ARTICLE

Natural Gene-Expression Variation in Down Syndrome Modulates the Outcome of Gene-Dosage Imbalance

Paola Prandini, Samuel Deutsch, Robert Lyle, Maryline Gagnebin, Celine Delucinge Vivier, Mauro Delorenzi, Corinne Gehrig, Patrick Descombes, Stephanie Sherman, Franca Dagna Bricarelli, Chiara Baldo, Antonio Novelli, Bruno Dallapiccola, and Stylianos E. Antonarakis

Down syndrome (DS) is characterized by extensive phenotypic variability, with most traits occurring in only a fraction of affected individuals. Substantial gene-expression variation is present among unaffected individuals, and this variation has a strong genetic component. Since DS is caused by genomic-dosage imbalance, we hypothesize that gene-expression variation of human chromosome 21 (HSA21) genes in individuals with DS has an impact on the phenotypic variability among affected individuals. We studied gene-expression variation in 14 lymphoblastoid and 17 fibroblast cell lines from individuals with DS and an equal number of controls. Gene expression was assayed using quantitative real-time polymerase chain reaction on 100 and 106 HSA21 genes and 23 and 26 non-HSA21 genes in lymphoblastoid and fibroblast cell lines, respectively. Surprisingly, only 39% and 62% of HSA21 genes in lymphoblastoid and fibroblast cells, respectively, showed a statistically significant difference between DS and normal samples, although the average up-regulation of HSA21 genes was close to the expected 1.5-fold in both cell types. Gene-expression variation in DS and normal samples was evaluated using the Kolmogorov-Smirnov test. According to the degree of overlap in expression levels, we classified all genes into 3 groups: (A) nonoverlapping, (B) partially overlapping, and (C) extensively overlapping expression distributions between normal and DS samples. We hypothesize that, in each cell type, group A genes are the most dosage sensitive and are most likely involved in the constant DS traits, group B genes might be involved in variable DS traits, and group C genes are not dosage sensitive and are least likely to participate in DS pathological phenotypes. This study provides the first extensive data set on HSA21 gene-expression variation in DS and underscores its role in modulating the outcome of gene-dosage imbalance.

The clinical presentation of Down syndrome (DS) or trisomy 21 (MIM 190685) is complex and highly variable.[1] Cognitive impairment, muscle hypotonia at birth, and dysmorphic features occur to some extent in all affected individuals. In contrast, the majority of the other associated traits are present in only a fraction of individuals with trisomy 21. In addition, the severity of many phenotypic traits varies greatly.[2]

Theoretically, the supernumerary copy of human chromosome 21 (HSA21) is expected to result in a 50% increase in the level of transcripts of all genes mapping to HSA21. However, it has been recently observed that there is not always a direct correlation between genomic imbalance (deletion or duplication) and transcript level of genes within the aneuploid segment, suggesting that complex molecular mechanisms regulate RNA transcript levels.[3–5]

An additional level of complexity comes from the recent observations of extensive gene-expression variation among unaffected individuals and that a significant fraction of this variation is controlled by genetic variation, either in *cis* or *trans* to the individual gene.[6–8] In a previous study, we showed, by quantitative PCR, that there is also extensive expression variation for HSA21 genes, with some genes varying up to 40-fold among individuals.[9,10] These findings may have direct implications for the phenotypic variability of DS and underlie the need to re-evaluate our models of dosage imbalance and how they relate to human disorders.

To date, very little is known about the genes and pathways involved in DS pathology, although recently the involvement of the nuclear factor of activated T cells (NFAT)[11] pathway and Sonic hedgehog[12] pathway has been postulated. Several previous genomewide expression studies analyzed the pattern and extent of gene expression dysregulation in human trisomy 21 and its mouse models, to identify candidate genes responsible for DS phenotypes.[13–19] All these studies detected the expected up-regulation of a fraction of HSA21 transcripts. However, small

From the Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospitals of Geneva (P.P.; S.D.; R.L.; M.G.; C.G.; S.E.A.), and National Center for Competence in Research (NCCR) Frontiers in Genetics, Genomics Platform, University of Geneva (C.D.V.; P.D.), Geneva; Department of Medical Genetics, Ullevål University Hospital, Oslo, Norway (R.L.); Swiss Institute of Bioinformatics, Lausanne, Switzerland (M.D.); Swiss Institute of Experimental Cancer Research, Epalinges, Switzerland (M.D.); Emory University School of Medicine, Atlanta, Georgia (S.S.); Genetics Laboratory, Galliera Hospital, Genoa, Italy (F.D.B.; C.B.); and Istituti di Ricovero e Cura a Carattere Scientifico (IRCCS)-Casa Sollievo della Sofferenza (CSS), San Giovanni Rotondo and CSS-Mendel Institute (A.N.; B.D.), and Department of Experimental Medicine and Pathology, University of Rome "La Sapienza" (B.D.), Rome

Received March 26, 2007; accepted for publication April 27, 2007; electronically published June 20, 2007.

Address for correspondence and reprints: Dr. Stylianos E. Antonarakis, Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospitals of Geneva, 1 rue Michel-Servet, 1211 Geneva, Switzerland. E-mail: Stylianos.Antonarakis@medecine.unige.ch

Am. J. Hum. Genet. 2007;81:252–263. 0002-9297/2007/8102-0007$15.00
DOI: 10.1086/519248

Table 1. Characteristics of Cell Lines

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Table 2. TaqMan Assays

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

sample sizes and limitations inherent to microarray technology have precluded detailed analysis of interindividual variation.

We hypothesize that natural gene-expression variation of HSA21 genes in individuals with trisomy 21 contributes to the phenotypic variability in DS. We expect that, for a fraction of genes, there is a degree of overlap of expression levels between individuals with DS and unaffected individuals, whereas, for other genes, the distributions of expression levels are distinct. Those genes with expression overlap are thus candidates for the variable phenotypes of DS, likely the result of a threshold effect.[20] In contrast, genes with distinct distributions are candidates for the constant features of DS.

To classify transcripts according to the overlap of expression levels between DS and normal samples and to determine the impact of natural gene-expression variation in the context of aneuploidies, we studied gene-expression variation by quantitative real-time PCR. We analyzed 14 lymphoblastoid cell lines (LCLs) and 17 fibroblast cell lines from individuals with DS and an equal number of matched controls and assayed HSA21 annotated genes that are expressed in LCLs and fibroblasts.

Material and Methods

Lymphoblastoid and Fibroblast Cell Culture

Epstein Barr virus (EBV)–transformed LCLs and fibroblast cell lines were obtained from Coriell cell repositories (Coriell Institute for Medical Research) (14 LCLs and 28 fibroblasts), Galliera Genetic Bank of Genoa (4 LCLs and 5 fibroblasts), Emory University School of Medicine in Atlanta (8 LCLs), and CSS-Mendel Institute in Rome (3 LCLs). All cell lines were assessed for full trisomy by karyotyping. LCLs and fibroblasts were collected from different individuals, and the control samples for both cell lines were matched for sex and ethnicity. Age was matched for fibroblasts, but, for LCLs, individuals with DS were, on average, younger than control individuals (table 1). Informed consent was obtained for all human samples, and the study was approved by the ethics committee of the Geneva University Hospital.

