 Human Molecular Genetics, 2001, Vol. 10, No. 1 77–83

X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications

Ralf Sudbrak[1,+], Georg Wieczorek[1,+], Ulrike A. Nuber[1], Wolfgang Mann[1,§], Roland Kirchner[1,§], Fikret Erdogan[1], Carolyn J. Brown[2], Doris Wöhrle[3], Peter Sterk[4], Vera M. Kalscheuer[1], Wolfgang Berger[1], Hans Lehrach[1] and Hans-Hilger Ropers[1,¶]

[1]Max-Planck Institute for Molecular Genetics, Ihnestrasse 73, D-14195, Berlin, Germany, [2]Department of Medical Genetics, University of British Columbia, Vancouver, Canada and [3]Department of Medical Genetics, University of Ulm, Germany and [4]European Bioinformatics Institute, Hinxton, Cambridge, UK

Received 8 September 2000; Revised and Accepted 2 November 2000

Mutant alleles are frequently characterized by low expression levels. Therefore, cDNA array-based gene expression profiling may be a promising strategy for identifying gene defects underlying monogenic disorders. To study the potential of this approach, we have generated an X chromosome-specific microarray carrying 2423 cloned cDNA fragments, which represent up to 1317 different X-chromosomal genes. As a prelude to testing cell lines from patients with X-linked disorders, this array was used as a hybridization probe to compare gene expression profiles in lymphoblastoid cell lines from normal males, females and individuals with supernumerary X chromosomes. Measurable hybridization signals were obtained for more than half of the genes represented on the chip. A total of 53 genes showed elevated expression levels in cells with multiple X chromosomes and many of these were found to escape X-inactivation. Moreover, the detection of a male-viable deletion encompassing three genes illustrates the utility of this array for the identification of small unbalanced chromosome rearrangements.

INTRODUCTION

Since their first description <5 years ago (1), cDNA microarrays have become indispensable tools for monitoring gene expression profiles in a wide variety of species and tissues. In cultured cells, expression profiling with cDNA arrays has led to the identification of genes that are regulated in a coordinated manner during the cell cycle, under nutritional constraints or after treatment with hormones (2–4). cDNA-based gene expression profiling in normal and malignant tissues is beginning to shed light on the molecular mechanisms involved in tumorigenesis and tumour progression (5,6). Similar approaches are being employed in cell culture or animal models for drug testing (7,8) and another emerging field is the search for secondary effects of defined disease genes on the expression profile of cells to identify downstream target genes and to elucidate the pathogenesis of hereditary disorders (9). In contrast, cDNA array-based expression profiling has been employed in a sole publication so far to search for primary defects underlying genetic disorders (10).

To date, causative gene defects have been found for only 20% of the >5000 monogenic disorders known in man (11) and thus the molecular elucidation of Mendelian disorders is still a major challenge. As sequencing of the human genome approaches completion, mutation screening in candidate genes has replaced positional cloning as the strategy of choice for identifying disease genes, but for both approaches, precise mapping of the relevant genetic defect is a crucial prerequisite. Since most of the Mendelian disorders that have not yet been elucidated are rare or genetically heterogeneous, however, linkage intervals are often wide and may encompass hundreds of genes. Therefore, identification of the relevant disease gene by systematic mutation screening in patients may be very tedious and time-consuming even if the chromosomal location of the underlying defect is already known.

In contrast, the strategy to use chromosome-specific cDNA arrays to search for primary changes in the gene expression profile of patients is conceptually simple and very fast. There are, however, several problems that might limit the practical use of this approach, e.g. the fact that gene expression profiling in patients is confined to accessible cells and tissues such as blood cells and fibroblasts. So far, little is known about the proportion of all human genes that are expressed in these cells at sufficiently high levels to allow quantification with cDNA arrays, and it is not clear whether this approach is also feasible to search for gene defects that are primarily manifest in other tissues such as the brain. The most fundamental problem of this strategy is its limitation to mutations that significantly alter the intracellular concentration of the relevant mRNA transcript (12,13).

