleanup kit (Qiagen). To further rule out DNA contamination, 0.5  $\mu$ g of cleaned RNA was subjected to RT-PCR using  $\beta$ -actin-specific primers. RNA was considered DNA-free only if no product appeared in an identical control reaction where the reverse transcriptase (Superscript II; Life Technologies, Inc., Grand Island, NY) has been omitted. RNA quality and quantity were determined by measuring the absorbance at 260 and 280 nm and by denaturing gel electrophoresis.

Received 5/14/02; accepted 8/29/02.

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.

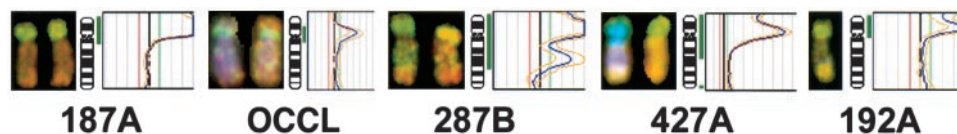
<sup>1</sup> This work was supported by grants from the NIH and the Byrne Research Fund.

<sup>2</sup> To whom requests for reprints should be addressed, at Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021. Phone: (212) 639-8121; Fax: (212) 717-3541; E-mail: chagantr@mskcc.org.

<sup>3</sup> The abbreviations used are: CGH, comparative genomic hybridization; gDNA, genomic DNA; EST, expressed sequence tag; RT-PCR, reverse transcription PCR; OMIN, Online Mendelian Inheritance in Man; TOGA, TIGR, Orthologous Gene Alignments; GCT, germ cell tumor; TGCT, testicular GCT; OCCL, ovarian carcinoma cell line; AECOM, Albert Einstein College of Medicine; TIGR, The Institute for Genomic Research.

<sup>4</sup> Internet address: [www.helsinki.fi/~Igl\\_www/HLA/HLA.html](http://www.helsinki.fi/~Igl_www/HLA/HLA.html).

Fig. 1. Chromosomal CGH analysis. gDNA from tumor (green) and placenta (red) were cohybridized to metaphase chromosomes. Four seminomas (187A, 287B, 427A, and 192A) and the OCCL cell line showed amplification at 12p11-p12.



**EST Arrays.** The EST arrays used were generated at the AECOM microarray facility (22). Per printing, 8254 T3/T7 PCR-amplified human Image clones (ESTs) were pin-printed on poly-L-lysine or amino-silane (CMT-GAPS; Corning, Corning, NY)-coated slides. A Toto-3 dye (Molecular Probe, Eugene, OR) staining of one slide per printing was performed routinely to monitor the quality of printing.

**Array Preparation and Prehybridization.** Array slides were prepared for hybridization as described in the AECOM microarray facility URL.<sup>5</sup> For array CGH, the slides were prehybridized for 1–5 h with 20  $\mu$ l of prehybridization solution consisting of 35% formamide, 4 $\times$  saline-sodium phosphate-EDTA, 0.5% SDS, 2.5 $\times$  Denhart, 0.2 mg/ml Salmon sperm DNA (Sigma, St. Louis, MO), and 1  $\mu$ g/ml Cot-1 DNA (Life Technologies, Inc.). For expression array, the slides were prehybridized for 1 h with cDNA hybridization buffer (Genisphere, Hatfield, PA) plus 0.1% BSA before cDNA hybridization and for another hour with the dye hybridization buffer (Genisphere, Hatfield, PA) plus 0.1% BSA before tagged dye hybridization.

**gDNA Labeling and Array Hybridization.** A total of 5–10  $\mu$ g of test and reference gDNAs was digested with *DpnII* or *AvaI* restriction enzymes for 1 h, purified using a PCR clean-up kit (Promega, Madison, WI), and extracted in 50  $\mu$ l of water. Digested DNAs were concentrated, if necessary, by ultrafiltration (Microcon YM-30, Amicon; Millipore, Bedford, MA). Equal amounts of test and reference DNAs (1.8–2.2  $\mu$ g) were labeled in 50  $\mu$ l with 20- $\mu$ g random nonamers (Genelink), 50-unit 3'-5' exo minus Klenow fragment (New England Biolabs, Beverly, MA), dATP, dGTP, dCTP (120  $\mu$ M each), and dTTP (60  $\mu$ M). Cy3-dUTP or Cy5-dUTP (60  $\mu$ M; Amersham, Piscataway, NJ) were added to the test and reference reactions, respectively. The reaction mixtures were pooled, purified, and concentrated by ultrafiltration (Microcon YM-30, Amicon; Millipore) and adjusted to contain 35% formamide, 0.5% SDS, 2.5 $\times$  Denhart, and 0.25 $\times$  saline-sodium phosphate-EDTA in a final volume of 19  $\mu$ l. To this, 1  $\mu$ l of block solution (10) was added. The probe was then denatured for 2 min at 100°C, annealed with Cot-1 DNA for 15 min at 50°C, and hybridized to the array for 12–15 h at 50°C. The slides were washed for 5 min at 55°C in 1 $\times$  SSC, 0.1% SDS, followed by three washes at room temperature: 10 min in 0.2 $\times$  SSC, 0.1% SDS; 20 min in 0.2 $\times$  SSC; and twice 20 min in 0.1 $\times$  SSC. The slides were then centrifuged for 5 min at 900  $\times$  g before scanning.