LCLs were grown in RPMI 1640 with Glutamax I medium (Invitrogen) supplemented with 10% fetal calf serum and 1% mix of penicillin, fungizone, and streptomycin (Invitrogen). Fibroblasts were cultured in Dulbecco's modified Eagle medium with Glutamax I plus Na pyruvate (Invitrogen) supplemented with the same antibiotic mix.

The cell lines were treated with a standardized procedure, to minimize environmental variation. Cell lines were harvested at a density of 0.6–1 × 10^6 cells/ml or 6.5–10 × 10^6 cells/dish and at least 80% viability. LCLs and fibroblasts (after trypsinization) were spun for 5 min at 1,000 *g,* and the resulting pellets were rinsed with PBS and were lysed with 1 ml lysis solution containing β-mercaptoethanol (RNeasy kit [Qiagen]). Cell pellets were stored at −80°C.

Total RNA was extracted using the RNeasy kit (Qiagen), including the DNAse step in accordance with the manufacturer's protocol. RNA samples were then quantified with NanoDrop (NanoDrop Technologies) and were analyzed for quality control on a 2100 BioAnalyzer by use of the RNA 6000 Nano LabChip (Agilent).

cDNAs were synthesized from total RNA by use of SuperScriptII reverse transcriptase (Invitrogen) and a poly d(T) primer. For each cell line, 10.5 μg of total RNA was used to perform three reverse-transcriptase reactions, and the resulting cDNA was diluted 1:14 before PCR.

Gene Selection

On the basis of a combination of the Hattori et al.[21] and Ensembl annotations, we initially considered 258 genes on HSA21. We excluded from our analysis (i) pseudogenes, (ii) gene predictions supported by spliced ESTs but not complete mRNAs, (iii) genes supported only by *ab initio* predictions, (iv) single-exon genes, and (v) two genes for which it was not possible to design assays by use of our default parameters. We also excluded 27 genes from the *KRTRAP* cluster that are almost exclusively expressed in hair root and for which it was impossible to design specific primer-probe sets. This resulted in a total of 178 genes for which we designed a TaqMan assay.

Primers and probes were selected using Primer Express version 2.0 (Applied Biosystems) with default parameters. Assay efficiencies were calculated using five fourfold serial dilutions of a pool of human brain, liver, and testis cDNAs, as described elsewhere.[3,9] When transcripts were not expressed in brain, liver, and testis, efficiencies were tested in a pool of LCLs and fibroblasts. Of the 178 genes, 43 did not pass our efficiency criteria threshold (0.95–1.05) and were excluded from further analysis. The remaining 135 genes were tested in a pool of five LCL cDNAs and five fibroblast cDNAs, to determine their expression in these cell types. A total of 117 and 114 genes were expressed in LCLs and fibroblasts, respectively, and were retained for subsequent analysis.

In addition, we designed assays for 30 non-HSA21 genes. Among these, five genes for LCLs and four genes for fibroblasts were selected for normalization, and the remaining genes were used as additional controls. Selection of normalization genes was performed with GeNorm software.[22] A list of all the assays used is provided in table 2.

Quantitative Real-Time PCR

Real-time quantitative PCR was performed as described elsewhere,[3,9] with minor modifications. Each gene was amplified in six replicates per individual. The assays for each individual were performed in 3 × 384-well plates. To assess possible technical errors generated by interplate variations in amplification, we placed three control genes (*AGPAT1, EEF1A1,* and *B2M*) in each plate. We performed different normalization procedures, using either genes in the first plate or genes present in each plate. Different normalization methods were highly correlated (Spearman's rank correlation $\rho = 0.8955$; $P = 2.2 \times 10^{-16}$), suggesting that interplate variation gives a negligible contribution to the expression variability observed.

Moreover, on the basis of two previous studies, we have estimated that the variance of gene-expression levels due to culture conditions is extremely low. For example, Merla et al.[4] established six independent LCLs for the same individual and compared expression levels for 25 HSA21 genes. For all genes, correlations were between 0.8 and 0.92. In addition, LCLs from the HapMap collection were cultured, and RNA was extracted from two different laboratories; the gene-expression levels showed very high overall correlation, suggesting that the uncontrolled culturing parameters did not significantly contribute to the measured gene-expression variation.[10,23] In total, 53,346 quantitative real-time PCR reactions were performed.

Data Analysis

Raw cycle threshold (C_T) values were obtained using SDS 2.1 software (Applied Biosystems). Baseline values were automatically determined, and threshold values were manually adjusted for each gene. Values with a deviation of 0.25 C_T with respect to the median (which corresponds to the 99% CI) were considered outliers and were excluded.

Transcripts that amplified with a C_T value >37.9 and in fewer than seven individuals per sample set were not included in the analysis. Each gene was rescaled using the mean expression value of control individuals, to give a relative normalized value. Data handling and normalizations were performed using Excel (Microsoft Corporation) and R (The R Project for Statistical Computing) software.

To assess the differences in gene-expression values between individuals with DS and unaffected individuals in LCLs and fibroblast cell lines, we performed the Kruskal-Wallis (KW) test. *P* values were corrected for multiple testing by use of the false-discovery rate (FDR) method of Benjamini and Hochberg.[24] We applied a conservative significance threshold of 1% FDR.

As an alternative method to compare the distributions of the expression values between individuals with DS and unaffected individuals and to compare their degree of overlap, we used the Kolmogorov-Smirnov (KS) test. The KW and KS tests were implemented using Minitab and R software, respectively.

To determine whether there is contiguous gene region on HSA21 that significantly departs from the expected 1.5-fold gene-expression dysregulation in DS, we performed a sliding-window analysis. We used the averaged ratio of expression in DS samples versus in euploid samples (DS:euploid) or the log *P* value (from the KW test) in a fixed window of four genes as the test statistic and assessed significance by a permutation test randomizing the order of the HSA21 genes.

Pathway analysis was performed using IPA software (Ingenuity System). In brief, gene lists were created using the *D* value classification from the KS test, to identify potential enrichment for certain functional categories.

Results

HSA21 Gene Expression in LCLs and Fibroblasts

To analyze HSA21 gene-expression differences between individuals with DS and euploid individuals, we designed and tested 178 TaqMan assays (see the "Material and Methods" section for the criteria used for selection). A total of 135 assays met our efficiency (*E*) criteria of $0.95 < E < 1.05$ in a pool of human brain, liver, and testis RNA; these assays constitute our HSA21 TaqMan set (fig. 1*A*).

To select appropriate gene sets to study in each cell type, we tested the HSA21 TaqMan set in RNAs from a pool of five LCLs and five fibroblast cell lines. We used C_T of 37.9 as the threshold for declaring a gene expressed, and, on the basis of these criteria, we selected 117 and 114 genes for LCLs and fibroblasts, respectively (fig. 1*A*). In addition, 30 non-HSA21 genes were incorporated into the analysis (table 2).

HSA21 expression levels were measured in 29 LCLs (14 normal and 15 trisomy 21) and 33 fibroblasts (16 normal and 17 trisomy 21). Genes that were not detected in at least seven individuals per group were eliminated, leaving a total of 128 LCL-expressed genes (100 on HSA21) and 136 fibroblast-expressed genes (106 on HSA21) for the final analyses (table 2).

