



HAL
open science

Classification of variations of HSA21 gene expression in down syndrome lymphoblastoid cell lines : impact on disease phenotype

E. Aït Yahya-Graison, Julie Aubert, L. Dauphinot, Isabelle Rivals, Marguerite Prieur, G. Golfier, J. Rossier, L. Personnaz, N Créau, Stephane Robin, et al.

► To cite this version:

E. Aït Yahya-Graison, Julie Aubert, L. Dauphinot, Isabelle Rivals, Marguerite Prieur, et al.. Classification of variations of HSA21 gene expression in down syndrome lymphoblastoid cell lines : impact on disease phenotype. *American Journal of Human Genetics*, 2007, 81 (3), pp.475-491. hal-01208111

HAL Id: hal-01208111

<https://hal.science/hal-01208111v1>

Submitted on 1 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification of variations of HSA21 gene expression in Down syndrome lymphoblastoid cell lines: impact on disease phenotype.

E. Aït Yahya-Graison^{1,2}, J. Aubert³, L. Dauphinot¹, I. Rivals⁴, M. Prieur⁵, G. Golfier¹, J. Rossier¹, L. Personnaz⁴, N. Créau², S. Robin³, J. Delabar², and M.-C. Potier¹

¹Neurobiologie et Diversité Cellulaire, CNRS UMR 7637, ESPCI, 10 rue Vauquelin 75005 Paris, France ; ²EA3508 Université PARIS 7-Denis Diderot, 2, place Jussieu 75251 Paris Cedex, France ; ³UMR ENGREF/INA PG/INRA 518, 16, rue Claude Bernard 75231 Paris Cedex 05, France ; ⁴Equipe de Statistique Appliquée, ESPCI, 10 rue Vauquelin 75005 Paris, France, ⁵Service de Cytogénétique, Hôpital Necker Enfants Malades 75015 PARIS, FRANCE.

Rajouter ef Reiner

INTRODUCTION

Down syndrome (DS) results from the triplication of chromosome 21 (HSA21) and is the most common genetic cause of mental retardation in humans, occurring in ~1 in 800 newborns. The phenotype of DS is characterized by more than 80 clinical features such as cognitive impairments, muscle hypotonia, short stature, facial dysmorphism and several other anomalies [1]. The complexity of the disease is due to the fact that these clinical features can vary considerably in number and in severity [2]. Moreover, several pathologies such as congenital heart disease, acute megakaryoblastic leukemia or Hirschsprung disease occur with higher frequencies in DS patients than in the general population.

The trisomy of HSA21 as the cause of DS has been known since 1959 when Lejeune and colleagues demonstrated the presence in three copies of a G group chromosome now known as HSA21 in trisomic patients [3]. However, molecular mechanisms by which genomic dosage imbalance results in disruption of the phenotype remain not yet completely understood. Two different hypotheses have been proposed to explain the phenotype of DS patients: "developmental instability" (loss of chromosomal balance) and "gene dosage effect". According to the "developmental instability" hypothesis, the presence of a supernumerary chromosome globally disturbs the correct balance of gene expression in DS cells during development [4;5]. However, this hypothesis is weakened by the fact that autosomal trisomy syndromes do not lead to the same clinical pattern (Schinzel A. 2001). Moreover, correlations between genotype and phenotype in patients with partial trisomies indicate that a restricted region in 21q22.2 is associated with the main features of DS including hypotonia, short stature, facial dysmorphies and mental retardation [6;7]. This "DS critical region" supports the alternative "gene dosage effect" hypothesis postulating that

the restricted number of genes from chromosome 21 that are over-expressed in patients with partial trisomies contributes to the phenotypic abnormalities.

In order to find which hypothesis participates in the etiology of DS, several gene expression studies in human DS cells or tissues have been conducted [8-15]. Most of these studies have concluded to a primary global up regulation of the three copy genes mapping to the trisomic chromosome. However, the limited number of samples restricted the analysis to a global view and could not allow the identification of specific deregulated genes in DS with relevant statistical power. Moreover, these studies have been performed on a small number of three copy genes.

Several other experiments have been done on animal models of DS on a higher number HSA21 gene orthologs using microarray and quantitative PCR experiments [16-19]. In these studies three copy genes were over-expressed with a mean 1.5 ratio proportional to the gene dosage imbalance. But several triplicated genes appeared to escape the "1.5-fold" rule. However, these animal models are not trisomic for all HSA21 orthologs. Thus a complete and comprehensive classification of all HSA21 genes according to their level of expression in DS did not exist yet.

The goal of the present study was to fulfill this gap. For this purpose, an oligonucleotide microarray was designed containing all HSA21 genes, orf and predictions listed in most common HSA21 databases (GenBank, Ensembl, Eleanor Roosevelt Institute database, Max Planck Institute database) except the gene clusters corresponding to the keratin-associated proteins (i.e 53 genes) . The study was conducted on lymphoblastoid cell lines (LCLs) from 10 DS patients and 11 control individuals since these cells are easy to obtain and are widely used for genotype-phenotype correlations. To our knowledge, this is the largest study so far

realized on triplicated genes in DS human cells. In addition we applied a new statistical method for finding genes that are significant (ie significantly differentially expressed) but also genes that are unchanged (ie significantly not differentially expressed).

Our data show a global overall dosage-dependant expression of HSA21 genes in LCLs. However HSA21 genes could be classified into four classes: Class I - genes over-expressed with a mean ratio very close to 1.5 corresponding to the gene dosage effect; Class II - genes over-expressed with ratios above 1.6 reflecting an amplification mechanism; Class III - genes over-expressed with ratios below 1.4 corresponding to compensated genes and Class IV -genes that are significantly unchanged but with expression levels highly variable between individuals.

