Transcriptome analysis of human autosomal trisomy

David R. FitzPatrick1,*, Jacqueline Ramsay1, Niolette I. McGill1, Mary Shade2, Andrew D. Carothers1 and Nicholas D. Hastie1

1MRC Human Genetics Unit, Edinburgh EH4 2XU, UK and 2Lothian Regional Cytogenetics Laboratory, WGH, Edinburgh EH4 2XU, UK

Received June 10, 2002; Revised and Accepted October 10, 2002

We present transcriptome analyses of primary cultures of human fetal cells from pregnancies affected with trisomy 21 (t21) and trisomy 13 (t13). Pooled mRNA samples from t21 and t13 cases were used for comparative hybridizations to cDNA arrays with pooled mRNA from normal cells. When the array cDNAs were grouped by chromosomal location the relevant trisomic chromosome could be clearly identified as showing the most significant misregulation. The average level of transcription on the trisomic chromosome was increased only \( \approx 1.1 \)-fold compared to normal cells on array analysis. Since the karyotype could be accurately predicted by the transcriptome this could provide a novel method of detecting aneusomy of unknown position. Subsequent analysis of individuals cases demonstrated that variation in transcriptional profiles between samples within each class made transcriptional karyotyping difficult without pooling or the use of arrays with a higher proportion of all human cDNAs. Interestingly, consistent differences in the relative expression levels between chromosomes were detected suggesting that genomic control mechanisms may act over larger distances than previously thought. Most (>95%) \( > \pm 2 \) SD misregulated genes did not map to the trisomic chromosome and significant misregulation was more common in t13 than t21. These data support a model of a subtle primary upregulation of genes on the trisomic chromosome resulting in a secondary, generalized and more extreme transcriptional misregulation. It seems likely that the degree of this misregulation determines the severity of the phenotype in most aneuploidy.

INTRODUCTION

Human autosomal trisomies are common causes of early pregnancy loss, neonatal death and multiple congenital anomalies (1,2). In postnatal life only trisomies of chromosomes 21, 18 or 13 are consistently detected with livebirth incidences of 1.23, 0.15 and 0.046 per 1000, respectively (3). Each produces a distinct clinical syndrome recognizable at or before birth. However, each trisomy shows considerable variability in the severity and pattern of associated malformations, which cannot be accurately predicted by the karyotype. The primary genetic defect in affected embryos is assumed to be a 1.5-fold increase in the dosage of genes on the trisomic chromosome (4), which alters signalling pathways to produce a specific and malign effect on development. Transcription factors on the trisomic chromosome will produce a secondary genome-wide transcriptional misregulation, which involves downregulation in addition to upregulation. These indirect effects are widely used in antenatal screening programs for trisomy 21 and trisomy 18, which detect abnormal levels of 

fetal proteins in maternal serum. Alphafetoprotein (AFP) is reduced in both t21 (5) and t18 (6) while human chorionic gonadotrophin (hCG) is elevated in t21 (7) and reduced in t18 (6). The genes encoding AFP and hCG map to 4q11–13 and 19q13, respectively.

The aim of the present study was to identify the pattern and extent of transcriptional misregulation in human autosomal trisomy and to determine if such information may improve our understanding of trisomy-associated pathology and inform the choice of proteins for maternal serum screening.

RESULTS

Microarray analysis of pooled samples

Table 1 summarizes the basic data on the hybridisation experiments and Table 2 lists the ‘top ten’ up- and down-regulated genes in both trisomies. The full results are available electronically (http://www.hgu.mrc.ac.uk/Research/Cellgen/).
Table 1. Summary of comparative hybridization results

<table>
<thead>
<tr>
<th></th>
<th>t21 versus Normal</th>
<th>Chr21 genes</th>
<th>t13 versus Normal</th>
<th>Chr13 genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>All genes</td>
<td>8020 (87.6%)</td>
<td>79 (92.9%)</td>
<td>8663 (94.9%)</td>
<td>113 (94.6%)</td>
</tr>
<tr>
<td>cDNA spots &gt; 2 SD Up-Regulated</td>
<td>187 (2.3%)</td>
<td>10</td>
<td>311 (3.6%)</td>
<td>5</td>
</tr>
<tr>
<td>cDNA spots &gt; 1 SD Up-Regulated</td>
<td>965 (12.0%)</td>
<td>27</td>
<td>1350 (15.6%)</td>
<td>27</td>
</tr>
<tr>
<td>cDNA spots &gt; 1 SD Down-Regulated</td>
<td>697 (8.7%)</td>
<td>2</td>
<td>1368 (15.8%)</td>
<td>9</td>
</tr>
<tr>
<td>cDNA spots &gt; 2 SD Down-Regulated</td>
<td>85 (1.1%)</td>
<td>0</td>
<td>210 (2.4%)</td>
<td>1</td>
</tr>
</tbody>
</table>