HSA21 Gene Overexpression in Trisomy 21

Comparison of expression levels between individuals with trisomy 21 and euploid individuals revealed a general overexpression of HSA21 transcripts in affected individuals compared with unaffected individuals, both in LCLs and in fibroblasts ($P = 3.783 \times 10^{-7}$ for LCLs and $P = 2.2 \times 10^{-16}$ for fibroblasts, by KW test). We observed that 39% (39/100) of HSA21 transcripts in LCLs and 62% (66/106) in fibroblasts showed a significant expression difference between the DS and euploid samples. The significance threshold was set at .01 after correction for multiple testing by use of the Benjamini-Hochberg FDR.[24] The average up-regulation ratio (DS:euploid) was 1.44 in LCLs and 1.67 in fibroblasts; these values are similar to the expected up-regulation of gene expression of 1.5-fold in trisomy 21 (fig.1*B* and table 3). None of the genes tested in both cell types showed significant down-regulation in DS samples compared with in normal samples. These results are consistent with previously published data showing an overall up-regulation of HSA21 genes in individuals with DS and in trisomy 21 mouse models.[5,13,15,17,18,25,26]

Examples of some of the statistically significantly overexpressed genes in LCLs (with their corresponding fold overexpression and corrected *P* values for multiple testing) are *GABPA* (1.48-fold; $P = .0006$), *PFKL* (1.57-fold; $P = .0006$), *ITGB2* (1.65-fold; $P = .0006$), and *TMPRSS3* (8.1-fold; $P = .0006$). For fibroblasts, examples include *U2AF1* (1.76-fold; $P = 4.81 \times 10^{-5}$), *USP16* (1.77-fold; $P = 4.81 \times 10^{-5}$), *DONSON* (2.24-fold; $P = 4.81 \times 10^{-5}$), and *GART* (1.54-fold; $P = 4.81 \times 10^{-5}$).

As an additional control, we assessed whether transcript

Table 3. Average Up-Regulation and Associated *P* Values of Genes in Trisomy Samples

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

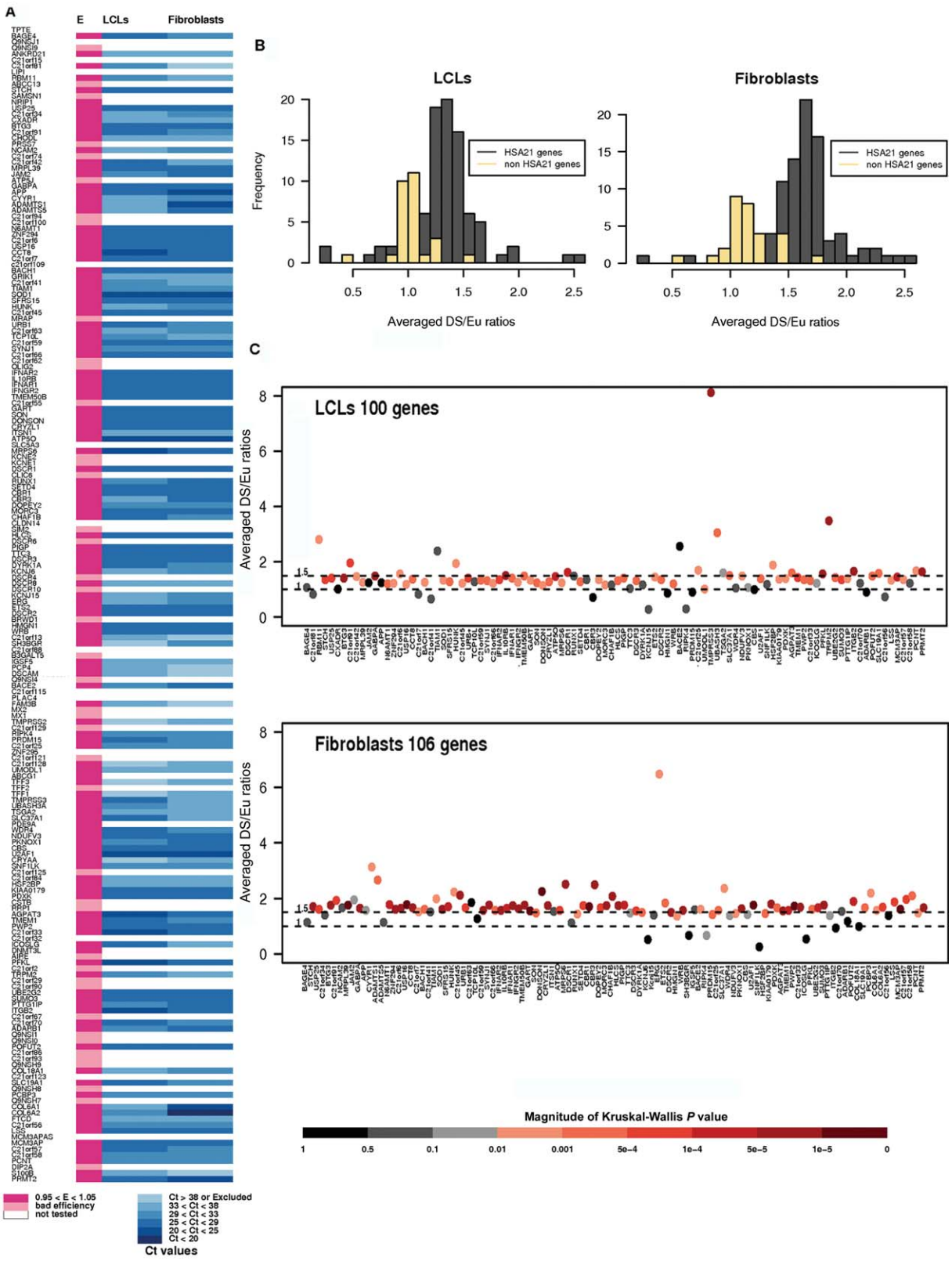


Figure 1. *A,* Representation of HSA21 genes analyzed. In the first column, dark pink indicates assays with efficiency (*E*) 0.95–1.05, light pink indicates assays with efficiency <0.95 (not used in this study), and white indicates assays not designed (see selection criteria in the "Material and Methods" section). The different shades of blue represent the average level of expression for each gene in LCLs and fibroblasts; genes with no expression data available are in white. *B,* Histograms of average ratio of expression in DS versus in euploid samples (DS/Eu) for HSA21 genes (*gray*) and non-HSA21 genes (*yellow*) in LCLs (*left panel*) and fibroblasts (*right panel*). *C,* Average DS/Eu expression ratios of 100 HSA21 genes in LCLs (*upper panel*) and of 106 HSA21 genes in fibroblasts (*lower panel*). Each dot corresponds to the average of normalized expression values for HSA21 genes according to their order along the chromosome. The range of *P* values is shown; red indicates lower *P* values, and black indicates higher *P* values.

up-regulation specifically involved HSA21 genes. We measured the expression level of 23 genes expressed in LCLs and 26 genes expressed in fibroblasts localized outside HSA21. The gene-expression ratio (DS:euploid) for these non-HSA21 transcripts was 1.02 in LCLs and 1.14 in fibroblasts, which was significantly different from that for HSA21 genes in the two cell types ($P = 1.305 \times 10^{-6}$ for LCLs and 2.865×10^{-9} for fibroblasts) (fig. 1*B*). Although non-HSA21 genes did not show a general up-regulation, *WBSCR18* (HSA7) was significantly upregulated in DS fibroblast samples compared with in normal fibroblast samples (corrected $P = .00366$, by KW test) (table 3).