[+]These authors contributed equally to this work
[§]Present address: MWG Biotech AG, Anzinger Strasse 7, D-85560 Ebersberg, Germany
[¶]To whom correspondence should be addressed. Tel: +49 30 8413 1240; Fax: +49 30 8413 1383; Email: ropers@molgen.mpg.de

To address some of these problems, we have employed an X chromosome-specific cDNA microarray representing up to 1317 X-chromosomal genes, to study their expression in Epstein–Barr virus (EBV)-transformed lymphoblastoid cells from normal males and females and from individuals with supernumerary X chromosomes. The human X chromosome was chosen as a model for this approach because of the large number of known X-linked disorders. Moreover, in affected males, aberrant expression of X-linked genes is easier to detect because it is not 'diluted' by heterozygosity. X-chromosomal genes were identified by computer-assisted comparison of cDNA sequences derived from UniGene expressed sequence tag (EST) clusters (http://www.ncbi.nlm.nih.gov:80/UniGene/) with genomic sequences from the human X chromosome, including ~600 EST clusters and genes that had been mapped to the X chromosome previously. For each of these clusters, one or several probes were selected from the IMAGE collection of cDNA clones (14) and were deposited on activated glass slides to generate the X chromosome-specific cDNA microarray.

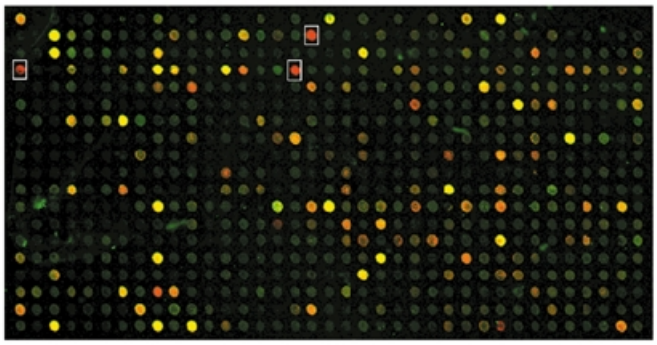

Figure 1. Segment of the X chromosome-specific cDNA array hybridized to Cy5-labelled cDNA from a lymphoblastoid cell line with multiple X chromosomes and Cy3-labelled cDNA from a male lymphoblastoid cell line. Pseudo-coloured red spots indicate genes which are overexpressed in cells with multiple X chromosomes. *XIST* cDNA clones are marked by a white box.

RESULTS

Differential expression of X-chromosomal genes in male and female cells

By reverse transcription of mRNA from male and female lymphoblastoid cell lines, cDNA was generated and labelled with fluorochromes Cy3 (pseudo-coloured green) and Cy5 (pseudo-coloured red), respectively. Differentially labelled cDNA preparations were co-hybridized to the arrays in several independent experiments. Consistently, close to 50% of the 2423 X-specific EST clones on the chip yielded hybridization signals exceeding the background, defined as two standard deviations above the average signal of 15 plant genes which were used as negative controls. These results were independent from the fluorochrome employed. These data indicate that in lymphoblastoid cells, expression profiling is possible for at least 50% of all genes represented on the chip. In a limited number of analogous experiments with cDNA from male and female fibroblasts, similar results were obtained (data not shown). Further analysis revealed that certain genes showed higher hybridization signals with female than with male cDNA, including *XIST* and several genes that had been shown previously to escape from X-inactivation (15–19).

Many genes showing dosage-dependent expression escape from X-inactivation

To study this finding in more detail, analogous analyses were performed with Cy3-labelled cDNA from five different human lymphoblastoid cell lines, carrying four or five X chromosomes. In duplicate experiments, pools of Cy5-labelled cDNA from five cell lines with normal male and normal female karyotypes, respectively, were co-hybridized to the microarray as internal standard (Fig. 1). In this way, 20 different data sets were obtained for each of the 2423 X-chromosomal cDNA clones on the chip. After background subtraction and normalization of the data by comparison with signal intensities of 49 clones representing human housekeeping genes, global analysis revealed that in cell lines with supernumerary X chromosomes, hybridization signals were skewed towards higher values. To identify individual genes showing elevated expression levels in cells with multiple X chromosomes, areas including 95% of the signal intensity ratios (i.e. the ratios of signal intensities obtained with cDNA from cells with four or five X chromosomes, versus hybridization intensities with cDNA from normal male or female cells) were calculated. Signal intensity ratios that fell outside the 95% thresholds, indicating elevated expression levels in cells with supernumerary X chromosomes in at least 10 of the 20 data sets, were considered as biologically relevant; this was the case for 88 different clones corresponding to a minimum of 59 different EST clusters. The identity of 53 of the 59 clusters could be confirmed by sequencing of the respective IMAGE clones (listed in Table 1).