RESULTS

Design of a comprehensive HSA21 oligoarray

The HSA21 oligoarray is a tool designed for the exhaustive study of HSA21 gene expression in DS. This microarray contains 664 sequences representing 145 genes, 58 orf, 118 predictions with their reverse sequences for 20 of them and 18 antisense transcripts allowing the expression analysis of all putative genes mapped on HSA21. To increase the robustness of the results, when possible, at least two probes per gene were designed (84% of the HSA21 oligoarray content). The description of the HSA21 oligoarray content according to BLAST results performed on the latest version of the human genome sequence (www.ncbi.nlm.nih.gov, version 36.2) is summarized in Table 1. Oligonucleotide sequences spotted on the array have been designed according to four main criteria: specificity for the

represented sequence, GC content equal to 50%, melting temperature allowing an optimal match between the probe and the target at the hybridization temperature, and no stable predicted secondary structure. All these criteria are compiled by the Selection of OLigonucleotides (SOL) software (Golfier *et al.*, under submission à ajouter dans Biblio) which was used in the present study.

Using this new analysis tool, 40 differential hybridizations comparing DS and control LCLs were performed. The mean signal intensities on each array spot indicated the expression levels of HSA21 genes. Out of the 359 genes represented on the array, 132 showed signal intensity above a background cut-off level (mean A > 7). LCLs thus express approximately one third of the HSA21 gene content.

Biological material from patients and control individuals

LCLs have been obtained after immortalization by Epstein-Barr virus (EBV) of B lymphocytes collected from blood samples of DS patients and control individuals. In order to make sure that EBV-transformation did not induce any chromosomal rearrangement, all cell lines were karyotyped after immortalization. In order to limit technical bias, culture conditions were standardized. Cell lines were always maintained in exponential growth phase. No significant difference on cell morphology or cell proliferation has been observed between DS and control LCLs.

Experimental design and statistical analysis

The main objective of the microarray experiments was to detect differentially expressed genes between DS and control samples taking into account the variability between

individuals. A second objective was to find genes whose expression was affected by sex alone or by sex and DS. The aim of the experimental design was to respect the experimental constraints and adapt them to the study. Experimental design (see Material and Methods) for the 40 differential hybridizations is summarized in Table 2.

First, a mixed model was constructed to highlight the effects of DS, sex, sex and DS and to take into account the variability between individuals in HSA21 gene expression. We used the Benjamini-Hochberg procedure to adjust the p-values obtained and to limit false positives due to multiple testing [27]. We set the False Discovery Rate (FDR) at 0.05, thus the list of significant genes was expected to contain 5% false positives.

This analysis identified HSA21 genes that were significantly differentially expressed between DS and control LCLs. No HSA21 genes had significant p-values when sex, or DS and sex combined together were tested. In other words, HSA21 gene expression was not significantly different between men and women. In addition, DS effects on HSA21 gene expression were not dependant on sex. The two effects sex and sex/DS were thus dropped from the model and a simplified mixed model testing the effects of DS on HSA21 gene expression was finally used. We set again the FDR at 0.05. One hundred and thirty two HSA21 genes gave signal intensities above the background cut off ($A \text{ mean} > 7$) and 58 exhibited a significant over-expression in DS LCLs as compared to controls (Fig.1). For these genes the DS/control expression ratios ranged from 1.25 to 2.27 with a mean ratio of 1.5 reflecting the gene dosage imbalance in DS (Fig.2A). Significantly over-expressed genes (SOE) are listed in Table 3. None of them had ratio below 1, indicating that no HSA21 genes are significantly under-expressed in DS LCLs.

The goal of this study was also to demonstrate whether HSA21 gene expression profiles could differentiate DS from control samples. We therefore performed two distinct principal component analyses (PCA) on HSA21 genes and non HSA21 genes used as references.

PCA applied to reconstructed data based on the expression levels of HSA21 genes clearly separated DS samples from controls suggesting that the effects of DS on HSA21 gene expression prevails over any other effect including biological variability (Figure 3A). Indeed, no distinction could be obtained between DS and control individuals when PCA was conducted on non HSA21 genes (Figure 3B). Non HSA21 genes had mean DS/control expression ratio of 1 (Fig. 2C).

In order to classify HSA21 genes according to their expression variations in DS, we tested HSA21 gene expression data to detect genes that were significantly unchanged in DS LCLs. For this purpose, we used a statistical procedure that calculates local FDR for each tested gene [20]. The local FDR is the probability that a gene, whose raw p-value is P_i , is not differentially expressed, taking into account the whole set of tests. Genes with local FDR equal to 1 can be considered as unchanged, ie not differentially expressed (NDE) between DS and control samples. We used raw p-values deduced from the simplified mixed model analysis to deduce local FDR for each gene on the array. Results indicated that 42 HSA21 genes and predictions were significantly unchanged (NDE) (Table 4). The mean DS/control ratio for these invariant genes is equal to 1.08. Distribution of DS/control ratios is shown in Figure 2B.

All the other genes that were neither significantly over-expressed (SOE) nor non differentially expressed (NDE) were non significant (NS) (Supplementary data Table 1). DS/control expression ratio of NS genes ranged from 0.81 to 1.60.

Quantitative PCR validation experiments

To confirm variations in HSA21 gene expression between DS and control LCLs obtained from microarray experiments, we performed quantitative real-time PCR (QPCR) on 10 HSA21 genes selected in the 3 following categories: 4 significantly over-expressed genes (SOE), 2 non differentially expressed (NDE) and 4 non significant genes (NS) (Table 5). QPCR were conducted on all LCL from DS patients and controls. Normalized mean DS/control ratios from QPCR were in agreement with the ratios obtained on HSA21 microarrays with a correlation coefficient of 0.73. Statistical t-tests performed on QPCR values confirmed the significant over-expression of 3 genes in DS samples as compared to controls (*DYRK1A*, *SNF1LK*, *STCH*). Moreover, NDE genes were also confirmed (*H2BFS*, *DSCR1*). These genes had QPCR ratios close to 1 and were not significant with t-tests. Two additional genes were found to be over-expressed by QPCR but were non significant (NS) on HSA21 microarray (*MX1*, *SOD1*). *TMEM1* and *CSTB* were non significant (NS) with both methods while *DSCR1*, *H2BFS* and *CHAF1B* were NS only by QPCR. However *DSCR1*, *H2BFS* were non differentially expressed on HSA21 microarrays. With QPCR, we were not able to distinguish between NDE and NS genes.