To test whether there were any regional patterns of expression dysregulation, we performed a sliding-window permutation test of the analyzed HSA21 genes. Permutation analysis did not reveal statistically significant evidence of a cluster of contiguous genes showing locally higher overexpression or underexpression among the studied set of HSA21 genes. This suggests that the transcriptional dysregulation of HSA21 genes in trisomy 21 is uniformly spread along the chromosome (fig. 1*C*) and is not preferentially located in certain areas, which is consistent with previous reports.[16]

One interesting outlier is *TMPRSS3,* for which we found an average DS:euploid ratio of 8.1 (fig. 1*C*), which is much higher than would be expected from the genomic imbalance alone. This suggests that additional mechanisms, such as positive feedback loops, might be operating for this gene. *TMPRSS3* was not expressed in fibroblasts, precluding a comparison between the two cell types.

Gene-Expression Variation in DS and Normal Samples

Since we found that a substantial fraction of HSA21 genes are not significantly overexpressed in the trisomy 21 population in the two cell types analyzed, we surmised that interindividual differences in the levels of expression could partly explain this observation. To evaluate the extent of gene-expression variation among individuals, we measured the coefficient of variation (CV) of the normalized expression values for each gene in both cell types.

We observed a wide distribution of CVs, in the range 0.14–1.29 in LCLs and 0.14–1.49 in fibroblasts, with medians of 0.31 and 0.34, respectively. Within each cell type, the CVs were not significantly different between unaffected individuals and individuals with DS ($P = .052$ for LCLs and $P = .084$ for fibroblasts). For some genes (e.g., *PWP2*), the CV is low in both cell types (0.17 and 0.19 in LCLs and fibroblasts, respectively), suggesting that the expression of the gene is tightly regulated, whereas, for other genes (e.g., *KCNJ15*), the CV is approximately seven times higher (1.03 and 1.25), suggesting a less controlled transcriptional regulation for the gene in the two cell types analyzed (fig. 2*A*).

The distributions of the CVs were not significantly different between the two cell types ($P = .146$, by KW test), and, interestingly, we observed a significant correlation in the levels of gene-expression variation between fibroblasts and LCLs (Spearman's $\rho = 0.58$; $P = 1.717 \times 10^{-9}$), suggesting that a majority of the genes show a similar pattern of expression variation across cell types (fig. 2*B*), which is consistent with previous observations.[3]

To determine whether the degree of interindividual gene-expression variation could partly explain why certain HSA21 genes are not significantly overexpressed in the DS samples, we performed a regression analysis for all genes, taking the CV as the "predictor" and the $-\log P$ value of the expression difference (euploid vs. DS) as the "response." The results of the regression analysis were highly significant for both cell types ($P = 1.0 \times 10^{-12}$ for LCLs and $P = 2.3 \times 10^{-16}$ for fibroblasts), showing that, as CV increases, the P values tend to increase (with R^2 values of 0.42 and 0.58 in LCLs and fibroblasts, respectively) (fig. 2*C*). Hence, the degree of gene-expression variation contributes, to a large extent, to the determination of whether HSA21 genes are significantly overexpressed in the DS samples.

Estimating the Degree of Overlap of Gene Expression between DS and Normal Samples

Given the considerable gene-expression variation observed for many genes (fig. 3), we decided to approach the problem of overexpression, not only in terms of the average expression dysregulation in DS samples compared with in normal samples, but also in terms of the amount of overlap in the distribution of expression values between the two groups. For this purpose, we used the KS test, which measures the distance between empirical cumulative distributions of expression levels in DS samples compared with normal samples for each gene.

A D value and an associated P value are assigned to each gene according to their degree of overlap between the two distributions of expression values (in DS and normal samples). The D values range from 0 (i.e., no difference between the two distributions) to 1 (i.e., nonoverlapping distributions).

Using this approach, we classified all HSA21 genes into three groups. Group A had D values of 0.8–1, corresponding to genes with little or no overlap in the distribution between DS and normal samples (fig. 4). This group of genes—including, for example, *IFNAR2, GABPA,* and *SUMO3*—are highly dosage sensitive and are the most easily identifiable as upregulated by use of microarray analysis.[19] Group B had D values of 0.5–0.8, corresponding to genes with partially overlapping distributions (for example, *USP25, DSCR2, PIGP,* and *HUNK*); for these genes, only a fraction of trisomy 21 samples show expression values above the distribution of the normal (euploid) samples. Group C had D values <0.5, corresponding to genes that are effectively dosage insensitive, partly as a result of the substantial gene-expression variation in the population. For these genes, we observed an extensive expression over-