To substantiate our assumption that many of these genes escape from X-inactivation, RT–PCR experiments were performed to amplify transcripts of these genes from five independent human–rodent somatic cell hybrids carrying a single inactive human X chromosome. Of 19 PCR primer pairs, 11 representing different EST clusters from this list yielded human-specific amplification products on mRNA from one or several of these somatic cell hybrids (Table 1). Three of these 11 genes were already known to escape from X-inactivation, whereas none had been reported previously as being subject to X-inactivation. Altogether, prior information about the X-inactivation state was available for 19 of the 53 'outliers'. Of these, 12 escaped from X-inactivation, 1 showed a polymorphic inactivation pattern, and 6 were not expressed when carried on inactive X chromosomes. Thus, 63% (12 of 19), or ~33, of the genes showing elevated mRNA concentrations in cells with supernumerary X chromosomes may escape from X-inactivation.

Regional localization of genes that are overexpressed in cells with supernumerary X chromosomes

Genes that escape from X-inactivation are thought to cluster on the short arm of the human X chromosome, which is evolutionarily 'younger' than the long arm (17). Of the 53 genes identified in this study, 27 map to Xp, whereas 26 map to Xq. Of the 27 genes on Xp, 11 had previously been reported to escape from X-inactivation and none had been found to be inactivated. In

contrast, only 2 of the 26 genes on Xq escape from X-inactivation or have a polymorphic inactivation pattern, respectively, but 6 of the 26 genes on Xq are known to be subject to X-inactivation. Our RT–PCR analyses were in keeping with these findings: all genes expressed in at least three of the five rodent–human somatic cell hybrids carrying an inactive human X chromosome were located on Xp, whereas none of the genes on Xq was expressed in more than two of these cell lines (Table 1).

Identification of disease-causing mutations with cDNA microarrays

To demonstrate the utility of the X chromosome-specific cDNA microarray for detecting molecular defects in patients with X-linked disorders, EBV-transformed lymphoblastoid cells from a mentally retarded male patient with an interstitial deletion on the long arm of the X chromosome (20) were screened for aberrant gene expression patterns. mRNA from this cell line was reverse transcribed, labelled with Cy3 and co-hybridized to the array with Cy5-labelled control cDNA, derived from normal male lymphoblastoid cells. Image analysis revealed differential hybridization signals for three genes (*P2Y10*, *ITM2A* and *SH3BGRL*), indicating lack of the corresponding transcripts in the patient's cell line (Fig. 2). Subsequent PCR analyses confirmed that all three loci are indeed deleted in the patient (data not shown). In contrast, *CHM*, *POU3F4* and *RSK4*, three genes known to be deleted in the patient, could not be identified as such by expression profiling because of their low expression levels.

DISCUSSION

In this article, we have described the generation of a human X chromosome-specific cDNA microarray carrying probes from 1317 different EST clusters. In a first series of experiments we have used this array, the only one of its kind published to date, to systematically search for differences in the expression profiles of cultured cells from normal males, females and patients with supernumerary X chromosomes. In this way, we have identified 53 genes showing significantly elevated expression in cells with multiple X chromosomes. Two lines of evidence support the assumption that many of these are not subject to X-inactivation. (i) Of these 53 genes, 12 were already known to escape from inactivation, whereas normal inactivation had been reported for only 6; and (ii) we could show that 11 of 19 genes with unknown X-inactivation status are expressed in somatic cell hybrids carrying an inactive human X chromosome but no active human X chromosome. However, according to our data, escape from X-inactivation may not be the only reason why the expression of some of the X chromosomal genes represented on microarray seems to be dosage dependent. Instead, the over-expression of certain X-chromosomal genes in cells with supernumerary X chromosomes may be due to other regulatory mechanisms.

In the present study, some of the non-inactivated genes could not be identified as such since they were not represented as IMAGE clones on the array or showed no expression in lymphoblastoid cells. Others may have been overlooked because our selection criteria (i.e. significantly elevated expression in at least 10 of 20 independent experiments) are too strict, and it may be possible to identify additional genes by modest relaxation of these criteria. Indeed, several of the previously described non-inactivated genes show elevated expression in 8 or 9 of 20 experiments. It may be that some of these genes have been missed because of polymorphic inactivation patterns, i.e. because they are expressed on some inactive X chromosomes but not on others (17,21) or, alternatively, because they are much less active when carried on an inactive X chromosome. Semi-quantitative RT–PCR studies have revealed that expression levels of genes on inactive X chromosomes are often far lower than on active X chromosomes (21). For the steroid sulphatase (STS) gene, one of the first genes known to escape from X-inactivation, previous enzyme activity measurements have shown that in female somatic cells with an active and an inactive X chromosome, enzyme levels are only 30% higher than in male cells (22). Our current experiments have identified the STS gene as an 'outlier', albeit only just: in 11 of 20 data sets obtained by cDNA-based expression profiling, STS gene expression levels were significantly higher in cells with multiple X chromosomes than in male and female controls. Assuming that transcription rates of STS genes on inactive X chromosomes correspond to 0.3 times the transcription rates on active X chromosomes, signal intensities in cell lines with four X chromosomes should be 1.46–1.9 times higher than in normal female or male cells, respectively, which agrees well with the ratios observed (1.43 and 1.70, respectively).