MATERIAL AND METHODS

Cell lines and culture conditions

Lymphoblastoid cell lines (LCLs) were derived from B lymphocytes of 10 DS patients (DS) from the cytogenetic service of *Necker Enfants Malades* Hospital and the *Jérôme Lejeune* medical center. Patient's parents from the *Jérôme Lejeune* medical center gave their informed consent and the French bio-medical ethic committee gave its approval for this study (CCPPRB n°03025). Written informed consent was obtained from the participants or from their families by the cytogenetic service of *Necker Enfants Malades* Hospital. Cell lines from 11 control individuals were also obtained for comparison of chromosome 21 gene expression profiles. Culture media consisted of Opti-MEM® with GlutaMax® supplemented with 5% fetal bovine serum and 1% penicillin/streptomycin (10000 u/ml). Cell lines were grown at 37°C in humidified incubators, in an atmosphere of 5% CO₂. Each culture was grown until having at least 60x10⁶ cells. All 21 cell lines were karyotyped in order to confirm either trisomic or euploid status and also to verify that immortalization by the Epstein-Barr virus did not produce any visible chromosomal rearrangement other than trisomy 21 (TS21). Cells were harvested by centrifugation, washed in 5 ml PBS followed by another centrifugation and stored at -80°C.

HSA21 oligoarray

A dedicated oligonucleotide microarrays, named HSA21 oligoarray, containing 664 50mers aminomodified oligonucleotides representing 145 genes, 58 orf, 118 predictions (20 of these were represented in both orientations) and 19 antisense transcripts assigned to chromosome 21 (HSA21) were used in the present study. Among predictions, cDNAs and isolated exons

in the CBR-ERG region on HSA21q deduced from cDNA isolation and exon trapping experiments [6] were represented on the array and other gene or exon predictions were produced from *in silico* analysis of the complete sequence of HSA21 [21]. Unpredicted transcript sequences and antisense transcripts isolated either by Kathleen Gardiner or by Alexander Reymond were also included in this oligoarray [22-24]. Thirty nine genes assigned to chromosomes other than HSA21 represented by 58 oligonucleotides were added for data normalization. All probes present on the array were designed using the SOL software (Golfier *et al.*, submitted A AJOUTER DANS BIBLIO). Sequences were then synthesized by EuroGentec and spotted onto CodeLink® activated glass slides (Amersham Biosciences) using a MicroGrid II spotter (Biorobotics). Each array contained 2 matrices with 8 blocks each, in which probes were present in duplicates so that each oligonucleotide was present in four replicates on each slide.

Experimental design

The experimental constraints were the following: 10 DS patients (7 men and 3 women) and 11 controls (4 men and 7 women). Samples from the same individual were sometimes used in different hybridizations.

For each gene, we used the following linear model:

$$y_{ijklm} = \mu + D_i + S_j + A_l + F_m + DS_{ij} + DF_{im} + SF_{jm} + I(DS)_{ijk} + \varepsilon_{ijklm}$$

where y_{ijklm} is the normalized expression of the gene in \log_2 for factor i (DS or control), sex j (male or female), patient number k ($k=1, \dots, 21$), labeled with dye m ($m=$ red for Cy5 or green for Cy3) on the HSA21 oligoarray l . The symbols D, S, A, F represented the fixed effects due to the disease, sex, array and fluorochrome respectively. The symbol $I(DS)$ referred to the

patient (nested within disease and sex) random effect. This last effect accounted for the correlation between samples used in different hybridizations but harvested within the same patient.

We assumed the independence between all $I(DS)_{ijk}$ and E_{ijklm} . We also assumed that $I(DS)_{ijk}$ were independent with a distribution $N(0, s^2)$ and that E_{ijklm} were independent with distribution $N(0, \sigma^2)$.

The model can be rewritten under the following matrix form:

$$\text{(Matrix 1) } Y = X\theta + ZU + E$$

where θ was the vector of fixed effects (D, S, A, F, DS, DF, SF), U was the vector of $I(DS)_{ijklm}$ and E the vector of E_{ijklm} . $Y \sim N(X\theta, \Sigma)$ where $\Sigma = 2\sigma_g^2 \text{Id} + s^2_g ZZ^T$. Y had n (total number of samples) rows and 1 column, X was the matrix describing the status of the patient (disease, sex) in which the sample had been harvested. Z had n rows and I (total number of patients) columns and described the correspondence between samples and patients.

On each array l , we actually observed the differential expression (red signal - green signal):

$$\begin{aligned} \mathcal{Y}_{ii'jj'kk'lm} &= \mathcal{Y}_{ijklm} - \mathcal{Y}_{i'j'k'l'm'} \\ &= (D_i - D_{i'}) + (S_j - S_{j'}) + (DS_{ij} - DS_{i'j'}) + (F_m - F_{m'}) \\ &\quad + (DF_{im} - DF_{i'm'}) + (SF_{im} - SF_{i'm'}) + (I(DS)_{ijk} - I(DS)_{i'j'k'}) \\ &\quad + (\varepsilon_{ijklm} - \varepsilon_{i'j'k'l'm'}) \end{aligned}$$

This model can be rewritten under another matrix form: we define matrix Δ describing the comparisons performed on each array. This matrix had L (total number of arrays) rows and n columns. The l -th row of Δ was full of zero except a +1 in the column corresponding to the sample labeled in red and a -1 in the column corresponding to the sample labeled in green.

The model for the differential expression is obtained by multiplying all terms of matrix 1 by

Δ :

$$\text{(Matrix 2)} \Delta Y = \Delta X \theta + \Delta Z U + \Delta E$$

The vector of differential expression ΔY had a distribution $N(\Delta X \theta, \Delta \Sigma \Delta^T)$.

The experimental design was defined by the three matrices X , Z and Δ . X and Z basically depended on the number of samples for each patient. Because the microarrays we used were two-color assays, the total number of samples was twice the number of slides. We thus chose 4 replicates for each patient.

The important remaining choice was the comparison to be made, i.e. the choice of Δ . Considering the differential expression between two patients on an array eliminated the array effect and the constant which were not of interest. We were neither interested in other technical effects like fluorochrome or interactions of fluorochrome with other effects. In order to eliminate DF_{im} and SF_{jm} effects, we proposed a balanced design for the fluorochrome effect. Finally, data were normalized across genes in order to remove the dye effect F_m .