Two hybridizations were performed to the InCyte UniGene human cDNA microarray; t21 versus Normal and t13 versus Normal. 600 ng of each mRNA pool was labelled using either Cy3 (t13 and t21) or Cy5 (Norm). >250 and >150 differences are presented. >250 in the balanced Cy3 : Cy5 ratio is considered to represent true differential expression (InCyte Genomics). Hybridization signal was defined as >2.5-fold higher than background fluorescence and covered >40% of the spot area.

The mean and standard deviation (SD) of the ratios from the t21 versus Normal and t13 versus Normal comparative hybridizations was 1.01 and 0.19 and 0.26, respectively. The number of cDNAs showing >1 SD and >2 SD change in expression are summarized in Table 1. There were significantly more up- and down-regulated genes in t13 cells compared to t21 (P < 0.0001) both using 2 SD and 1 SD. If 1 SD change is considered then almost one third (31.4%) of all scorable genes are misregulated in t13 compared to just over a fifth (20.8%) for t21 (Table 1). Almost all the significantly upregulated genes differed between the two trisomies. Only six genes were >2 SD upregulated in both T21 and T13; PDGFRα (platelet-derived growth factor receptor-like), GFPT2 (glutamine-fructose-6-phosphate transaminase 2), MBD3 (methyl-CpG binding domain protein 3), TIMP3 (tissue inhibitor of metalloproteinase 3), NPPB (atriuretic peptide precursor B) and SA4 (serum amyloid A4). Significantly downregulated genes were more commonly shared and these fell into the following main groups: 1) Signal transduction; OXTR (oxytocin receptor), CTGF (connective tissue growth factor), TGFβ2 (transforming growth factor, beta 2), IGFFBP4, IGFFBP5 (insulin-like growth factor binding protein 4 and 5), EDN1 (endothelin 1), RGSS5 (regulator of G-protein signalling 5); 2) Extracellular matrix; FBN1 (fibrillin 1), COL4A1 (collagen, type IV, alpha 1), THBS1, THBS2 (thrombospondin 1 and 2); 3) Cell junctions and adhesion; CLDN1 (claudin 1), DSP (desmoplakin); 4) Apoptosis; CASP3 (caspase 3 and 5). Unknown function; BRAP (BRCA1 associated protein), DLC1 (deleted in liver cancer 1).

Transcriptosome analysis

Analysis of the fluorescent ratios for all cDNA on the array averaged by chromosome of origin (‘transcriptosome’) for the pooled mRNA experiments is shown in Figure 1. The relevant trisomic transcriptosome could be easily distinguished. In t21 chromosome 21 showed the largest increase and was the only one to differ significantly from the mean for all other chromosomes. In t13 there was much more variability in the transcriptosome ratios with several chromosomes showing significant misregulation, however, chromosome 13 showed the largest upregulation and was the only transcriptosome where the standard error did not overlap with any other chromosome. The mean levels of up-regulations on normalized ratios were 1.12 (95% CI: 1.08, 1.16) for t21 and 1.07 (95% CI: 1.03, 1.11) for t13.

Analysis of individual samples

Quantitative RT–PCR analysis of the individual cDNA samples was used both to confirm microarray ratios and to gain an indication of the variability within pooled samples. IGFBP3, IGFBP5 and RGSS5 were chosen as they were among the top ten up- (IGFBP3) or down- (IGFBP5 and RGSS5) regulated genes in one or both trisomies. GAPD was used as a control and the results are presented in Table 3. The RT–PCR results are generally consistent with the misregulation detected by microarray analysis of pooled samples. Two of the genes (RGSS5 and IGFBP5), however, show very significant variability within the pools and overlap with normal samples. IGFBP3 shows consistent upregulation in t21 cells and t13 (microarray ratio = 1.4) and no overlap with the normal samples.