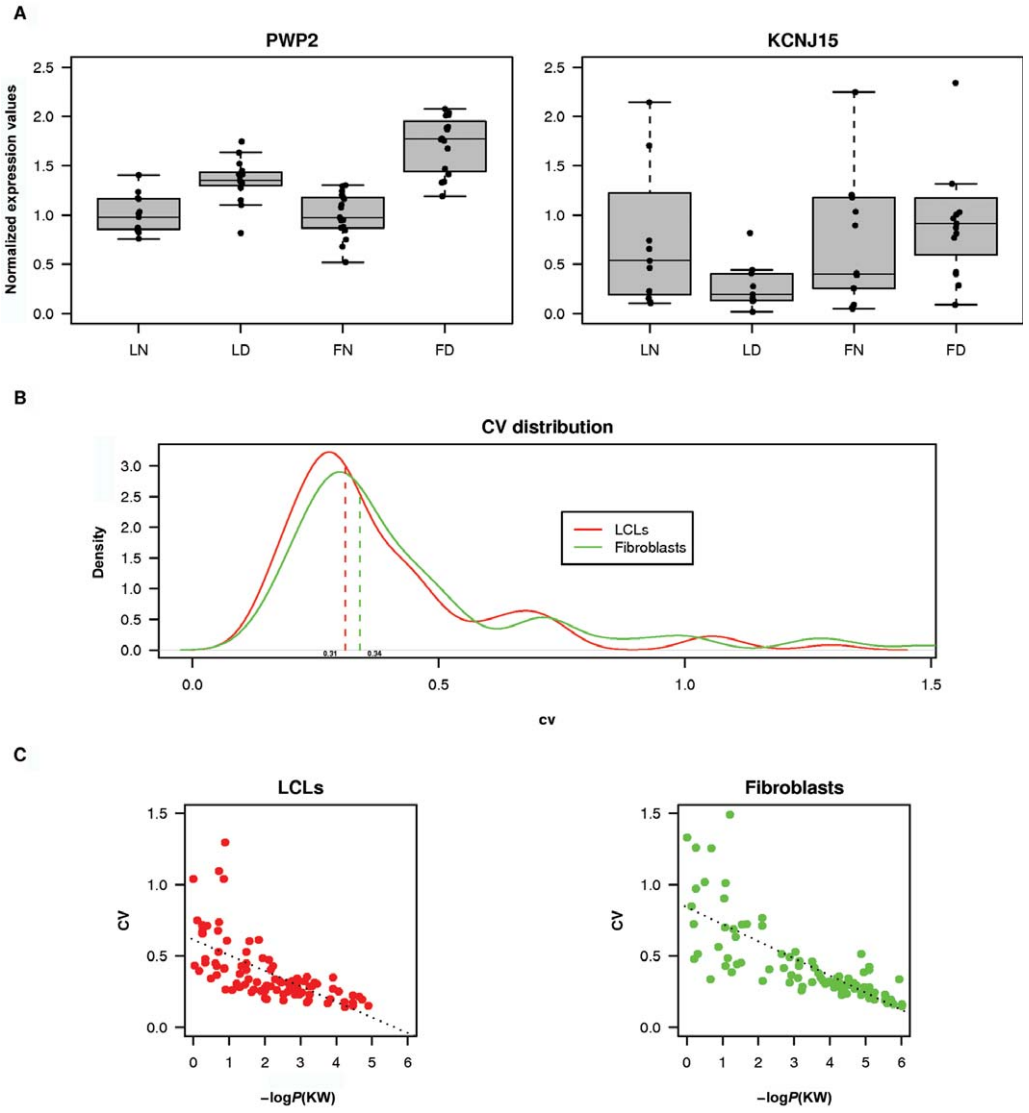


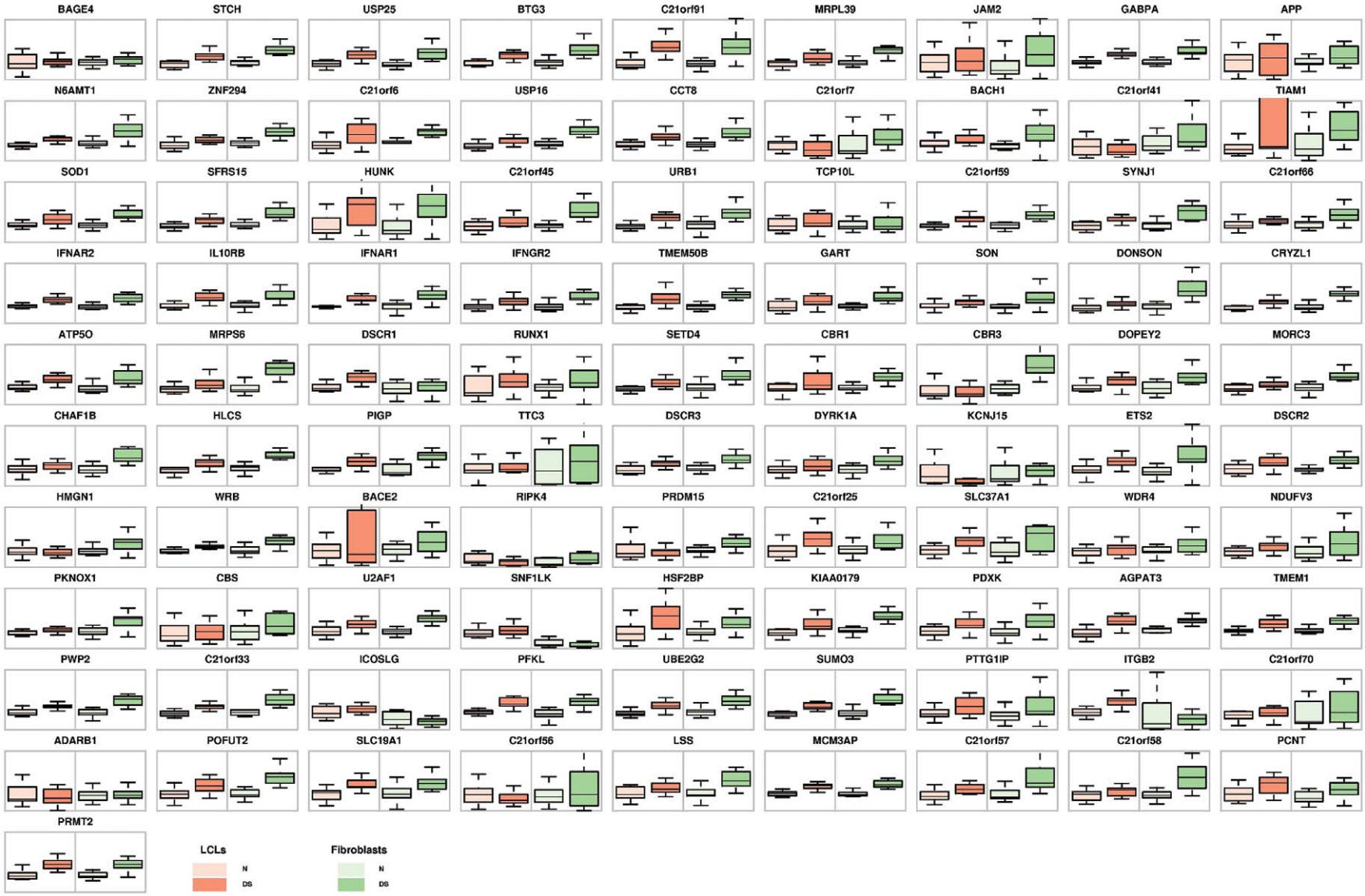
Figure 2. *A,* Examples of box plots of gene-expression values for two genes, *PWP2* and *KCNJ15*. The *Y*-axis is normalized expression values; the *X*-axis is the 62 samples, grouped by cell type (LN = normal LCLs; LD = DS LCLs; FN = normal fibroblasts; FD = DS fibroblasts). The left panel shows an example of a gene with low expression variance (<0.20), and the right panel shows a gene with high expression variance (>1). *B,* Distribution of CV of gene expression in LCLs (*red*) (median 0.31) and fibroblasts (*green*) (median 0.34). *C,* Regression curve of CVs versus the −log*P* of the KW test in LCLs (*red*) and fibroblasts (*green*), showing an inverse correlation between the CV and the −log*P* of the KW test.

lap between the two groups. Examples in this group include *JAM2, PCNT, BAGE4,* and *CBS*.

Sixty (60%) of the 100 HSA21 genes studied in LCLs and 73 (69%) of the 106 HSA21 genes studied in fibroblasts were in groups A and B (*D* values >0.5; *P* values <.05), showing that the majority of HSA21 genes are sensitive to the dosage imbalance, even in the presence of gene-expression variation (fig. 4). Interestingly, the number of group A genes is significantly higher in fibroblasts than in LCLs (fig. 4*A*). This may be because of the EBV transformation process, which may cause a transcriptional program in LCLs altered from that in fibroblasts that are primary cell lines.[27]

To systematically compare the degree of overlap between the two cell types studied, we focused on 91 genes that are commonly expressed in LCLs and fibroblasts. Information about the extent of overlap for genes that were specifically expressed either in LCLs or in fibroblasts is summarized in figure 5. Overall, there was a high correlation of *D* values between the two cell types ($\rho = 0.518$; $P = 1.452 \times 10^{-7}$). In addition, 76% of genes are in the same or similar (A or B) gene group in both cell types (fig. 4*B*), showing that a substantial proportion of genes display a similar degree of dosage sensitivity in both cell types.