The last criterion to choose was Δ , the precision of the estimated effects (gathered into the vector θ). According to the mixed linear model theory, this precision is given by its variance matrix

$$V(\hat{\theta}_A) = (\Delta^T X^T (\Delta \Sigma \Delta^T)^{-1} \Delta X)^{-1}$$

which depends on Δ and on ratio σ^2/s^2 . The diagonal contained all the informations about the quality of the estimates. It gave the variance of the estimates of all effects of interest D_i , S_j , DS_{ij} . This matrix was the ultimate tool to compare designs.

We calculated the variance of the disease effect (which is of main interest) and the determinant of the variance covariance matrix (which gives a global measure of the

precision of the estimates) for a certain number of designs Δ . To do that, we used a value given a priori for the ratio σ^2/s^2 equal to 2.

We finally chose a design involving 40 arrays. Samples from a same patient are marked the same number of times with each fluorescent dye (Cy3 or Cy5). On each array a DS LCL and a control LCL were compared to increase the precision of the disease effect. Ten arrays compared a man with DS versus a man without DS, 10 a woman with DS versus a woman without DS, 10 a man with DS versus a woman without DS and 10 a woman with DS versus a man without DS. The design is described in Table 2.

mRNA extraction, HSA21 oligoarray hybridization, data filtering and normalization

mRNAs were extracted from frozen individual cell samples using FastTrack 2.0 mRNA Isolation kit (Invitrogen) according to the manufacturer's instructions. To eliminate DNA contamination, the appended DNase protocol of RNeasy mini kit (Qiagen) was used according to manufacturer's protocol. Samples were further tested for purity and quantity on RNA 6000 NanoChips using the Agilent 2100 Bioanalyzer (Agilent Technologies). Two μg of mRNA were converted into either Cy3 or Cy5 labeled cDNA by incorporation of fluorescent dUTP (Amersham) using the Reverse-iT RTase Blend kit (ABGene). Labeled targets were then purified on Qiaquick column according to the manufacturer's protocol (Qiagen). Hybridizations of samples pairs on HSA21 oligoarrays (one DS and one control) according to the experimental design were carried out in hybridization buffer (formamide 50%, SSC 4X, SDS 0.1%, Denhart 5X) at 42°C over night. Data were acquired with GenePix 4000B scanner and using the GenePix Pro 6.0 software (Axon). For each array, the raw data comprised the median feature pixel intensity at wavelength 635 and 532nm for Cy5 and Cy3

labeling respectively. After subtraction of the background signal, LOWESS normalization of the M/A plots was applied to all oligonucleotides representing non HSA21 genes and used to calculate a correction factor applied to M values for HSA21 probes. Normalized data from each slide were then filtered using two criteria: i) for each oligonucleotide, at least two values among the four replicates had to be available, and ii) standard deviation of A values corresponding to the geometric mean in \log_2 of Cy3 and Cy5 signal intensities had to be below 1. Arithmetic means of normalized and filtered M and A values were calculated for each oligonucleotide and submitted to the statistical analysis.

Expression data analysis: statistical testing

Mixed model

To determine differentially expressed genes for DS, sex and DS*sex effects, we performed a mixed-model analysis of variance according to the experimental design. This method was chosen in order to distinguish between inter-individual variability and experimental variability [25]. We used the Mixed Procedure of the SAS software® [26] with the Restricted Maximum Likelihood (REML) method of estimation.

After the filtering and normalization steps, the number of observations per spot varied between 8 and 40 which was enough to calculate the variance for each gene.

We first tested the effects of the complete model (Model 1). As the sex and the sex*DS effects were not significant for any gene, these two effects were dropped from the model.

We finally analyzed the simplified model:

$$\begin{aligned}
 (Model\ 2) \quad y_{i'j'k'k'l} &= y_{ijkl} - y_{i'j'k'l} \\
 &= (D_i - D_{i'}) + (I(DS)_{ijk} - I(DS)_{i'j'k'}) + (\epsilon_{ijklm} - \epsilon_{i'j'k'l'm'})
 \end{aligned}$$

The raw p-values were adjusted by the Benjamini Hochberg procedure [27], which controls the False Discovery Rate (FDR).

PCA Analysis

Results from microarray experiments were obtained as differential expression between TS21 and control samples. M values correspond to $\log_2(\text{DS}) - \log_2(2\text{N})$ and A values to $(\log_2(\text{DS}) + \log_2(2\text{N}))/2$. For each probe p, the mean value of its expression (in log2) in DS cell lines and in controls can hence be expressed as:

$$\begin{cases} E_p^i = \frac{1}{N_i} \sum_{k \in S_i} \left(A_p^k + \frac{M_p^k}{2} \right) & \text{for } i=1 \text{ to } 10 \text{ (DS)} \\ E_p^i = \frac{1}{N_i} \sum_{k \in S_i} \left(A_p^k - \frac{M_p^k}{2} \right) & \text{for } i=11 \text{ to } 21 \text{ (controls)} \end{cases}$$

where i denotes the index of the individual, S_i the set of slides on which all samples from the individual i have been hybridized, and N_i the size of this set. Since the overall expression level of the genes of one individual varied from one slide to another, we rather considered equal A's across slides for each individual, i.e. $A_p^k = A_p$. Since PCA subtracts their mean values to the descriptors of each individual, it is equivalent to consider $A_p = 0$. We thus simplified the equation system and reconstructed relative mean values of expression E for each individual as:

$$\begin{cases} E_p^i = \frac{1}{N_i} \sum_{k \in S_i} \left(\frac{M_p^k}{2} \right) & \text{for } i=1 \text{ to } 10 \text{ (DS)} \\ E_p^i = \frac{1}{N_i} \sum_{k \in S_i} \left(-\frac{M_p^k}{2} \right) & \text{for } i=11 \text{ to } 21 \text{ (controls)} \end{cases}$$

Principal component analysis (PCA) on HSA21 genes and on non HSA21 genes were performed separately using E values deduced from all expressed HSA21 probes and all expressed non HSA21 probes respectively.

Non differentially expressed genes: local FDR

The raw p-values deduced from statistical analysis on microarray data were used to calculate an estimate of the FDR attached to each gene, called local FDR[20]. This method is implemented in R (www.r-project.org).