**RNA Reverse Transcription, cDNAs Labeling, and Array Hybridization.** Equal quantities (8–10  $\mu$ g) of test and reference RNAs were reverse transcribed, labeled with the 3DNA Expression Array Detection kit (Genisphere, Hatfield, PA), and hybridized to the arrays as described in the manufacturer's protocol. After reverse transcription, the cDNA pellets were dried with compressed N<sub>2</sub> followed by incubation at 37°C for 3–5 min.

**Array Image Analysis.** The arrays were scanned with either the AECOM (22) or an Axon dual color laser scanner (GenePix 4000A; Axon, Union City, CA). At the time of the scanning, the laser power was adjusted to have <5% features saturated; the digitized Cy3 and Cy5 signals were pseudocolored in green and red, respectively (GenePix Pro 3.0; Axon). After gridding, each dot on the 24-bit ratio image was visually inspected and, if necessary, unsatisfactory dots were manually flagged. A GenePix results (\*.gpr) file of the raw data (F635 median-B635 median, F532 median-B532 median) was used for additional data analysis as described.

**Copy Number Estimation by Southern Blot Analysis.** For each tumor and control, 5  $\mu$ g of gDNA were digested with *EcoRI*, separated by electrophoresis, transferred to nylon membranes (Sure Blot; Intergen, Purchase, NY), and hybridized with specific probes for *GCT1* (R44861) and *GCT2* (R70583). Signals were quantified using ImageQuant (Molecular Dynamics, Sunnyvale, CA), and values for *GCT1* and *GCT2* were normalized with the values obtained for D2S48 (ATCC 59212).

**Semiquantitative RT-PCR Analysis of RNA Expression.** A multiple tissue cDNA panel consisting of 16 normal human tissues generated using

normalized first strand poly(A)<sup>+</sup> RNA (Clontech, Palo Alto, CA) was used for analysis of normal expression pattern of *GCT1* and *GCT2*. Total RNA isolated from normal testis, tumor tissues, and the tumor cell lines was reverse transcribed using random primers and Pro-Star first strand RT-PCR kit (Stratagene, La Jolla, CA) following manufacturer-recommended protocol. Semi-quantitative analysis of gene expression was performed by 26–28 cycles of multiplex RT-PCR with  $\beta$ -actin plus *GCT1* or *GCT2*-specific primers. Primers for RNA amplification were selected to span at least two exons to rule out gDNA contamination. *GCT1*-forward 5'-TGTTGACAGCAGTGGAACTC-3' (580–599) and *GCT1*-reverse 5'-ACCAACTGGGTAGGTGTGGA-3' (822–841) amplified 262 bp of cDNA. *GCT2*-forward 5'-GATCTCCTTAACGGACACGC-3' (96–115) and *GCT2*-reverse 5'-TCAAAGGCCAACCAATAAGG-3' (327–346) amplified 251 bp of cDNA. The PCR products were run on 1.5% agarose gels, visualized by ethidium bromide staining, and quantitated using a Kodak digital image analysis system. The values obtained for each gene in four normal testis RNA samples were used to calculate the SD, and their mean was used to calculate the ratio (test/normal). A gene was considered overexpressed if the ratio was >(1.00 + 2SD).

## RESULTS

**Mapping of the Sequences Spotted on the Arrays.** Among the 8254 ESTs spotted on the array, 252 were singletons, whereas the remaining were included in 7366 UniGene clusters, including 6167 that had mapping information. Of these, 352 were mapped to chromosome 12, 118 that were assigned to 12p, and 33 that were further assigned to 12p11-p12. On the basis of the comparison of sequences mapped to 12p in several databases (UniGene, UniSTS, GeneMap'98, GeneMap'99, GeneBridge 4, OMIM Gene Map, the working draft sequence build 22, the AECOM 12p map, and TIGR<sup>6</sup>), we estimated that 12p contains coding sequences of ~400 genes.<sup>7</sup> The EST array used therefore represents approximately one-third of the 12p genes.