We performed a pathway analysis to determine whether the different groups of genes (A, B, and C) were enriched


BAGE4
STCH
USP25
BTG3
C21orf91
MRPL39
JAM2
GABPA
APP
N6AMT1
ZNF294
C21orf6
USP16
CCT8
C21orf7
BACH1
C21orf41
TIAM1
SOD1
SFRS15
HUNK
C21orf45
URB1
TCP10L
C21orf59
SYNJ1
C21orf66
IFNAR2
IL10RB
IFNAR1
IFNGR2
TMEM50B
GART
SON
DONSON
CRYZL1
ATP5O
MRPS6
DSCR1
RUNX1
SETD4
CBR1
CBR3
DOPEY2
MORC3
CHAF1B
HLCS
PIGP
TTC3
DSCR3
DYRK1A
KCNJ15
ETS2
DSCR2
HMGN1
WRB
BACE2
RIPK4
PRDM15
C21orf25
SLC37A1
WDR4
NDUFV3
PKNOX1
CBS
U2AF1
SNF1LK
HSF2BP
KIAA0179
PDXK
AGPAT3
TMEM1
PWP2
C21orf33
ICOSLG
PFKL
UBE2G2
SUMO3
PTTG1IP
ITGB2
C21orf70
ADARB1
POFUT2
SLC19A1
C21orf56
LSS
MCM3AP
C21orf57
C21orf58
PCNT
PRMT2
LCLs
N
DS
Fibroblasts
N
DS

for involvement in particular pathways or biological processes. Interestingly, genes in group A and/or B, both in LCLs and fibroblasts, showed an enrichment in the *interferon-IL10RB* signaling pathway (P range 1.79×10^{-5}–1.52×10^{-2}). This network includes *IFNAR1, IFNAR2, IL10RB,* and *IFNGR2,* genes known to cluster in ~250 kb on 21q22.11, known to be conserved down to chicken as a syntenic block.[28]

Discussion

DS is considered a disorder of gene-expression imbalance in which allelic differences are likely to be important determinants of the phenotypic variability. Several recent studies, mainly using LCLs and array-based transcriptome analyses, concluded that there is substantial gene-expression variation in unaffected individuals.[7,29,30] In addition, a considerable fraction of this normal gene-expression variation is genetically determined.[8–10] Thus, a likely molecular mechanism for the variability of phenotypic manifestations of trisomy 21 is a threshold effect of expression of HSA21 genes that show variable levels of expression in the population.[20]

In this study, we have looked at HSA21 gene overexpression in trisomy 21 in the context of natural gene-expression variation. This enabled us to determine genes for which the additional copy results in significant overexpression in individuals with DS (dosage-sensitive genes) and genes for which the additional copy does not result in overexpression outside of the range in unaffected individuals (dosage-insensitive genes).

We measured HSA21 gene-expression levels, using two different cell types from samples originating from different DS-affected individuals (15 LCLs and 17 fibroblasts) and matched controls (14 LCLs and 16 fibroblasts). In total, we measured gene-expression levels for 100 and 106 expressed HSA21 genes in the two cell types.

On average, the steady-state RNA levels of genes on HSA21 in trisomy 21 is expected to be 1.5-fold that in unaffected individuals. This is close to what we observed here: the mean overexpression of HSA21 genes was 1.44-fold and 1.67-fold in LCLs and fibroblasts, respectively. Interestingly, however, only 39% and 62% of genes in the two cell types were found to be significantly overexpressed ($P < .01$, by KW test, after correction for multiple testing) in the trisomy 21 samples compared with in the euploid samples. Several mechanisms could explain this apparent discordance between the genomic dosage imbalance and the expression levels, such as negative feedback, epigenetic dosage compensation, or gene-expression variation. Because we and others previously reported extensive levels of interindividual gene-expression variation of HSA21 genes,[9,30,31] we have assessed the relationship between HSA21 gene overexpression and natural expression variation (measured as the CV). We found a wide distribution of CVs for HSA21 gene expression in both cell types (range 0.14–>1), which is consistent with previously published data,[3,16] with ~75% of genes having a CV <0.5. Regression analysis clearly showed that genes with higher CV tended to be less overexpressed in DS samples, and vice versa, suggesting that gene-expression variation explains, to a large extent, why many HSA21 genes are not significantly upregulated in trisomy 21.

One major objective of this study was to categorize HSA21 genes according to the degree of overlap of gene expression between trisomy 21 and euploid samples, because we hypothesized that a threshold effect of gene expression might partially explain the phenotypic variation among individuals with DS. We classified the genes tested into three groups, using the D statistic of the KS test that measures the degree of overlap. Genes in group A have minimal expression overlap between DS and normal samples and are candidates for involvement in DS phenotypes that are present in all affected individuals, such as mental retardation, muscle hypotonia, and Alzheimer disease. This is because the expression levels of this gene group in DS samples are consistently higher than those in control samples. Genes in this group (in both cell types tested) include *PRMT2,* involved in nuclear factor κB signaling and apoptosis[32]; *SUMO3,* encoding a protein involved in posttranslational modification of proteins (including p53), many of which are linked to senescence, DNA repair, and apoptosis[33]; and *MCM3AP,* involved in the regulation of DNA replication.[34] In addition, four genes involved in the *interferon-IL10RB* signaling pathway (*IFNAR1, IFNAR2, IL10RB,* and *IFNGR2*), which have been kept in a single syntenic block throughout vertebrate evolution, are in group A in fibroblasts.[27]

Group B genes have partially overlapping expression distributions between DS and normal samples and are likely to be involved in the variable features of DS. The working hypothesis for these genes is that, if the level of expression reaches a given threshold (that must be above the highest value observed for unaffected individuals), then the probability of the manifestation of a given phenotype is very high. This group includes a substantial number of genes (about half in LCLs and one-third in fibroblasts), such as *SON,* encoding a DNA-binding protein likely to be involved in cellular defense against hepatitis B virus[35]; *ETS2,* an oncogene involved in the normal embryonic anteroposterior axis and skeletal development[36,37]; and *PDXK,* which is involved in vitamin B6 metabolism.[38,39]

Genes in group C show extensive overlap in the distri-

Figure 3. Box plots of normalized expression levels of 91 HSA21 genes expressed in both LCLs and fibroblasts. The *Y*-axis is normalized expression values, with data points in the range 0–3.5; the *X*-axis is the four cell/genotype groups (LN = normal LCLs; LD = DS LCLs; FN = normal fibroblasts; FD = DS fibroblasts). Each panel represents a gene (*shown on top*).