The analysis was extended by hybridizing seven individual mRNA samples to separate arrays of 3280 different cDNA probes. Six of these mRNA samples had also been used in the pooling experiment. The seventh sample was amniocyte mRNA from a case of trisomy 18 (t18). The number of genes on the Human GEN1 array mapping to chromosomes 13, 18 and 21 were 64, 60 and 47, respectively. The transcriptosome results are summarized in Figure 2. In all trisomic individuals the trisomic chromosomes showed higher relative expression levels than the two controls. However these differences were not significant since the variability between individuals within each class was as great as the variability between classes. In spite of this, a remarkably stable pattern of relative expression levels between chromosomes was observed across all experiments with an average correlation coefficient of 0.88 (range 0.75–0.93) for all comparisons (Fig. 3). For example chromosomes 7 and 10 had consistently higher ratios than chromosome 9 in all hybridizations (Fig. 2).

DISCUSSION

Studies of gene expression based on RNA extracted from cell lines must be interpreted with care as spurious ratios may result from minor variations in the culture conditions. We tried to minimize these effects by using commercial media optimized for amniocyte culture, extracting RNA from quiescent cells and by pooling RNA within the diagnostic categories. Our subsequent analyses of individual samples demonstrated both the strength and weakness of this pooling strategy. The RT–PCR analysis of individual samples confirmed the average
misregulation found in the microarray analysis of pools and in one case (IGFBP3) has shown that the upregulation is consistent in and specific to trisomic samples. Following this finding a literature search revealed that IGFBP3 has been investigated as a maternal serum marker (8) and implicated in postnatal growth failure (9) in Down syndrome. The results from IGFBP5 and RGS5, however, show extremely variable transcriptional activity and overlap between trisomic and normal samples. These may therefore be simple cultural artefacts. In this regard it was, however, encouraging that a relatively small proportion of genes showed substantial misregulation and that these mostly differed between the t21 versus Normal and t13 versus Normal experiments.

cDNA microarrays have been extensively used for massive parallel analysis of gene expression. The human genome sequence has allowed us to determine the map position of almost all cDNAs analysed. Linkage of these two sets of information led to the most interesting and unexpected finding in this study. The trisomic chromosome could be easily distinguished on statistical examination of the expression ratios of all cDNAs when grouped by their chromosomal origin. This was surprising because the vast majority of the >2 SD misregulated genes did not map to the trisomic chromosome.
Although the average misregulation of genes on the trisomic chromosome could be clearly distinguished, the overall level of change was surprisingly low. This may simply reflect a limitation of the microarray assay in under-reporting the degree of misregulation. However, it could reflect important dosage compensation mechanisms acting in trisomic cells that lead to a deviation from the expected 1.5-fold change in transcription. Such mechanisms are likely to be mediated via cis-acting non-coding elements that may themselves be good candidates in the study of phenotypic variation in human trisomies. It is interesting that the trisomy 13 pattern was more chaotic than trisomy 21 and this difference was due to a larger number of both upregulated and downregulated genes in t13 compared to t21. This may be a non-specific effect of the culture of ‘sicker’ cells, however, there appears to be a roughly linear relationship between the degree of misregulation and the number of trisomic genes (~260 for Chr21; ~460 for Chr13). It is interesting that t13 is equivalent to a duplication of ~4% of the haploid genome which is approaching the 4.3% level of tolerance observed for segmental duplication of the human genome (11). Thus the lethality associated with larger aneuploid regions may be the direct result of exponentially increasing transcriptional chaos.

A consistent pattern of relative expression levels between chromosomes was observed in the course of analysing the, otherwise disappointing, microarray experiments of individual RNA samples (Figs. 2 and 3). It is not known whether this effect is specific to amniocytes or if patterns will change between tissues. Meta-analysis of SAGE data from various human tissues identified genomic clusters of highly expressed genes (12). It has also been shown that genes, which are transcriptionally active in muscle, cluster on C. elegans chromosomes. These data imply that there is regional variation in transcription levels across the genome which may be the result of physical clustering of genes of similar function and expression profile. Our data suggests that these regions may be much bigger than previously thought.