**Chromosomal CGH Analysis.** Thirty-eight TGCT biopsy samples (10 seminomas and 27 nonseminomas) and the OCCL were subjected to chromosomal CGH analysis. The 12p amplification status in 34 of these tumors by CGH has been published elsewhere (23). High-level amplification at 12p11-p12 was observed in 4 seminomas (187A, 192A, 287B, and 427A) and the OCCL cell line (Fig. 1).

**Array CGH Analysis.** We performed array CGH analysis on the DNAs from the five samples that showed 12p11-p12 amplification by chromosomal CGH. As controls, we similarly analyzed tumor 227A, a seminoma that did not show 12p11-p12 amplification and placenta. Test (tumor) and reference (placenta) gDNAs were labeled with Cy3 and Cy5, respectively, and cohybridized to the EST array. The signals obtained after laser excitation of the dyes were digitized (Fig. 2), and the raw data (median feature pixel intensity with the median local background intensity subtracted at each wavelength) were then subjected to statistical analysis. To correct for systematic errors introduced by the intensity-dependent dye efficiencies, the hybridization signal data from each slide were normalized using a local regression of the log-ratio variable  $Y = \log_2(G/R)$  versus the log-product  $X = \log_{10}(R \times G)/2$  ( $R$  and  $G$  represent the intensities of the Cy3 and Cy5, respectively; Ref. 20). Examples of pre- and postnormalized data are shown in Fig. 3, A and B. In practice, the noise structure and magnitude of the signals varied from slide to slide, as can be seen by

<sup>5</sup> Internet address: sequence.aecom.yu.edu/bioinf/funcgenomic.html.

<sup>6</sup> Internet address: www.tigr.org.

<sup>7</sup> V. Bourdon, et al., manuscript in preparation.