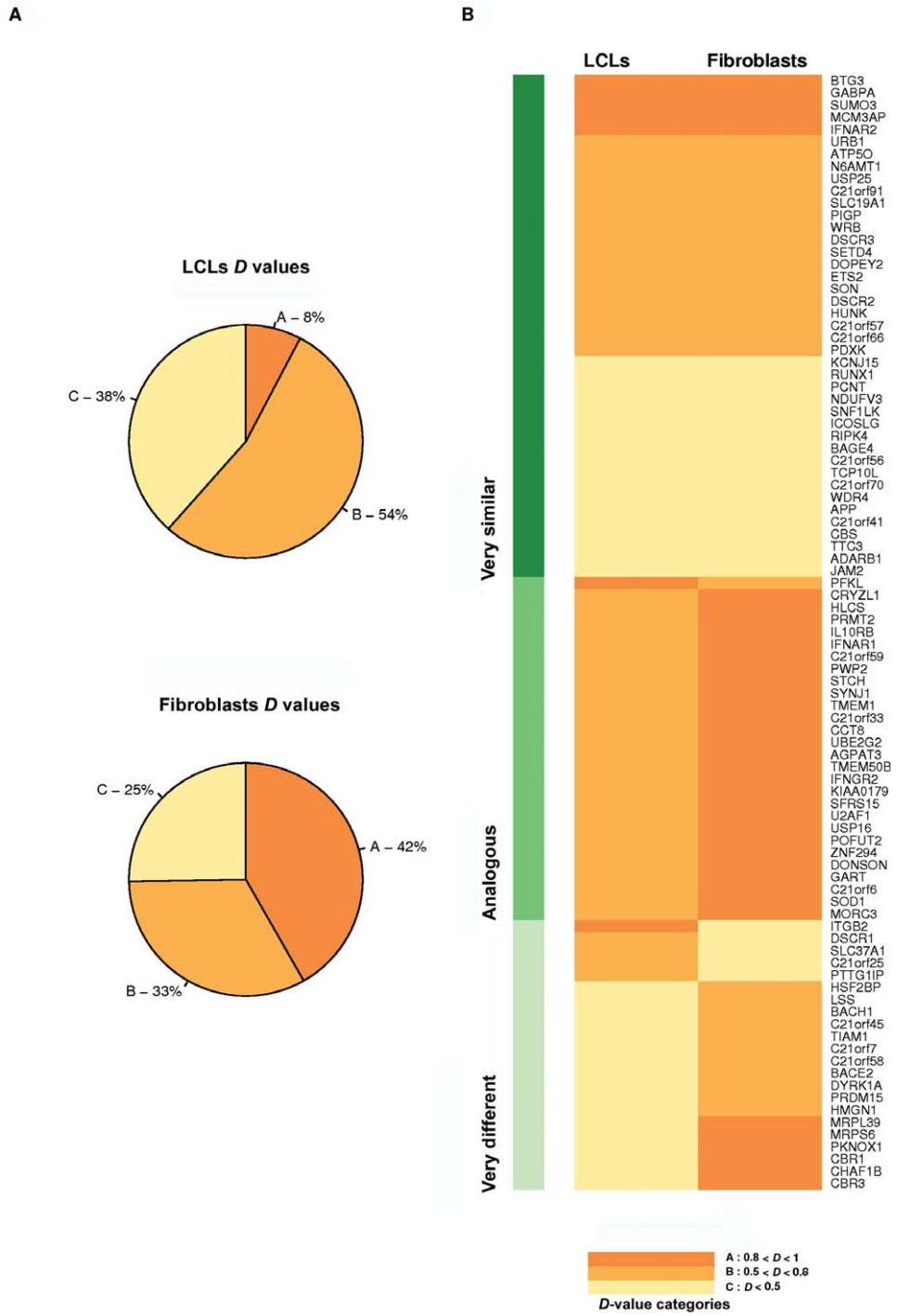


Figure 4. *A,* Pie charts of *D* values between the three groups of genes A, B, and C in LCLs and fibroblasts. Group A contains genes with minimal expression overlap between DS and normal samples, group B contains genes with partial overlap, and group C contains genes that show extensive overlap of expression values between DS and control samples. *B,* Classification of 91 HSA21 genes on the basis of their expression overlap between trisomy 21 and the euploid samples in the two cell types (LCLs and fibroblasts). Each gene is grouped into one of three categories of similarity—"very similar," "analogous," and "very different"—on the basis of the similarity of *D* values between the two cell types.

The figure is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Figure 5. Box plots of normalized expression levels of 9 HSA21 genes expressed only in LCLs and of 15 HSA21 genes expressed only in fibroblasts. The legend is available in its entirety in the online edition of *The American Journal of Human Genetics*.

bution of expression values between DS and control samples and are effectively dosage insensitive in the cell types analyzed. Examples include *ADARB1,* encoding an RNA-editing enzyme[40]; *JAM2,* encoding an adhesion molecule involved in the formation of tight junctions and transendothelial migration[41]; and *ICOSLG,* involved in T-cell regulation and acute immunoresponse.[42] Although genes in this category are less likely to be involved in DS phenotypes, it would be naive to exclude completely their involvement in DS, since some of these may have regulatory mechanisms that are specific to cell type or developmental stage. Clear examples of genes that are in group C in the two cell types studied but are known to contribute to DS features include *APP,* which has been shown to be dosage sensitive in the brain,[43] with overexpression resulting in early-onset Alzheimer disease, and *CBS,* which alters homocysteine metabolism in plasma of individuals with DS.[44] This emphasizes the need to perform similar studies in additional cell types and tissues at different developmental stages. An additional limitation of this study is that not all known spliced isoforms of genes have been examined; it is possible that specific isoforms could be categorized in different expression-overlap groups.

Comparison of gene-expression overlap between DS and control samples (*D* values from KS test) for the 91 genes expressed in the two cell types revealed a significant level of correlation for a majority of genes (76%). The remaining genes, on the other hand, showed cell type–specific regulation—for example, *PKNOX1,* encoding a homeobox transcription factor involved in development,[45] showed consistent levels of overexpression in fibroblasts but not in LCLs, and the opposite was true for *ITGB2,* which is involved in cell adhesion and cell-surface–mediated signaling.[46]

The results of this study show, for the first time, the importance of considering natural gene-expression variation in the context of aneuploidies and provide a framework for gene classification. We propose to prioritize genes according to the degree of overlap in the distribution of expression levels between trisomic and control samples; genes in group A are most likely associated with constant DS phenotypes, and genes in group B with variable DS features.

Although *D* values were consistent across cell types for a majority of genes, performance of a similar analysis of well-matched human tissue collections would be highly informative. However, there are limitations precluding such studies with human samples. Mouse models of aneuploidy, both for inbred and outbred populations, could be used to perform detailed analyses in a wide variety of tissues and developmental stages. One potential drawback of this approach is that the gene-expression variation, much of which is genetically determined, may be different between human and mouse; thus, results obtained from mouse samples will not be necessarily applicable to human trisomy 21.

Acknowledgments

The laboratory of S.E.A. is supported by grants from the Swiss National Science Foundation, the NCCR Frontiers in Genetics, the European Union AnEUploidy, T21 targets, and BioSapiens projects, and the Child Care foundation. P.P.'s fellowship was partially funded by the Blanceflor Boncompagni-Ludovisi Foundation. The Galliera Genetic Bank is supported by the Italian Telethon project GTF4003. The laboratory of S.S. is supported by National Institutes of Health (NIH) grant R01 HD38979 and NIH/NCRR grant M01 RR00039.