Genes mapping to the chromosome 21 Down syndrome critical region have been the subject of intense study (13–20). However, relatively few global analyses of transcription in human aneuploidy have been published and all of these relate to trisomy 21. Differential display has been used to compare transcription in t21 and control adult brains (21) but this approach is likely to have significant confounding effects given the common association of a neurodegenerative process (Alzheimer disease) in Down syndrome. The same group have used subtractive hybridization to detect Down syndrome-specific alterations in fetal brain expression (22). Analysis of brain tissue of the mouse model Ts65Dn has also been performed using the SAGE technique and demonstrated that 0.72% of the unique tags generated showed significant differences from control mice (23). This mouse contains three copies of a region of mouse chromosome 16 from App to Mx1, a region of conserved synteny with human chromosome 21. There appears to be very little overlap between mouse and human misregulation. This may not be surprising given that different tissues are under study in different species using different techniques. Amniocytes have no obvious trisomic phenotype and thus may have fewer transcriptional confounders associated with the secondary structural or degenerative pathology seen in the brain. However, our approach may reduce the chances of identifying important genes involved in the embryopathy.

The identification of gene networks that may be effectors of the trisomic phenotype is at an early stage. The most upregulated genes in t21 cells included several metalloproteinases (MMP10, MMP7, MMP1, TIMP3). While no direct link could be established between these specific genes and Down syndrome, other metalloproteinases have been implicated in the pathophysiology of periodontal disease in Down syndrome (24,25) and in Alzheimer disease (26,27) which is common in Down syndrome. Several growth factor-associated or regulated genes (IGFBP4, IGFBP5, CTGF, TGFβ2 and TGFBR1) were down-regulated in t21 which may have an influence on many embryological processes. In t13 the potential gene networks provided no obvious clues to the pathogenesis of the condition. It is interesting to note that Δ5-sterol reductase (seladin), a gene involved in cholesterol biosynthesis and mutated in desmosterolosis (MIM 602398) (28), is down regulated in t13. Mutations in another gene in this pathway have been implicated
in the aetiology of holoprosencephaly (29,30), which is a common malformation in t13.

We believe that transcriptome analysis holds great potential for unravelling the molecular basis of phenotypic variation and embryopathology in chromosomal disorders. Amniocytes are of fetal origin and easily available following routine diagnostic testing. It is not clear whether their expression profile will be able to accurately predict the phenotype and this will require further study where outcome data are available (e.g. microphthalmia in t13 or atrioventricular septal defects in t21). It is interesting that IGFBP3 is consistently upregulated in t21 amniocytes and has been previously implicated in postnatal growth failure in trisomy 21 (9) and as targets for maternal serum screening (8). Analysis of individual rather than pooled samples may then identify a subset of genes that can be used to improve diagnosis, prognosis and treatment.

**METHODS**

**Cell culture**

Primary cultures of amniocytes were collected from male fetuses between 16–18 weeks of pregnancy, following routine diagnostic cytogenetic analysis of singleton pregnancies. The following cell lines were used in the study; three cases of trisomy 21 (47,XY + 21); two cases of trisomy 13 (47,XY + 13 and 46,XY,−14,rob (13;14)mat); and four unrelated control cases with normal karyotypes (46,XY). A single case with trisomy 18 (47,XY + 18) was used for microarray analysis of individual cases. For ethical reasons the single case with trisomy 18 (47,XY + 18) was used for quantitative PCR analysis of individual samples. The results are presented as calculated numbers of double-stranded DNA molecules in the starting template. GAPD is used to normalize the results for IGFBP3, IGFBP5 and RGS5 and these results are presented as the number of double stranded DNA molecules per 1000 GAPD molecules.