Not unexpectedly, the most striking results were obtained for the *XIST* gene, which yielded significantly elevated signal intensity ratios in 20 of 20 data sets. Since this gene is only expressed on inactive X chromosomes, transcription levels in female cells should be directly proportional to the total number of X chromosomes minus one. In cells with four or five X chromosomes, intracellular mRNA concentrations should therefore be 3–4 times higher than in normal female cells (assuming that mRNA concentrations are proportional to the transcription rate). Therefore, these analyses suggest that 3-fold differences in the expression level of individual genes can be reliably detected even in single experiments.

Thus, although originally designed as pilot experiments for gene expression profiling in patients with X-linked disorders, these studies have opened a new and fast route to completion of the map of non-inactivated genes on the human X chromosome. At the same time, we have shown here that with mRNA from lymphoblastoid cells, expression levels can be monitored for at least 50% of all X-linked genes represented on the chip. In part, these results may be due to over-representation of ubiquitously expressed genes among known ESTs, the primary source of cDNA probes used for this array. However, it should be possible to attain even higher coverage of X-chromosomal genes by using mRNA from other cells and tissues, such as cultured keratinocytes or hair bulbs, that are routinely accessible, or can be obtained by biopsies in a diagnostic context. Moreover, the number of genes represented on the chip can be further expanded. With sequencing of the human genome approaching completion, novel genomic sequences from the X chromosome are being released every day, and the generation of a second, larger X chromosome-specific cDNA array is already in progress.

After having demonstrated its usefulness for identifying X chromosomal deletions, this array is being employed to systematically search for specific changes in the gene expression

Table 1. Genes that showed elevated expression levels in cells with supernumerary X chromosomes in at least 10 of 20 independent microarray hybridization experiments (indicated as frequency)

UniGene cluster	IMAGE ID	Gene description	Cytogenic band	Frequency	RT–PCR	Reference
Hs.9933	1963103	*ALTE* (*TRAMP*)	Xp22.33, Yp11	12		Active (15)
Hs.164280	2046256	*SLC25A6* (*ANT3*)	Xp22.32, Yp	13		Active (16)
Hs.79876	52726	*STS*	Xp22.32	11		Active (17)
Hs.54941	1915452	*PHKA2*	Xp22.2–p22.1	12		
Hs.58521	1892272		Xp22	16		
Hs.88764	46405	*MSL3L1*	Xp22.3	11	3 of 5	
Hs.269839	1565590		Xp22	10	4 of 5	
Hs.174038	2000675	*SEDL*	Xp22	14	5 of 5	Active (17)
Hs.78361	26820	*GPM6B*	Xp22.2	13	1 of 5	
Hs.269933	628599		Xp22	10	5 of 5	
Hs.19978	926000	*ARSD*	Xp22	15		Active (17)
Hs.18625	248308	*MT-ACT48*	Xp22	10	0 of 5	
Hs.78991	685427	*DXF68S1E* (*GS1*)	Xp22.3	15		Active (18)
Hs.25625	713301		Xp22	13		
Hs.221797	35788		Xp22	10		
Hs.120769	488555		Xp22	11		
Hs.77578	645018	*USP9X*	Xp11.4	11	5 of 5	Active (17)
Hs.13980	2046395	*UTX*	Xp11.2	14		Active (17)
Hs.765	1952126	*GATA1*	Xp11.23	12		
Hs.211602	46800	*DXS423E* (*SMC1*)	Xp11.22–11.21	18	5 of 5	Active (17)
Hs.260100	144981		Xp11.2	11		
Hs.193363	2005914		Xp11.2	10		
Hs.231195	1640561		Xp11.2	10		
Hs.55823	1854668	*SMCX* (*XE169*)	Xp11.22–11.21	15		Active (17)
Hs.147916	469217	*DDX3*	Xp11.3–p11.23	12		Active (19)
Hs.103104	703526		Xp11.3	11	0 of 5	
Hs.192846	1554002		Xp11	10		

Continued opposite

profile of patients with disease-associated balanced rearrangements of the X chromosome (http://www.molgen.mpg.de). Subsequent experiments will include patients with a wide variety of X-linked conditions, including 150 samples from unrelated patients with unspecific and syndromic forms of X-linked mental retardation (MRX), which have been collected by the European MRX Consortium in France, Belgium, The Netherlands and Germany. Results of these studies should enable us to reliably estimate the proportion of clinically relevant mutations which significantly alter the intracellular concentration of specific transcripts and, in particular, to determine the possibilities and limitations of this approach for the identification of gene defects underlying disorders of brain function.