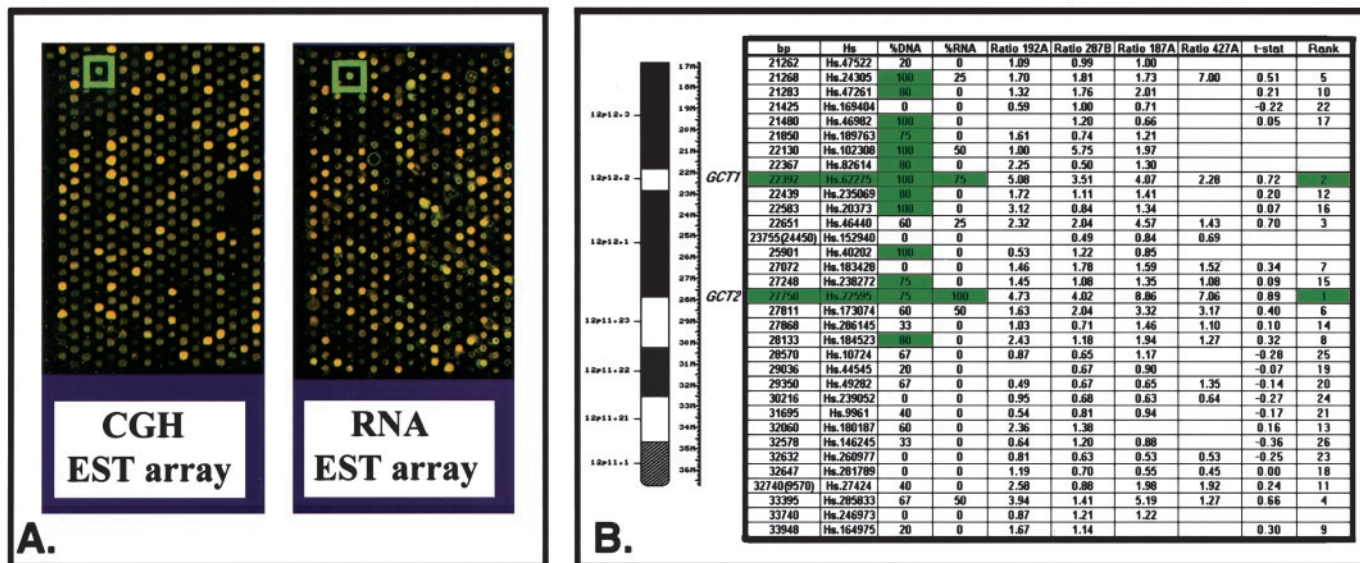


Fig. 2. Combined CGH and Expression EST array analysis. A, pseudocolored views of an identical portion of the EST arrays (392 of the 8254 features spotted). Left: gDNA from tumor 187A (green) and placenta (red) were cohybridized. Right: cDNAs from tumor 187A (green) and normal testis (red) were cohybridized. In both panels, the green box indicates the common amplified and overexpressed EST. B, summary of the results. Green boxes represent the ESTs considered to be amplified and overexpressed after statistical analysis. Thirteen were amplified (LR/SD > 3 in  $\geq 75\%$  of the tumor tested). Two of them (*GCT1* and *GCT2*) were also found overexpressed (LR/SD > 2 in  $\geq 75\%$  of the tumor tested). All of the ESTs on the array that mapped at 12p11-p12 are shown from the distal (top) to proximal (bottom) of 12p. The positions on the chromosomal arm (bp) and UniGene cluster number (Hs.) are given. The expression array ratio values are given in the ratios column. The percentage of tumors showing significant amplification and overexpression of a particular EST is given in the %DNA column. The percentage of tumors showing significant amplification and overexpression of a particular EST is given in the %RNA column. For comparison, the last two columns show the two-sample t-statistics computed following the approach described by Hughes *et al.* (26). The ranks are taken within the ESTs mapped at 12p11-p12; empty fields correspond to insufficient data for determining the t-parameter. Notice that when using the t-statistics scheme, the two first candidates (green) are identical to the one found using our approach.

the variation in the shapes of the point clouds (Fig. 3A). It was therefore important to construct an indicator to identify ESTs that exhibited significant signal deviation from normal in a given slide. To this end, we computed the intensity-dependent (local) variance  $\sigma(X_N)^2$  from a local regression of  $Y_N^2$  versus  $X_N$  after normalization ( $X_N$  and  $Y_N$  represent the normalized X and Y variables). We observed that the variable  $|Y_N|/\sigma(X_N)$  closely followed a Gaussian distribution in the control experiments, as shown by the straight quantile-quantile plot (Fig. 3C). As expected for tumor versus reference comparisons, these distributions were slightly longer-tailed on one side of the plot, reflecting biological amplification (indicated by the arrows in Fig. 3C). Because the bulk of the distributions were nearly Gaussian, it was felt justified to attribute significance to amplified or overexpressed ESTs according to the values of  $Y_N/\sigma(X_N)$  (LR/SD), independently for each slide.

After preprocessing the data as described above, an EST was considered amplified when the value LR/SD was >3 (0.1% probability of a false-positive call). Among the 8254 sequences on the array, 111 were amplified in one or more of the tumors tested. Of these, 57 (51%) were mapped to chromosome 12, and of these, 49 (86%) were mapped to 12p. This number was highly significant when compared with the number of amplified ESTs mapped to regions of the genome other than 12p, thus ruling out the possibility of finding the 12p-amplified ESTs by chance alone ( $P < 10^{-10}$ ). When selecting only for sequences showing amplification in  $\geq 75\%$  of the acceptable hybridization data from the five tumors tested, 19 sequences were identified, of which 13 (68%) were mapped to 12p11-p12 (Fig. 2). None of these ESTs showed a significant amplification in the control experiments, which comprised three placenta versus placenta and the placenta versus tumor 227A hybridizations (LR/SD < 1.13).

**Array Expression Analysis.** We were able to isolate good quality RNA from the four GCTs tested by array CGH (187A, 192A, 287B, and 427A). Test (tumor) and reference (normal testis) RNAs were reverse-transcribed, labeled, and cohybridized to the EST array as

described above. The data were normalized also as described above, and the expression data for the 13 amplified ESTs mapped to 12p11-p12 were analyzed (Fig. 2). Considering ESTs to be overexpressed when the LR/SD was >2 in  $\geq 75\%$  of the acceptable hybridization data from the four tumors (2.3% probability of a false-positive call), we found two amplified and overexpressed sequences belonging to the UniGene<sup>8</sup> clusters Hs.22595 and Hs.62275. We designated these genes as *GCT1* and *GCT2*. Neither of these amplified and overexpressed ESTs showed significant overexpression in five independent control experiments comprising normal testis versus normal testis RNA (cDNA) hybridizations (LR/SD < 1.3).

**Southern Blot Analysis.** We performed Southern blot analysis of the five tumors that were subjected to array CGH, using *GCT1* and *GCT2* probes, to confirm their genomic copy number status (Fig. 4). The *GCT1* and *GCT2* probes were selected from the 3' end of their cDNAs, and the probe D2S48, mapped to chromosome 2, was used to quantify the amount of DNA loaded. The amplification of *GCT1* and *GCT2* observed by array CGH was confirmed by Southern blot analysis, their copy number ranged from 10.2 to 26.2.