Quantitative PCR experiments

In order to validate expression ratios between DS and control samples obtained from HSA21 oligoarray data, 100 ng of mRNA were reverse transcribed into cDNA using Reverse-iT RTase Blend kit (ABgene). Quantitative real-time PCR (QPCR) on diluted cDNA were conducted in the presence of 0.6 mM of each specific primer (designed by Primer3 software) and 1X Quantitect™ SYBR® Green PCR master mix (Qiagen) containing 2.5 mM MgCl₂, HotstartTaq® Polymerase, dNTP mix and the fluorescent dye SYBR Green I. QPCR experiments were performed in a Lightcycler system (Roche Molecular Biochemicals) on ten HSA21 genes: *CHAF1B*, *CSTB*, *DSCR1*, *DYRK1A*, *H2BFS*, *Mx1*, *SNF1LK*, *SOD1*, *STCH* and *TMEM1*. The Ubiquitin-activating enzyme E1 (L13852) mRNA mapping to HSA3 and the Zinc finger protein (AB000468) mRNA mapping to HSA4 were used as endogenous control genes as described by Janel *et al.* [28]. For each sample, mean Ct was corrected by subtracting the mean of the Ct obtained with the two reference genes. Unpaired t tests ($p < 0.05$) were then performed between DS and control groups for each gene of interest.

Submission of microarray data to GEO repository

The GEO accession number for all of the microarray data is xxxxxxxxxxxx

DISCUSSION

The aim of the study was to analyze HSA21 gene expression in lymphoblastoid cell lines (LCLs) established from DS patients and control individuals. Differential hybridizations comparing DS LCLs to control LCLs were performed on a dedicated HSA21 oligoarray designed from the complete HSA21 gene catalogue. Approximately one third (n=132) of all HSA21 genes, orf and predictions were expressed in LCLs.

Among the expressed transcripts, nearly one half (n=58) were significantly over-expressed (SOE) in DS cell lines. One third (n=42) corresponded to non differentially expressed genes (NDE) with DS/control ratios significantly not different from 1. The remaining 33 expressed genes gave non significant (NS) expression ratios.

Based on the expression levels of HSA21 genes, DS samples were clearly distinct from controls thus reflecting the prevalent effect of DS on HSA21 gene expression. However, non HSA21 genes could not separate DS from control LCLs.

HSA21 genes can now be categorized into three classes according to their expression levels in DS cell lines relative to control.

The first class (Class I) contains 23 HSA21 genes with expression ratio DS/control close to 1.5 (1.4 to 1.6), correlated to the presence of three genomic copies (Table 3, Fig. 3). These class I genes could be responsible for the phenotype observed in DS either directly or indirectly through a secondary effect of *cis* or *trans* acting genes.

The second class (Class II) contains 15 HSA21 genes with expression ratio DS/control above 1.6 corresponding to the amplification of the initial gene dosage (Table 3, Fig. 3). Among these genes, *SAMSN1*, *SNF1LK*, *STCH*, *BTG3* show the highest expression ratio ranging from 1.79 to 2.27.

Gene dosage amplification could result from a cascading effect through regulation networks involving *trans*- or *cis*-acting genes [29]. Some Class II genes might be regulated at the transcriptional level by mechanisms which are themselves disturbed in DS hence the amplification of target genes initiated by one of the numerous regulation factors mapping to HSA21 such as the transcription factor *GABPA* which belongs to class I genes. Kinases assigned to HSA21 such as *BTG3*, *SNF1LK*, *DYRK1A*, involved in various signaling pathways could as well amplify the expression of HSA21 targets through a cascading effect. Pellegrini and colleagues [30] identified *in silico* a putative mitogen-activated kinase cascade involving *DYRK1A*, *SNF1LK*, *RIPK4* and *DSCR3*. In our study, *DYRK1A*, *SNF1LK* and *DSCR3* were significantly over-expressed (SOE) in LCL but not *RIPK4* which was not found to be expressed in LCLs. *DYRK1A* and *DSCR3* belong to the class I genes whereas *SNF1LK* is submitted to an amplification of the initial gene dosage (class II). Based on this putative mitogen-activated kinase cascade, amplification of *SNF1LK* gene expression could thus result from the over-expression of *DYRK1A* acting as a regulatory factor on *SNF1LK* in the cascade, *DSCR3* acting as a scaffolding protein.

Deutsch and colleagues (2005) analyzed gene expression variation of 34 HSA21 genes in normal individuals and found 26 genes with significant inter-individual differences. Of these genes, 7 (*ITGB2*, *CBS*, *PDXK*, *SLC37A1*, *BTG3*, *IFNAR2* and *CCT8*) belong to the Class II genes. Since these genes are significantly over-expressed with an amplification of the

gene dosage effect, we thus conclude that these class II genes with high expression variability might contribute to the DS phenotype. Significant eQTLs have been identified for nine highly variable genes using genome wide linkage analysis [31]. In particular, a *cis*-eQTL in 21q22.11 was identified for *CCT8* indicating a direct *cis*-acting gene for which *CCT8* is a target gene and this could explain the amplification of the initial gene dosage imbalance for this gene.

Alternatively the existence of allelic variants with higher levels of expression could result in the amplification of the gene dosage effect in DS, overtaking the phenotype threshold (Fig.4, ratio DS/control around 2). However, this hypothesis underlies the existence of linkage disequilibrium in DS that remains to be shown. Further investigations on a larger cohort of DS patients will be necessary to confirm the existence of such linkage disequilibrium in DS for these particular genes with elevated DS/control ratio.

The third class (Class III) contains 20 SOE HSA21 genes with DS/control expression ratio below 1.4 (Table 3, Fig. 1). These class III genes are under gene dosage compensation in DS. This compensation mechanism has already been suggested in previous transcriptome studies both in DS patients [9;13;14] and in mouse models [4;17-19]. Here, using a robust statistical method such as the mixed model analysis, we confirmed the existence of gene dosage compensation in DS. Compensation is most likely due to negative feedbacks that would modulate transcriptional activity or mRNA stability of class III genes. Thus the expression of compensated genes is strongly regulated by mechanisms that are not impaired in DS.