Although the X chromosome and X-chromosomal gene defects are particularly appealing targets for this approach, it should also be feasible for other chromosomes and chromosomally assigned genetic disorders. Above all, this holds for the analysis of balanced and unbalanced chromosome rearrangements, which are available through the Mendelian Cytogenetic Network (http://mcndb.imbg.ku.dk/index.php), and the construction of other chromosome-specific arrays is therefore envisaged. Given the encouraging results of recently published pilot experiments (23), these cDNA arrays may also be valuable diagnostic tools for the detection of small deletions and duplications in somatic cells and tumour tissues.

MATERIALS AND METHODS

Cells and cell culture

Female lymphoblastoid cell lines containing four or five X chromosomes were obtained from the Coriell Institute for Medical Research and grown according to the manufacturer's instructions.

Table 1. *Continued*

UniGene cluster	IMAGE ID	Gene description	Cytogenic band	Frequency	RT–PCR	Reference
Hs.16875	2029878		Xq13.2	11		Inactive (17)
Hs.289060	347682	*XIST*	Xq13	20		
Hs.606	1668404	*ATP7A*	Xq13.2–q13.3	12	0 of 5	
Hs.171723	1646942		Xq13–q21.1	11	0 of 5	Inactive (17)
Hs.75344	182219	*RPS4X/0*	Xq13.1	17		Active (17)
Hs.17109	1644560	*ITM2A*	Xq13.1–q21.2	13	2 of 5	
Hs.239663	70349	*MLLT7* (*AFX1*)	Xq13.1	15		Heterogeneous (17)
Hs.75621	1978913		Xq13.2–q21.2	10		
Hs.108029	244012		Xq11.1–q13.2	10	2 of 5	
Hs.176920	2174742		Xq21.1	10	0 of 5	
Hs.159156	663609	*PCDH11*	Xq21.3	10		
Hs.270538	1184177		Xq23	13		
Hs.5367	1977072		Xq25	11		
Hs.38264	209840		Xq25–q27	10	1 of 5	
Hs.48924	1930224	*KIAA0512*	Xq21.33–q22.3	10		Inactive (17)
Hs.157438	283233		Xq21.33–q22.3	10	2 of 5	
Hs.90998	1740854	*KIAA0128*, Septin-2 homologue	Xq24–q26	16	0 of 5	Inactive (17)
Hs.79172	50396	*SLC25A5* (*ANT2*)	Xq24–q26	12		Inactive (17)
Hs.152663	548086	*PAK3*	Xq22.3–q23	11		
Hs.250911	897821	*IL13RA1*	Xq23	10		Inactive (17)
Hs.47209	1552493		Xq23	14	0 of 5	
Hs.163929	212155		Xq23	11	0 of 5	
Hs.184993	503725		Xq23–q26	11		
Hs.274461	2012441		Xq25–q26	11		
Hs.144300	609419		Xq26.2–q27.2	13		
Hs.168213	1676248		Xq25–q26	13		

All clones were confirmed by sequencing.

Human–rodent somatic cell hybrid cell lines containing an inactive human X chromosome were provided by Dr Carolyn Brown (Department of Medical Genetics, University of British Columbia, Vancouver, Canada), Prof. Dr Peter Steinbach (Department of Medical Genetics, Ulm, Germany) and Dr Frans Cremers (Department of Human Genetics, Nijmegen, The Netherlands). Cells were grown in Dulbecco's modified Eagle's medium/F12 containing 2 mM L-glutamine, 10% fetal calf serum, and 100 U/ml each of penicillin and streptomycin and harvested at 80–90% confluency.