**Expression of *GCT1* and *GCT2* in Normal and Tumor Tissues.** *GCT1* has 17 exons spanning ~32.8 Kb of genomic sequence. The ESTs were assembled in a UniGene (24) cluster Hs.22595 and a TIGR cluster THC847867 (THC Version 8.0; Ref. 25) and included cDNA derived from germ cells, testis, and GCTs. The corresponding human hypothetical protein sequence (DKFZp434L032.1, GenBank accession T46457) showed 70–88% identity with protein sequences of different species grouped in a tentative orthologue group TOGA96125 (TOGA<sup>9</sup> version 5.0). Blast analysis of this protein sequence against FlyBase<sup>10</sup> showed 40% identity with the *Mat89Bb* gene (*ovary2*). This gene, located on the right arm of the third chromosome (89B16-

<sup>8</sup> Internet address: www.ncbi.nlm.nih.gov/UniGene.

<sup>9</sup> Internet address: www.tigr.org/tdb/tooga.shtml.

<sup>10</sup> Internet address: flybase.bio.indiana.edu.

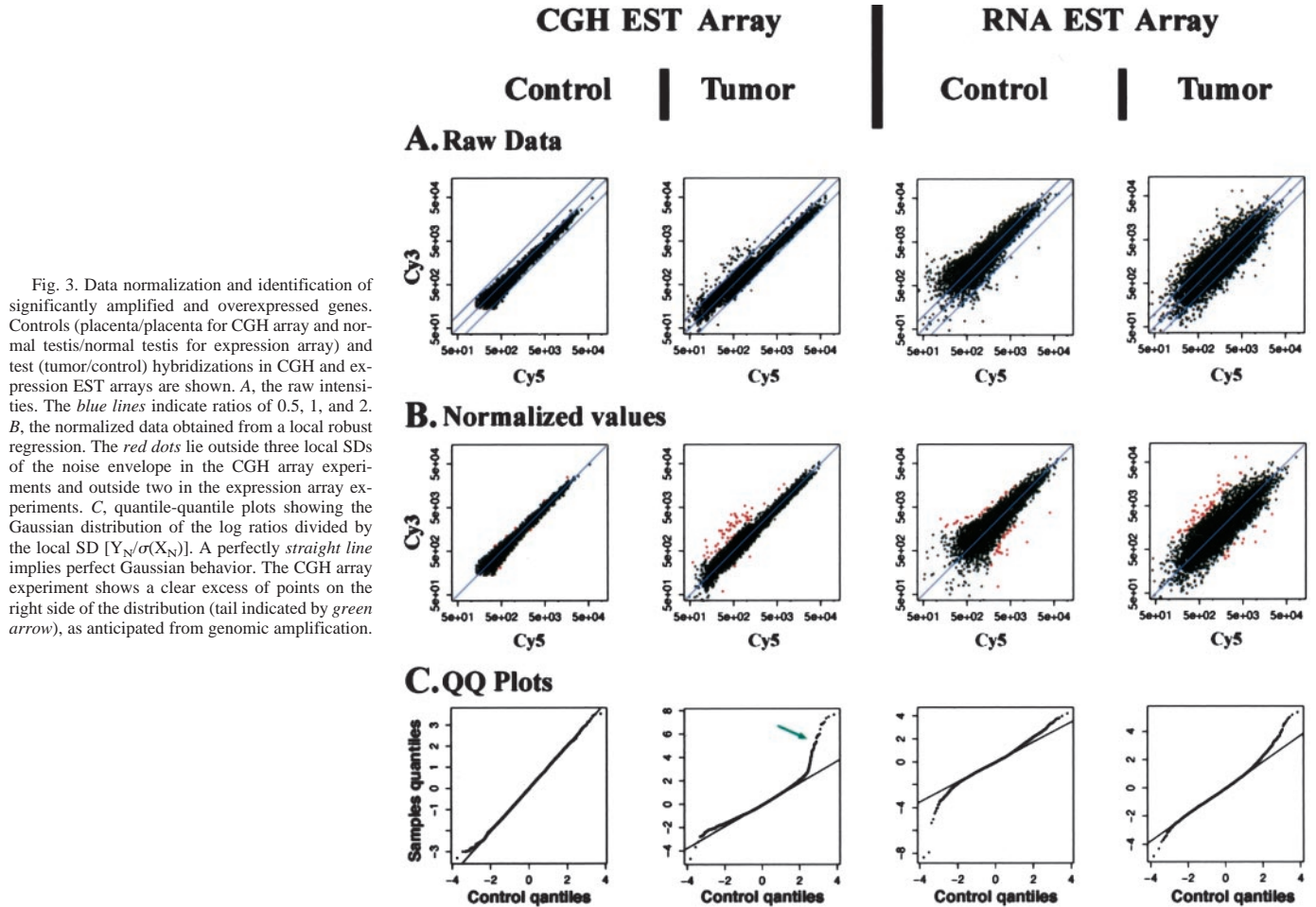


Fig. 3. Data normalization and identification of significantly amplified and overexpressed genes. Controls (placenta/placenta for CGH array and normal testis/normal testis for expression array) and test (tumor/control) hybridizations in CGH and expression EST arrays are shown. A, the raw intensities. The blue lines indicate ratios of 0.5, 1, and 2. B, the normalized data obtained from a local robust regression. The red dots lie outside three local SDs of the noise envelope in the CGH array experiments and outside two in the expression array experiments. C, quantile-quantile plots showing the Gaussian distribution of the log ratios divided by the local SD [ $Y_N/\sigma(X_N)$ ]. A perfectly straight line implies perfect Gaussian behavior. The CGH array experiment shows a clear excess of points on the right side of the distribution (tail indicated by green arrow), as anticipated from genomic amplification.