Fourty two HSA21 genes that are not differentially expressed (NDE) according to local FDR have highly variable DS/control expression ratio (var(M) in Table 4). These genes have

variable expression levels between individuals. One of them (*DSCR1*) has already been shown to be variable in control LCLs [31].

Variance distribution of NDE genes was compared to the variance of SOE and NS genes (Fig. 5). The three variance distributions were significantly different ($p < 0.001$) with the mean variance of SOE genes smaller than the NDE and NS genes. Interestingly among the NDE genes 12 had variance comparable to SOE genes (< 0.3). We suggest that these 12 genes with DS/control expression ratio from 0.82 to 1.22 belong to the Class-III compensated genes.

The remaining 30 NDE genes have highly variable DS/control ratio and are the Class IV-genes whose expression is highly variable between individuals.

Finally the last 33 genes that are neither significantly over-expressed (SOE) nor non differentially expressed (NDE) are non significant genes (NS) genes according to the statistical analysis used in our study (Fig. 1, Supplementary data Table X).

Among these, 8 genes had already been tested for inter individual expression variations by Deutsch and colleagues [31]. One of them, *SOD1*, was found to be non significant in our study because of high variations while inter individual differences were weak according to Deutsch *et al.* [31]. This could suggest that inter individual variability is higher in DS samples than in control samples. However *SOD1* is the only gene for which microarray data are not consistent with QPCR experiments probably because of differences in the alternative transcripts detected with the two methods (see below). The other seven genes tested by Deutsch *et al.* [31] (*CBR1*, *ETS2*, *GART*, *SLC19A1*, *TMEM1*, *UBASH3*, *WRB*) were found to be variable between individuals thus explaining why they were not significant (NS) in our

microarray study. The 25 remaining NS genes that were not analyzed by Deutsch and colleagues had high variance that could explain the absence of significant results and may reflect inter individual expression variations. Indeed distributions of $\text{var}(M)$ for the NS genes is shifted to higher values as compared to SOE genes (Fig. 5) . More data will be necessary for classifying NS genes into Class-III or Class-IV genes.

Only one of the 132 genes expressed in LCLs, *C21orf108* was represented by the two probes present on HSA21 oligoarray which gave different statistical results. One probe (name Supplementary data Table X avec toutes les datas) mapping to exon 39 of *C21orf108* was significantly over-expressed (SOE) with a DS/control ratio of 1.3. The other probe mapping to exon 26 of *C21orf108* was non differentially expressed with a DS/Control ratio of 1.09 with $\text{var}(M)$ under 0.3. This difference could result from the existence of two alternative transcripts containing either exon 26 or exon 39. However, the expression ratios obtained for the two probes allow us to classify this gene in the third Class as compensated gene.

The new classification described in our study was confirmed using QPCR assays on ten HSA21 genes. Ranking genes according to their DS/control expression ratios obtained from the HSA21 oligoarray experiments was validated by QPCR except for *SOD1*.

Microarray experiments showed that *SOD1* was not significantly over-expressed in DS LCLs (ratio 1.2) while *SOD1* was significantly SOE-expressed (2-fold) according to QPCR experiments. This difference could be due to the presence in LCLs of the two *SOD1* transcripts which are both amplified by QPCR whereas only the long and less abundant form can be detected on HSA21 oligoarray [32]. With a DS/control ratio of 2 *SOD1* should thus belong to the Class II genes (amplified genes).

The most complete published study measuring transcripts in DS patients was carried out by Mao and colleagues [14]. They performed transcriptome analysis using pangenomic Affymetrix U133A chip on fetal heart and cerebellum and astrocyte cells. These microarrays contain 253 probes representing about 200 genes assigned to HSA21. Of their most dysregulated HSA21 genes with high TS21/control ratios, 17 were in common with our 58 HSA21 genes that were significantly upregulated (SOE). The class I genes contained all 17 genes plus three other genes (*STCH*, *PDXK*, *SOD1*) showing that amplification of the initial gene dosage in DS occurs in all tissues or cells where these genes are expressed and that have been studied so far. This confirms that class I genes are very likely to be involved in the DS phenotype.

The use of LCLs appears relevant since 74% of the HSA21 genes dysregulated in TS21 fetal tissues and astrocyte cells - which are pathologically affected in DS - are also dysregulated in DS LCLs. In addition LCLs are much easier to obtain. For three DS patients, transcriptome comparisons between fresh blood samples and LCLs obtained from the same individuals were conducted on pangenomic microarrays from the *Réseau National des Génomiques* (20 000 human transcripts, [33]). From these experiments, no major alteration of the transcriptome by the EBV-transformation could be detected (only 1 to 2% of the represented genes exhibited significant differential expression, data not shown).

We have analyzed the distribution of expressed genes as well as SOE and NDE genes along HSA21. Fig. 6 shows that expressed genes map preferentially to the distal part of HSA21. In addition, among the expressed genes, the percentage of SOE genes is generally higher than the percentage of NDE genes. However two regions including the DS chromosomal region (DCR) contain more NDE genes than SOE genes. Finally the 5Mb interval proximal to the

DCR has a very high proportion of SOE genes. This region contains 3 clusters of genes belonging to the same class of SOE genes that could be co-regulated: IL10RB/IFNAR1/IFNGR2, SYNJ1/C21orf59/C21orf66/ C21orf49 and SON/DONSON.

The HSA21 oligoarray was also designed to analyze the effects of DS on the expression of antisense transcripts as well as alternative transcripts.

Fourteen antisense transcripts are present with their nesting genes on the HSA21 oligoarray. Among them, ten have been extracted from the HSA21 database established by Kathleen Gardiner at the Eleanor Roosevelt institute. The four remaining antisense transcripts correspond to transcribed sequences in the DCR that have been generated from various cDNA mapping and exon trapping experiments [6;34;35]. Six genes (*HLCS*, *C21orf25*, *CHAF1B*, *DYRK1A*, *KIAA0179*, *TTC3*) are expressed in LCL and are significantly over expressed (SOE) except *KIAA0179* which was found to be non differentially expressed (NDE). Four of the corresponding antisense transcripts were also found to be expressed in LCLs thus confirming their existence (*as-KIAA0179*, *as-DYRK1A*, *as-HLCS*, *as-TTC3*). *TTC3* and its antisense transcript were SOE whereas the other antisense sequences did not belong to the same class as their corresponding genes. In addition, two antisense sequences (*as-C21orf56*, *as-KCNJ6*) were expressed in LCLs whereas their corresponding genes were not. Probe referred as antisense transcript *as-KCNJ6* mapping in intron 3 of *KCNJ6* on the opposite strand corresponds to one of the transcribed sequences isolated in the DCR by exon trapping experiments [6]. Since there is no evidence that this sequence is an antisense transcript of *KCNJ6*, it could thus belong to a gene locus that has not yet been identified and map on the opposite orientation to *KCNJ6*.