RNA isolation, cDNA synthesis and PCR amplification

Total RNA from lymphoblastoid cells and fibroblasts was isolated using the Trizol reagent (Gibco BRL), and labelled target cDNA was generated by direct incorporation of fluorescent nucleotide analogues during reverse transcription. For each labelling experiment, 25 µg of total RNA was used in an oligo(dT) primed reaction in the presence of 100 µM Cy3- or Cy5-dUTP (Amersham Pharmacia Biotech Europe), 200 µM dTTP and 500 µM dATP, dCTP and dGTP (Roche Molecular Biochemicals). The labelled cDNA was purified using the Qiaquick PCR purification kit (Qiagen). Deletion in patient RvD was investigated by PCR amplification using primer set: 18a3.1for, 5′-TTTAATCATAGCATACTACATAGCTCA-3′, and 18a3.1rev, 5′-AAGAAATCTATTTCATACCATGGAG-3′, for *PGK1*;
10f1for, 5′-TTTCCCATCCTGAGAAGCAC-3′, and 10f1rev, 5′-TCCAAAAGGCAGCAGAGACT-3′, for *P2Y10*;
17h5for, 5′-GGCAAAAAGTAAAGACCAGCA-3′, and 17h5rev, 5′-TGCCCGTTTAAAACAACTCA-3′, for *SH3BGRL*;
5c2.1for, 5′-GCTGTTTGGGCTAACAGGAT-3′, and 5c2.1rev, 5′-TTGCCGGTCAGTATCTGAGTT-3′, for *ITM2A*.
Amplifications consisted of a total of 35 cycles at 94°C for 45 s, 55°C for 45 s and 72°C for 1 min.

Selection and isolation of cDNA clones

A total of 2423 putative human X chromosome-specific cDNA clones, representing 1317 different human genes (UniGene clusters) were isolated, including all EST clusters from the UniGene database (24–26) known to map to the human

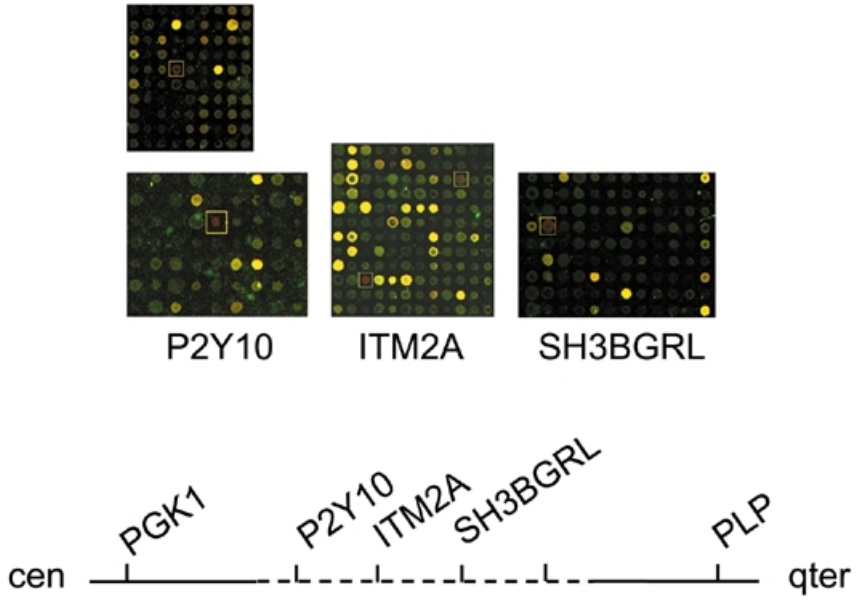

Figure 2. Genes located within a deletion of the proximal long arm of chromosome X in a male lymphoblastoid cell line show decreased expression levels in a hybridization experiment with the X chromosome-specific cDNA array. cDNA from the cell line carrying the deletion between the PGK1 and PLP locus (indicated by a dotted line) was labelled with Cy3 (green), cDNA from a male lymphoblastoid control cell line with Cy5 (red). The absence of cDNA corresponding to genes within the deletion which show detectable expression levels in lymphoblastoid cells such as P2Y10, ITM2A and SH3BGRL results in a high ratio of red to green signal (boxed spots on the chip, genes *P2Y10* and *ITM2A* are represented by two cDNA clones). The precise order of genes within the deletion is not known.

X chromosome, and others which were identified through alignment with known genomic sequences from the human X chromosome. These sequences were downloaded from http://www.ebi.ac.uk/~sterk/genome-MOT/ into a local database, which is updated at weekly intervals to incorporate new sequence information from the relevant EMBL database. Repeats, including simple repeats and low complexity regions, were detected by RepeatMasker (27), and the respective regions were masked. Masked sequences were used in a blastn (28) search against the human UniGene database (hs-unigene; ftp://ncbi.nlm.nih.gov/repository/unigene/). The output was filtered using MSPcrunch (29), which allows restriction of the output to blast hits with a minimum of 95% identity. In a second filtering step the blast results were further reduced by eliminating all sequences where the 95% sequence identity did not extend beyond an adjacent region of 150 nucleotides. These cut-off values represent a compromise based on experience. More stringent cut-offs result in a lower rate of false positives, but a higher risk of missing some X-specific clusters. The estimated rate of false positives with the selected cut-off values is 15%, mainly due to highly conserved domains of gene families and wrongly annotated genomic sequences. In this way we identified 1451 UniGene clusters assumed to be X chromosome specific. If existent, two representative IMAGE clones for each cluster were selected. IMAGE clones for 1317 of the 1451 UniGene clusters were available through the resource centre of the German Human Genome Project (RZPD Berlin; http://www.rzpd.de/). The estimated error rate associated with the IMAGE clone library due to interchanges and cross-contamination of clones is 14% (estimated by sequencing 192 putative X chromosome-specific IMAGE clones). After hybridization, clones of interest were checked carefully by re-sequencing, as outlined in Results.