18), encodes a protein that is expressed in adult ovary (follicle cells, nurse cell, and oocyte) and in adult testis. *GCT2* has five exons spanning ~14.5 Kb of genomic sequence. The ESTs were assembled in a UniGene cluster Hs.62275 and a TIGR cluster THC758883 and included cDNA derived from germ cells and GCTs. The corresponding human hypothetical protein sequence (DKFZp761G2423.1, GenBank accession T46908) showed 60–91% identity with protein sequences of different species grouped in TOGA 102326 and TOGA 86129.

A RT-PCR analysis of a panel of human tissues was performed

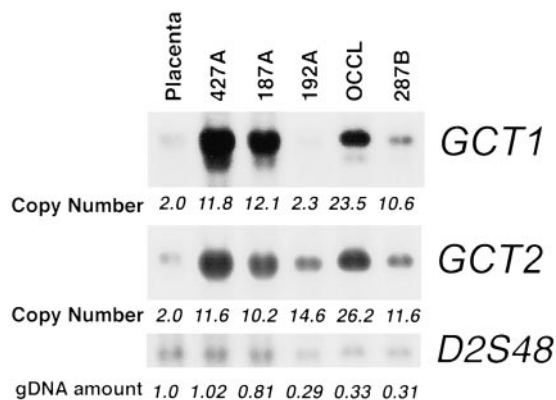


Fig. 4. Southern blot analyses of *GCT1* and *GCT2* in the 4 seminomas (427A, 187A, 287B, and 192A) and the OCCL with 12p11-p12 chromosomal amplification plus placenta. Tumor 192A did not show amplification of *GCT1*, possibly indicating a partial deletion. The copy number and relative gDNA amount per lane are given.

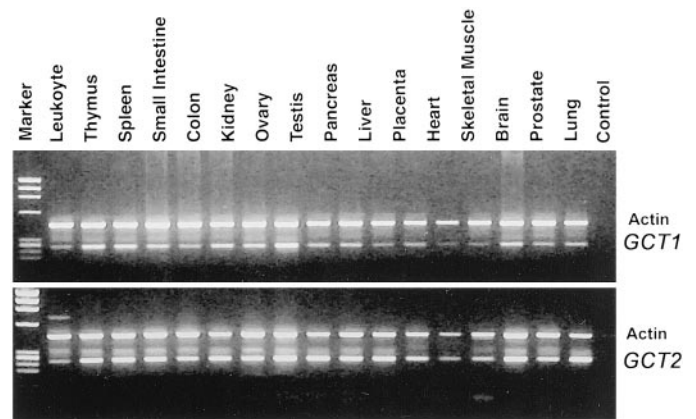


Fig. 5. RT-PCR analyses of *GCT1* and *GCT2* in a tissue panel.

using *GCT1* and *GCT2* primer pairs as described in “Materials and Methods.” Both genes were expressed in all of the tissues tested, including testis and ovary (Fig. 5). Their expression was then assayed in a panel of 47 GCT samples by semiquantitative RT-PCR that included 22 nonseminomas (10 cell lines and 12 tumor samples), 24 seminomas, and the OCCL cell line using the same primer pairs. *GCT1* was overexpressed compared with normal testis (1.27× to 2.83×) in 92% of the seminomas and 18% of the nonseminomas and the OGCT, whereas *GCT2* was overexpressed (1.38× to 3.07×) in 71% of the seminomas and 32% of the nonseminomas and the OGCT. These data confirm that both amplified genes are overexpressed



significantly in seminomas and less so in nonseminomas. This is consistent with the chromosomal CGH data, which showed 12p11-p12 amplification predominantly in seminomas.