According to the NCBI Gene database, *C21orf56* (GeneID=84221) currently maps on the negative strand of HSA21 but was previously annotated on the positive strand when the HSA21 oligoarray were designed. Thus probe sequence representing the antisense transcript *as-C21orf56* could rather correspond the actual sense of *C21orf56*. In conclusion, the expression of antisense transcripts is confirmed by our HSA21 oligoarray experiments and sense and antisense transcripts are not always similarly changed in DS

Lastly and in order to search for differential effects of DS on alternatively spliced transcripts we analyzed genes for which oligonucleotide probes present on the HSA21 oligoarray could differentiate between alternative transcripts. Seventeen genes had probes specific for alternative transcripts (Supplementary Table x). For 7 of those genes all the probes gave low signals indicating that these transcripts were not expressed in LCLs. Two genes (*C21orf33* and *MRPL39*) were expressed in LCLs as a unique transcript and belonged to class I genes that are significantly over-expressed (SOE) with a ratio close to 1.5. For the last eight genes (*ADARB1*, *C21orf34*, *C21orf66*, *DYRK1A*, *GART*, *PKNOX1*, *RUNX1* and *TMEM1*) oligonucleotide probes could distinguish between splicing variants that had DS/control ratio very similar. Only two of these genes (*C21orf66* and *DYRK1A*) were significantly SOE-expressed (SOE) in DS LCLs (Class I genes) while the others were not significant (NS). These results suggest that all transcripts belonging to the same gene and expressed in LCLs are similarly regulated in DS.

CONCLUSION

Using high content HSA21 oligoarray combined with a powerful statistical analysis of the data, we can now classify HSA21 genes according to their levels of expression in DS LCLs. We show that among the expressed genes, 17% are controlled by the gene dosage effect,

11% are amplified and 23% are compensated. Gene expression variations in DS are controlled by mechanisms involving gene regulation networks and perhaps allelic variations. This new hypothesis will need to be demonstrated for Class II or III genes by correlating levels of expression to allelic variants.

Finally the relevance of the LCL model is demonstrated and study of a larger cohort of DS patients should enable correlations between gene expression variations and DS phenotype.

REFERENCES

- [1] Epstein CJ, Korenberg J.R., Anneren G, Antonarakis SE, Ayme S, Courchesne E, Epstein LB, Fowler A, Groner Y, Huret JL, et al. Protocols to establish genotype-phenotype correlations in Down syndrome. *Am J Hum Genet* 1991;(49(1)):207-235.
- [2] Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. CHROMOSOME 21 AND DOWN SYNDROME: FROM GENOMICS TO PATHOPHYSIOLOGY. *Nat Rev Genet* 2004; 5(10):725-738.
- [3] LEJEUNE J., GAUTIER M., Turpin R. Etudes des chromosomes somatiques de neuf enfants mongoliens. *Comptes Rendus Academie des Sciences* 1959; 248:1721-1722.
- [4] Saran NG, Pletcher MT, Natale JE, Cheng Y, Reeves RH. Global disruption of the cerebellar transcriptome in a Down syndrome mouse model. *Hum Mol Genet* 2003; 12(16):2013-2019.
- [5] Shapiro BL. Developmental instability of the cerebellum and its relevance to Down syndrome. *J Neural Transm Suppl* 2001; (61):11-34.
- [6] Dahmane N, Ghezala GA, Gosset P, Chamoun Z, Dufresne-Zacharia MC, Lopes C, Rabatel N, Gassanova-Maugenre S, Chettouh Z, Abramowski V. Transcriptional Map of the 2.5-Mb CBR-ERG Region of Chromosome 21 Involved in Down Syndrome. *Genomics* 1998; 48(1):12-23.
- [7] Delabar J, Theophile D, Rahmani Z, Chettouh Z, Blouin J, Prieur M, Noel B, Sinet P. Molecular mapping of twenty-four features of Down syndrome on chromosome 21. *Eur J Hum Genet* 1993;(1(2)):114-124.
- [8] Chung IH, Lee SH, Lee KW, Park SH, Cha KY, Kim YS, Lee S. Gene Expression Analysis of Cultured Amniotic Fluid Cell with Down Syndrome by DNA Microarray. *J Korean Med Sci* 2005; 20(1):82-87.
- [9] FitzPatrick DR, Ramsay J, McGill NI, Shade M, Carothers AD, Hastie ND. Transcriptome analysis of human autosomal trisomy. *Hum Mol Genet* 2002; 11(26):3249-3256.
- [10] Giannone S, Strippoli P, Vitale L, Casadei R, Canaider S, Lenzi L, D'Addabbo P, Frabetti F, Facchin F, Farina A, Carinci P, Zannotti M. Gene Expression Profile Analysis in Human T Lymphocytes from Patients with Down Syndrome. *Annals of Human Genetics* 2004; 68(6):546-554.
- [11] Gross SJ, Ferreira JC, Morrow B, Dar P, Funke B, Khabele D, Merkatz I. Gene expression profile of trisomy 21 placentas: A potential approach for designing noninvasive techniques of prenatal diagnosis. *American Journal of Obstetrics and Gynecology* 2002; 187(2):457-462.