Radiation hybrid data were available for some of the UniGene clusters (http://www.ncbi.nlm.nih.gov/genemap/). Physical map positions and cytogenetic localizations were taken from http://www.ncbi.nlm.nih.gov/genome/seq/ (30) and http://ixdb.molgen.mpg.de/ (31).

Generation of cDNA microarrays

Amino-silane-coated slides (PE Biosystems) were treated according to Guo *et al.* (32).

EST clone inserts were amplified using standard conditions with 5′-amino-modified primers M13for (5′-GTA AAA CGA CGG CCA G-3′) and M13rev (5′-CAG GAA ACA GCT ATG AC-3′) obtained from Metabion. All PCR products were evaluated by agarose gel electrophoresis and classified as 1 (single product), 2 (weak product), 3 (no product) or 4 (multiple products). Purified PCR products were re-suspended in 5 µl of 100 mM sodium carbonate/bicarbonate buffer (pH 9.0) and printed on slides using a robotic spotting device (Beecher Instruments). Spotting volume was ~5 nl at each site, resulting in spots of ~200 µm in diameter.

Microarray hybridization

Labelled cDNA targets were re-suspended in hybridization solution (50% formamide, 6× SSC, 0.5% sodium dodecyl sulphate (SDS), 5× Denhardt's solution), to which 1 µl of poly(dA) (1 µg/µl) and 1 µl of human Cot-1 DNA (1.25 µg/µl) were added. The resulting reaction solution was concentrated to 25 µl, denatured at 95°C for 3 min and snap-cooled on ice. Hybridization took place under a coverslip at 42°C for 16 h. The slides were washed at room temperature in 0.2× SSC and 0.01% SDS for 5 min, followed by two washing steps in 0.2× SSC for 5 min each.

Image acquisition and data analysis

Fluorescence intensities of Cy3 and Cy5 were measured separately at 532 and 635 nm with a laser scanner (418 Array Scanner; Affymetrix). The resulting 16-bit data files were imported into a custom-made image analysis software (Y. Chen, personal communication), which runs as extensions of IPLab Spectrum Software (Scanalytics).

The software identifies and extracts the average fluorescence intensity of the target site by a pixel selection method based on the Mann–Whitney test. A normalization of the Cy3- and Cy5-image is done on the basis of 48 housekeeping clones and upper and lower limits of differential gene expression ratios are calculated using a maximum-likelihood estimation algorithm. The discrimination of differential gene expression was performed at a significance level of 95%.

RT–PCR with Xi hybrids

RNA was prepared from five mouse or hamster somatic cell hybrids retaining an inactive human X chromosome, a female human control cell line as well as from a mouse and a hamster control cell line using Trizol reagent (Gibco BRL). cDNA amplified with the SMART RACE cDNA amplification kit (Clontech) was used as a template for PCR reactions. To exclude contaminating genomic DNA as a template of the obtained PCR products, RNA diluted appropriately was used as a control. Specific PCR primers were designed on the basis of EST or mRNA sequences in the database for 19 randomly selected clones of the 53 outlier clones identified by microarray experiments.

ACKNOWLEDGEMENTS

We thank Hannelore Madle for dedicated cell culture work, Ute Schulz for assistance with RT–PCR experiments, Bettina Lipkowitz and Ralf Schulz for picking and PCR amplification of EST clones and the RZPD for support. We thank Robert Wagner for bioinformatics support.