Web Resources

The URLs for data presented herein are as follows:

Coriell Institute for Medical Research, http://www.coriell.org/
Ensembl, http://www.ensembl.org/
HUGO Gene Nomenclature Committee, http://www.gene.ucl.ac.uk/nomenclature/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for trisomy 21)
The R Project for Statistical Computing, http://www.r-project.org

References

1. Epstein CJ (ed) (1989) Down syndrome, trisomy 21. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. Metabolic basis of inherited disease. McGraw-Hill, New York, pp 291–326
2. Roizen NJ, Patterson D (2003) Down's syndrome. Lancet 361: 1281–1289
3. Lyle R, Gehrig C, Neergaard-Henrichsen C, Deutsch S, Antonarakis SE (2004) Gene expression from the aneuploid chromosome in a trisomy mouse model of Down syndrome. Genome Res 14:1268–1274
4. Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zabot MT, Antonarakis SE, Reymond A (2006) Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. Am J Hum Genet 79:332–341
5. Kahlem P, Sultan M, Herwig R, Steinfath M, Balzereit D, Eppens B, Saran NG, Pletcher MT, South ST, Stetten G, et al (2004) Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of Down syndrome. Genome Res 14:1258–1267
6. Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM (2007) Gene-expression variation within and among human populations. Am J Hum Genet 80:502–509
7. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet 33:422–425

8. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, Phillips JW, Sachs A, Schadt EE (2004) Genetic inheritance of gene expression in human cell lines. Am J Hum Genet 75:1094–1105
9. Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyan L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel CV, et al (2005) Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. Hum Mol Genet 14:3741–3749
10. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, et al (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1:e78
11. Arron JR, Winslow MM, Polleri A, Chang CP, Wu H, Gao X, Neilson JR, Chen L, Heit JJ, Kim SK, et al (2006) NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. Nature 441:595–600
12. Roper RJ, Baxter LL, Saran NG, Klinedinst DK, Beachy PA, Reeves RH (2006) Defective cerebellar response to mitogenic Hedgehog signaling in Down syndrome mice. Proc Natl Acad Sci USA 103:1452–1456
13. FitzPatrick DR (2005) Transcriptional consequences of autosomal trisomy: primary gene dosage with complex downstream effects. Trends Genet 21:249–253
14. Gross SJ, Ferreira JC, Morrow B, Dar P, Funke B, Khabele D, Merkatz I (2002) Gene expression profile of trisomy 21 placentas: a potential approach for designing noninvasive techniques of prenatal diagnosis. Am J Obstet Gynecol 187:457–462
15. Mao R, Wang X, Spitznagel EL Jr, Frelin LP, Ting JC, Ding H, Kim JW, Ruczinski I, Downey TJ, Pevsner J (2005) Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. Genome Biol 6:R107
16. Mao R, Zielke CL, Zielke HR, Pevsner J (2003) Global upregulation of chromosome 21 gene expression in the developing Down syndrome brain. Genomics 81:457–467
17. Tang Y, Schapiro MB, Franz DN, Patterson BJ, Hickey FJ, Schorry EK, Hopkin RJ, Wylie M, Narayan T, Glauser TA, et al (2004) Blood expression profiles for tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome. Ann Neurol 56:808–814
18. Chung IH, Lee SH, Lee KW, Park SH, Cha KY, Kim NS, Yoo HS, Kim YS, Lee S (2005) Gene expression analysis of cultured amniotic fluid cell with Down syndrome by DNA microarray. J Korean Med Sci 20:82–87
19. Li CM, Guo M, Salas M, Schupf N, Silverman W, Zigman WB, Husain S, Warburton D, Thaker H, Tycko B (2006) Cell type-specific over-expression of chromosome 21 genes in fibroblasts and fetal hearts with trisomy 21. BMC Med Genet 7: 24
20. Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S (2004) Chromosome 21 and Down syndrome: from genomics to pathophysiology. Nat Rev Genet 5:725–738
21. Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, et al (2000) The DNA sequence of human chromosome 21. Nature 405:311–319
22. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3:34.1–34.11
23. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, et al (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848–853
24. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Stat Med 9:811–818
25. Amano K, Sago H, Uchikawa C, Suzuki T, Kotliarova SE, Nukina N, Epstein CJ, Yamakawa K (2004) Dosage-dependent over-expression of genes in the trisomic region of Ts1Cje mouse model for Down syndrome. Hum Mol Genet 13:1333–1340
26. Saran NG, Pletcher MT, Natale JE, Cheng Y, Reeves RH (2003) Global disruption of the cerebellar transcriptome in a Down syndrome mouse model. Hum Mol Genet 12:2013–2019
27. Carter KL, Cahir-McFarland E, Kieff E (2002) Epstein-Barr virus-induced changes in B-lymphocyte gene expression. J Virol 76:10427–10436
28. Reboul J, Gardiner K, Monneron D, Uze G, Lutfalla G (1999) Comparative genomic analysis of the interferon/interleukin-10 receptor gene cluster. Genome Res 9:242–250
29. Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nature Genet 32:261–266
30. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG (2007) Common genetic variants account for differences in gene expression among ethnic groups. Nature Genet 39:226–231
31. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nature Genet 33: 422–425
32. Ganesh L, Yoshimoto T, Moorthy NC, Akahata W, Boehm M, Nabel EG, Nabel GJ (2006) Protein methyltransferase 2 inhibits NF-κB function and promotes apoptosis. Mol Cell Biol 26:3864–3874
33. Li T, Santockyte R, Shen RF, Tekle E, Wang G, Yang DC, Chock PB (2006) Expression of SUMO-2/3 induced senescence through p53- and pRB-mediated pathways. J Biol Chem 281: 36221–36227
34. Takei Y, Assenberg M, Tsujimoto G, Laskey R (2002) The MCM3 acetylase MCM3AP inhibits initiation, but not elongation, of DNA replication via interaction with MCM3. J Biol Chem 277:43121–43125
35. Sun CT, Lo WY, Wang IH, Lo YH, Shiou SR, Lai CK, Ting LP (2001) Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. J Biol Chem 276:24059–24067
36. Georgiades P, Rossant J (2006) Ets2 is necessary in trophoblast for normal embryonic anteroposterior axis development. Development 133:1059–1068
37. Raouf A, Seth A (2000) Ets transcription factors and targets in osteogenesis. Oncogene 19:6455–6463
38. Wrenger C, Eschbach ML, Muller IB, Warnecke D, Walter RD (2005) Analysis of the vitamin B6 biosynthesis pathway in the human malaria parasite *Plasmodium falciparum*. J Biol Chem 280:5242–5248
39. Shin JH, Weitzdoerfer R, Fountoulakis M, Lubec G (2004) Expression of cystathionine β-synthase, pyridoxal kinase, and ES1 protein homolog (mitochondrial precursor) in fetal Down syndrome brain. Neurochem Int 45:73–79
40. Mittaz L, Scott HS, Rossier C, Seeburg PH, Higuchi M, Antonarakis SE (1997) Cloning of a human RNA editing de-

aminase (ADARB1) of glutamate receptors that maps to chromosome 21q22.3. Genomics 41:210–217
41. Johnson-Leger CA, Aurrand-Lions M, Beltraminelli N, Fasel N, Imhof BA (2002) Junctional adhesion molecule-2 (JAM-2) promotes lymphocyte transendothelial migration. Blood 100: 2479–2486
42. Hu W, Janke A, Ortler S, Hartung HP, Leder C, Kieseier BC, Wiendl H (2007) Expression of CD28-related costimulatory molecule and its ligand in inflammatory neuropathies. Neurology 68:277–282
43. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, et al (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24–26
44. Pogribna M, Melnyk S, Pogribny I, Chango A, Yi P, James SJ (2001) Homocysteine metabolism in children with Down syndrome: in vitro modulation. Am J Hum Genet 69:88–95
45. Sanchez-Font MF, Bosch-Comas A, Gonzalez-Duarte R, Marfany G (2003) Overexpression of FABP7 in Down syndrome fetal brains is associated with PKNOX1 gene-dosage imbalance. Nucleic Acids Res 31:2769–2777
46. Mayadas TN, Cullere X (2005) Neutrophil β_2 integrins: moderators of life or death decisions. Trends Immunol 26:388–395