## DISCUSSION

Increase of gene dosage by DNA amplification of chromosomal regions is a frequent genetic alteration during tumorigenesis. Therefore, a high throughput screening method for identification of amplified as well as overexpressed sequences is of considerable value for the understanding of the molecular mechanisms of cancer, especially those that underlie progression and clinical behavior. Several studies have shown that many human neoplasms, including GCTs, exhibit overrepresentation of 12p sequences. Chromosomal CGH analysis of 38 GCTs and the OCCL showed high-level amplification at 12p11-p12 in 4 seminomas and in OCCL. We subjected these samples to a combined array CGH and expression analysis using an array spotted with 8254 ESTs. We found it important to normalize the data using a local regression technique as systematic errors introduced by the intensity-dependent dye efficiencies can lead to serious misinterpretation of data when using global normalization. In addition, local normalization is a necessary step before evaluating the local SD of the log ratios, which we used as criteria for attributing significance of variation. This criterion was justified because the log ratios divided by the local SD showed close Gaussian distributions and enabled to account for the intensity dependence of the noise clouds. Thus, a ratio value may be considered as highly significant at the highest intensities and insignificant toward the lower intensity end. The approach used here shares some similarity with a previously published approach (26), except that the local SD was determined in this analysis empirically without an *a priori* noise model. Because of the relatively few samples (four tumors and five controls) in the expression array, we were reluctant to use a t-statistics-based approach as our primary measure for identifying overexpressed ESTs. Nevertheless, the ranks of the t-statistics, as reported in Fig. 2, show a good correlation with results from the approach used. Merging data from five tumors, the analysis scheme used lead to the identification of amplified ESTs located at 12p11-p12, with very low false-positive rates. We found 13 ESTs with highly significant values ( $LR/SD > 3$ ) in  $\geq 75\%$  of the tumors giving a high biological as well as statistical confidence to the results. Two of these sequences (*GCT1* and *GCT2*) were also overexpressed in  $\geq 75\%$  of the tumor samples and therefore were considered as candidate amplified and overexpressed genes. We blasted (27) the 33 ESTs mapped to 12p11-p12 against the human draft genome sequences database (*Homo sapiens* build 24) and mapped them using Entrez Genome Map Viewer.<sup>11</sup> These two ESTs were mapped at ~23 and 29 Mb, respectively, from the distal 12p.

Several candidate genes on 12p have previously been studied or suggested because of the overall copy number increase of this chromosomal arm consistently associated with TGCTs. These included *SOX5*, *LRMP* (*JAW1*), *KRAS2*, *FGF6*, *TEL*, *CDKN1B* (*KIP1*), *PTHLH*, *LDHB*, *TNFRSF1A* (*TNFR1*), *LGS*, *SSPN* (*KRAG*), *ITPR2*, *CD69*, *CD94*, *GDF3*, *GRIN2B*, *PYGL*, *IAPP*, and *RECQL*. The array used by us contained ESTs representing *ITPR2*, *PYGL*, *SSPN*, *RECQL*, *TNFRSF1A*, *CDKN1B*, *TEL*, *KRAS2*, and *LRMP*. Among these, *RECQL*, *ITPR2*, *KRAS2*, and *LRMP* were found amplified; however, none of them were overexpressed by array analysis.

The two amplified and overexpressed sequences identified in this study were designated here as *GCT1* and *GCT2*. These novel genes code for hypothetical proteins without known functions. Both genes were expressed in all normal human tissues tested, including testis and

ovary. In a semiquantitative RT-PCR analysis of 47 GCT samples, both genes showed overexpression more frequently in seminomas than in nonseminomas consistent with chromosomal CGH analysis.

These studies thus demonstrate the feasibility of identifying candidate amplified and overexpressed genes in large regions of chromosomal amplification using EST arrays and robust data analysis methods. They identified here, for the first time, two novel genes, mapped at 12p, amplified and overexpressed. The elucidation of role of the *GCT1* and *GCT2* in normal and tumor cell biology is in progress.

## ACKNOWLEDGMENTS

We thank Katerina Dyomina for expert technical assistance.