REFERENCES

1. Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.
2. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
3. DeRisi, J., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
4. Feng, X., Jiang, Y., Meltzter, P. and Yen, P.M. (2000) Thyroid hormone regulation of hepatic genes *in vivo* detected by complementary DNA microarray. *Mol. Endocrinol.*, **14**, 947–955.
5. DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.*, **14**, 457–460.
6. Khan, J., Saal, L.H., Bittner, M.L., Chen, Y., Trent, J.M. and Meltzer, P.S. (1999) Expression profiling in cancer using cDNA microarrays. *Electrophoresis*, **20**, 223–229.
7. Debouck, C. and Goodfellow, P.N. (1999) DNA microarrays in drug discovery and development. *Nature Genet.*, **21**, 48–50.
8. Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P. and Rane, S. (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl Acad. Sci. USA*, **96**, 12833–12838.
9. MacLachlan, T.K., Somasundaram, K., Sgagias, M., Shifman, Y., Muschel, R.J., Cowan, K.H. and El-Deiry, W.S. (2000) BRCA1 effects on the cell cycle and the DNA damage response are linked to altered gene expression. *J. Biol. Chem.*, **275**, 2777–2785.
10. Aitman, T.J., Glazier, A.M., Wallace, C.A., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Al-Majali, K.M., Trembling, P.M., Mann, C.J., Shoulders, C.C. *et al.* (1999) Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.*, **21**, 76–83.
11. Antonarakis, S.E. and McKusick, V.A. (2000) OMIM passes the 1000-disease-gene mark. *Nature Genet.*, **25**, 11.
12. Frischmeyer, P.A. and Dietz, H.C. (1999) Nonsense-mediated mRNA decay in health and disease. *Hum. Mol. Genet.*, **8**, 1893–1900.
13. Culbertson, M.R. (1999) RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends Genet.*, **15**, 74–80.
14. Lennon, G.G., Auffray, C., Polymeropoulos, M. and Soares, M.B. (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics*, **33**, 151–152.
15. Esposito, T., Gianfrancesco, F., Ciccodicola, A., Montanini, L., Mumm, S., D'Urso, M. and Forabosco, A. (1999) A novel pseudoautosomal human gene encodes a putative protein similar to Ac-like transposases. *Hum. Mol. Genet.*, **8**, 61–67.
16. Schiebel, K., Weiss, B., Wohrle, D. and Rappold, G. (1993) A human pseudo-autosomal gene, ADP/ATP translocase, escapes X-inactivation whereas a homologue on Xq is subject to X-inactivation. *Nature Genet.*, **3**, 82–87.
17. Carrel, L., Cottle, A.A., Goglin, K.C. and Willard, H.F. (1999) A first-generation X-inactivation profile of the human X chromosome. *Proc. Natl Acad. Sci. USA*, **96**, 14440–14444.
18. Yen, P.H., Ellison, J., Salido, E.C., Mohandas, T. and Shapiro, L. (1992) Isolation of a new gene from the distal short arm of the human X chromosome that escapes X-inactivation. *Hum. Mol. Genet.*, **1**, 47–52.
19. Lahn, B.T. and Page, D.C. (1997) Functional coherence of the human Y chromosome. *Science*, **278**, 675–680.
20. van der Maarel, S.M., Scholten, I.H.J.M., Maat-Kievit, J.A., Huber, I., de Kok, Y.J.M., de Wijs, I., van de Pol, T.J.R., van Bokhoven, H., den Dunnen, J.T., van Ommen, G.J.B. *et al.* (1995) Yeast artificial chromosome cloning of the Xq13.3–q21.31 region and fine mapping of a deletion associated with choroideremia and nonspecific mental retardation. *Eur. J. Hum. Genet.*, **3**, 207–218.
21. Carrel, L. and Willard, H.F. (1999) Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc. Natl Acad. Sci. USA*, **96**, 7364–7369.
22. Muller, C.R., Traupe, H. and Ropers, H.H. (1980) X-linked steroid sulfatase: evidence for different gene-dosage in males and females. *Hum. Genet.*, **54**, 197–199.
23. 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. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.*, **23**, 41–46.
24. Schuler, G.D. (1997) Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J. Mol. Med.*, **75**, 694–698.
25. Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodreguez-Tome, P., Aggarwal, A., Bajorek, E. *et al.* (1996) A gene map of the human genome. *Science*, **274**, 540–546.
26. Boguski, M.S. and Schuler, G.D. (1995) ESTablishing a human transcript map. *Nature Genet.*, **10**, 369–371.
27. Smit, A. RepeatMasker (http://ftp.genome.washington.edu/RM/RM_details.html).
28. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
29. Sonnhammer, E. and Durbin, R. (1994) A workbench for large-scale sequence homology analysis. *CABIOS*, **10**, 301–307.
30. Jang, W., Chen, H.-C., Sicotte, H. and Schuler, G.D. (1999) Making effective use of genomic sequence data. *Trends Genet.*, **15**, 284–286.
31. Leser, U., Roest Crollius, H., Lehrach, H. and Sudbrak, R (1999) IXDB, an X chromosome integrated database (update). *Nucleic Acids Res.*, **27**, 123–127.
32. Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. and Smith, L.M. (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.*, **22**, 5456–5465.