**Microarray analysis**

Total RNA was extracted from each of the cell lines using Trisol (GibcoBRL) after they had been confluent for 5 days. mRNA was then isolated by two sequential rounds of hybridization, washing and elution using poly-dT magnetic beads (DYNAL). mRNA was quantitated and equivalent amounts were pooled in each of the following categories: Trisomy 21 (t21); Trisomy 13 (t13); Normal (Norm); and Trisomy 18 (t18). For the first experiments 600 ng of each pooled mRNA was labeled using either Cy3 (t13 & t21) or Cy5 (Norm) and then used for comparative hybridizations to a commercial human cDNA microarray Human UniGene 1 (InCyte Genomics) containing 9128 cDNAs representing 8466 unique genes/clusters. Two different comparative hybridizations were performed: t21 versus Normal and t13 versus Normal. The follow-up experiments on seven individual RNA sample (two controls, two t21, two t13 and one t18 sample) were not performed as comparative hybridizations due to the large number of combinations that would be required. In these experiments each RNA sample was post-labelled with Cy3 and Cy5 following amino allyl-dUTP incorporation during first-strand cDNA synthesis according to the manufacturers instructions (Amersham). The separate labelling reactions were then mixed and hybridized to seven Human GEN1 arrays (gifted by the MRC HGMP Resource Centre). The Human GEN1 array has 3280 different cDNA printed in duplicate on each slide. Each hybridization included synthetic RNA controls (InCyte Genomics) containing 9128 cDNAs representing 8466 unique genes/clusters. Two different comparative hybridizations were performed: t21 versus Normal and t13 versus Normal. The follow-up experiments on seven individual RNA sample were performed in triplicate or quadruplicate at different concentrations and different ratios (1:3, 3:1, 1:10, 10:1, 1:25, 25:1), which were mixed with the poly A RNA prior to labelling. The washed slides were scanned using the ScanArray 4000 apparatus (GSI Luminonics) and the fluorescence was quantified using GeneSpring software (Silicon Genetics).

**Quantitative PCR**

500 ng of each individual mRNA sample was heat denatured and cDNA produced using oligo-dT primers and PowerScript
Figure 2. Summary of the results from Human GEN1 microarray analysis using individual samples. Two t13, two t21, two control and one t18 were labelled with both Cy3 and Cy5 and hybridized to separate slides. For each experiment the recorded fluorescent levels were log transformed and averaged by chromosome. To facilitate comparison between mRNA samples the results were normalized by dividing the mean fluorescence of all spots in each experiment with the mean fluorescence of genes mapping to each chromosome to give a relative expression level. The same two control samples are shown in each graph. The position of relevant trisomic chromosome is shown with an arrow in each graph. (A) t13 versus Control; (B) t18 versus Control; (C) t21 versus Control. Although the relevant trisomic chromosomes are higher than the controls in each case this is not statistically significant due to the variation in pattern between individuals in each class i.e. t13, t21 and controls.
reverse transcriptase according to the manufacturer’s instructions (CLONTECH). 1/50 of the completed cDNA reaction was used as a template for RT–PCR using GAPD (32), IGFBP3 and IGFBP5 (33) primers that have been previously reported. The primers used to amplify RGS5 (RGS5-F 5′AGCCAAGACCCAGAAAACCT and RGS5-R 5′TTGCTTCTCAAGCCATCTT) were designed using Primer 3 software (http://www-genome.wi.mit.edu/genome_software/other/primer3.html). None of the primers amplified a product with a genomic DNA template. Real-time, quantitative fluorescent RT–PCR was performed using the LightCycler system (ROCHE) and SYBR green I fluorescent dye. A standard dilution series for each message was produced using the relevant purified DNA fragment. The PCR conditions consisted of an initial denaturation of 94°C, followed by 50 cycles of 94°C for 1 s, 55°C for 5 s and 72°C for 15 s. Fluorescence was measured at 85°C at the end of each cycle to avoid including primer dimer accumulation (34). Each quantitation was performed in triplicate on the individual mRNA samples and normalized to GAPD levels.

Bioinformatic analysis

For all microarray experiments the raw fluorescence data was imported into an Access database (MicroSoft). A hybridization signal was defined as a fluorescent measurement >2.5-fold higher than background fluorescence and covered >40% of the spot area. For the two comparative hybridizations >±2 SD from the mean of the balanced Cy3:Cy5 ratio was taken to represent significant differential expression. The chromosomal origin of each gene was determined by linking the accession number of the microarray cDNAs to a download of LocusLink (ftp://ncbi.nlm.nih.gov/refseq/LocusLink/). Statistical analysis by ANOVA of log (fluorescence ratio) versus chromosomal origin was performed on gene expression data with outliers removed (all those more than 3.5 SD from each chromosome mean). Data from cDNAs of unknown chromosomal origin were not included. The statistician performing this analysis (ADC) was blinded to the nature of the trisomy. Gene networks were identified using the PubGene web interface (http://www.pubgene.org/).

ACKNOWLEDGEMENTS

We thanks Professor van Heyningen for her helpful comments on the manuscript. This work was funded by the Medical Research Council and the Chief Scientist Office, Scotland grant number K/MRS/50/C2701. We gratefully acknowledge the gift of the Human GEN1 microarray slides from the MRC HGMP Resource Centre.

REFERENCES