## REFERENCES

- Schwab, M. Oncogene amplification in solid tumors. *Semin. Cancer Biol.*, 9: 319–325, 1999.
- Gray, J. W., and Collins, C. Genome changes and gene expression in human solid tumors. *Carcinogenesis* (Lond.), 21: 443–452, 2000.
- 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* (Wash. DC), 258: 818–821, 1992.
- Solinasto-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., Segraves, 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.*, 20: 207–211, 1998.
- Albertson, D. G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W., and Pinkel, D. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat. Genet.*, 25: 144–146, 2000.
- Hui, A. B., Lo, K. W., Yin, X. L., Poon, W. S., and Ng, H. K. Detection of multiple gene amplifications in glioblastoma multiforme using array-based comparative genomic hybridization. *Lab. Invest.*, 81: 717–723, 2001.
- Daigo, Y., Chin, S. F., Goringe, K. L., Bobrow, L. G., Ponder, B. A., Pharoah, P. D., and Caldas, C. Degenerate oligonucleotide primed-polymerase chain reaction-based array comparative genomic hybridization for extensive amplicon profiling of breast cancers: a new approach for the molecular analysis of paraffin-embedded cancer tissue. *Am. J. Pathol.*, 158: 1623–1631, 2001.
- Heiskanen, M. A., Bittner, M. L., Chen, Y., Khan, J., Adler, K. E., Trent, J. M., and Meltzer, P. S. Detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer Res.*, 60: 799–802, 2000.
- 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.
- Monni, O., Barlund, M., Mousseis, S., Kononen, J., Sauter, G., Heiskanen, M., Paaola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711–5716, 2001.
- Chaganti, R. S., and Houldsworth, J. Genetics and biology of adult human male germ cell tumors. *Cancer Res.*, 60: 1475–1482, 2000.
- 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: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107–1123, 1998.
- Suijkerbuijk, R. F., Sinke, R. J., Weghuis, D. E., Roque, L., Forus, A., Stellink, F., Siepmann, A., van de Kaa, C., Soares, J., and Geurts van Kessel, A. Amplification of chromosome subregion 12p11.2-p12.1 in a metastasis of an i(12p)-negative seminoma: relationship to tumor progression? *Cancer Genet. Cytogenet.*, 78: 145–152, 1994.
- Mostert, M. C., Verkerk, A. J., van de Pol, M., Heighway, J., Marynen, P., Rosenberg, C., van Kessel, A. G., van Echten, J., de Jong, B., Oosterhuis, J. W., and Looijenga, L. H. Identification of the critical region of 12p over-representation in testicular germ cell tumors of adolescents and adults. *Oncogene*, 16: 2617–2627, 1998.
- Roelofs, H., Mostert, M. C., Pompe, K., Zafarana, G., van Oorschot, M., van Gorp, R. J., Gillis, A. J., Stoop, H., Beverloo, B., Oosterhuis, J. W., Bokemeyer, C., and Looijenga, L. H. Restricted 12p amplification and RAS mutation in human germ cell tumors of the adult testis. *Am. J. Pathol.*, 157: 1155–1166, 2000.
- Heidenreich, A., Srivastava, S., Moul, J. W., and Hofmann, R. Molecular genetic parameters in pathogenesis and prognosis of testicular germ cell tumors. *Eur. Urol.*, 37: 121–135, 2000.
- Rosenberg, C., Van Gorp, R. J., Geelen, E., Oosterhuis, J. W., and Looijenga, L. H. Overrepresentation of the short arm of chromosome 12 is related to invasive growth of human testicular seminomas and nonseminomas. *Oncogene*, 19: 5858–5862, 2000.
- Schmidt, B. A., Rose, A., Steinhoff, C., Strohmeyer, T., Hartmann, M., and Ackermann, R. Up-regulation of cyclin-dependent kinase 4/cyclin D2 expression but

<sup>11</sup> Internet address: [www.ncbi.nlm.nih.gov/genome/seq](http://www.ncbi.nlm.nih.gov/genome/seq).

- down-regulation of cyclin-dependent kinase 2/cyclin E in testicular germ cell tumors. *Cancer Res.*, *61*: 4214–4221, 2001.
20. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, *30*: e15, 2002.
  21. Rao, P. H., Murty, V. V., Louie, D. C., and Chaganti, R. S. Nonsyntenic amplification of MYC with CDK4 and MDM2 in a malignant mixed tumor of salivary gland. *Cancer Genet. Cytogenet.*, *105*: 160–163, 1998.
  22. Cheung, V. G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G. Making and reading microarrays. *Nat. Genet.*, *21*: 15–19, 1999.
  23. Rao, P. H., Houldsworth, J., Palanisamy, N., Murty, V. V., Reuter, V. E., Motzer, R. J., Bosl, G. J., and Chaganti, R. S. Chromosomal amplification is associated with cisplatin resistance of human male germ cell tumors. *Cancer Res.*, *58*: 4260–4263, 1998.
  24. Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannikulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T. J., *et al.* A gene map of the human genome. *Science (Wash. DC)*, *274*: 540–546, 1996.
  25. Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Perteu, G., Sultana, R., and White, J. The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.*, *29*: 159–164, 2001.
  26. Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraburty, K., Simon, J., Bard, M., and Friend, S. H. Functional discovery via a compendium of expression profiles. *Cell*, *102*: 109–126, 2000.
  27. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.*, *215*: 403–410, 1990.