- [12] Li CM, Guo M, Salas M, Schupf N, Silverman W, Zigman W, Husain S, Warburton D, Thaker H, Tycko B. Cell type-specific over-expression of chromosome 21 genes in fibroblasts and fetal hearts with trisomy 21. *BMC Medical Genetics* 2006; 7(1):24.
- [13] Mao R, Zielke CL, Ronald Zielke H, Pevsner J. Global up-regulation of chromosome 21 gene expression in the developing down syndrome brain. *Genomics* 2003; 81(5):457-467.
- [14] Mao R, Wang X, Spitznagel E, Frelin L, Ting J, Ding H, Kim Jw, Ruczinski I, Downey T, Pevsner J. Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. *Genome Biology* 2005; 6(13):R107.
- [15] Tang Y, Schapiro MB, Franz DN, Patterson BJ, Hickey FJ, Schorry EK, Hopkin RJ, Wylie M, Narayan T, Glauser TA, Gilbert DL, Hershey AD, Sharp FR. Blood expression profiles for tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome. *Ann Neurol* 2004; 56(6):808-814.
- [16] Amano K, Sago H, Uchikawa C, Suzuki T, Kotliarova SE, Nukina N, Epstein CJ, Yamakawa K. Dosage-dependent over-expression of genes in the trisomic region of Ts1Cje mouse model for Down syndrome. *Hum Mol Genet* 2004; 13(13):1333-1340.
- [17] Dauphinot L, Lyle R, Rivals I, Dang MT, Moldrich RX, Golfier G, Ettwiller L, Toyama K, Rossier J, Personnaz L, Antonarakis SE, Epstein CJ, Sinet PM, Potier MC. The cerebellar transcriptome during postnatal development of the Ts1Cje mouse, a segmental trisomy model for Down syndrome. *Hum Mol Genet* 2005; 14(3):373-384.
- [18] Kahlem P, Sultan M, Herwig R, Steinfath M, Balzereit D, Eppens B, Saran NG, Pletcher MT, South ST, Stetten G, Lehrach H, Reeves RH, Yaspo ML. Transcript Level Alterations Reflect Gene Dosage Effects Across Multiple Tissues in a Mouse Model of Down Syndrome. *Genome Res* 2004; 14(7):1258-1267.
- [19] Lyle R, Gehrig C, Neergaard-Henrichsen C, Deutsch S, Antonarakis SE. Gene Expression From the Aneuploid Chromosome in a Trisomy Mouse Model of Down Syndrome. *Genome Res* 2004; 14(7):1268-1274.
- [20] Aubert J, Bar-Hen A, Daudin JJ, Robin S. Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinformatics* 2004; 5(1):125.
- [21] Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, Soeda E, Ohki M, Takagi T, Sakaki Y, Taudien S, Blechschmidt K, Polley A, Menzel U, Delabar J, Kumpf K, Lehmann R, Patterson D, Reichwald K, Rump A, Schillhabel M, Schudy A, Zimmermann W, Rosenthal A, Kudoh J, Shibuya K, Kawasaki K, Asakawa S, Shintani A, Sasaki T, Nagamine K, Mitsuyama S, Antonarakis SE, Minoshima S, Shimizu N, Nordsiek G, Hornischer K, Brandt P, Scharfe M, Schon O, Desario A, Reichelt J, Kauer G, Blocker H, Ramser J, Beck A, Klages S, Hennig S, Riesselmann L, Dagand E, Haaf T, Wehrmeyer S, Borzym K,

- Gardiner K, Nizetic D, Francis F, Lehrach H, Reinhardt R, Yaspo ML. The DNA sequence of human chromosome 21. *Nature* 2000; 405(6784):311-319.
- [22] Gardiner K, Fortna A, Bechtel L, Davisson MT. Mouse models of Down syndrome: how useful can they be? Comparison of the gene content of human chromosome 21 with orthologous mouse genomic regions. *Gene* 2003; 318:137-147.
- [23] Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR. Large-Scale Transcriptional Activity in Chromosomes 21 and 22. *Science* 2002; 296(5569):916-919.
- [24] Reymond A, Camargo AA, Deutsch S, Stevenson BJ, Parmigiani RB, Ucla C, Bettoni F, Rossier C, Lyle R, Guipponi M. Nineteen Additional Unpredicted Transcripts from Human Chromosome 21. *Genomics* 2002; 79(6):824-832.
- [25] Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. Assessing Gene Significance from cDNA Microarray Expression Data via Mixed Models. *Journal of Computational Biology* 2001; 8(6):625-637.
- [26] SAS Institute Inc. *SAS/STAT Software Version 8*. SAS Institute Inc., Cary, NC., 2000.
- [27] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JRSSB* 1995;(57(1)):289-300.
- [28] Janel N, Christophe O, Ait Yahya-Graison E, Hamelet J, Paly E, Prieur M, Delezoide AL, Delabar JM. Paraoxonase-1 expression is up-regulated in Down syndrome fetal liver. *Biochem Biophys Res Commun* 2006; 346(4):1303-1306.
- [29] Potier MC, Rivals I, Mercier G, Ettwiller L, Moldrich RX, Laffaire J, Personnaz L, Rossier J, Dauphinot L. Transcriptional disruptions in Down syndrome: a case study in the Ts1Cje mouse cerebellum during post-natal development. *J Neurochem* 2006; 97 Suppl 1:104-109.
- [30] Pellegrini-Calace M, Tramontano A. Identification of a novel putative mitogen-activated kinase cascade on human chromosome 21 by computational approaches. *Bioinformatics* 2006; 22(7):775-778.
- [31] Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyam L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel CV, Antonarakis SE. Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum Mol Genet* 2005; 14(23):3741-3749.
- [32] Sherman L, Levanon D, Lieman-Hurwitz J, Dafni N, Groner Y. Human Cu/Zn superoxide dismutase gene: molecular characterization of its two mRNA species. *Nucleic Acids Res* 1984; 12(24):9349-9365.
- [33] Brigand KL, Russell R, Moreilhon C, Rouillard JM, Jost B, Amiot F, Magnone V, Bole-Feysot C, Rostagno P, Virolle V, Defamie V, Dessen P, Williams G, Lyons P, Rios G,

Mari B, Gulari E, Kastner P, Gidrol X, Freeman TC, Barbry P. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. *Nucl Acids Res* 2006; 34(12):e87.

- [34] Chen H, Chrast R, Rossier C, Morris MA, Lalioti MD, Antonarakis SE. Cloning of 559 potential exons of genes of human chromosome 21 by exon trapping. *Genome Res* 1996; 6(8):747-760.
- [35] Ohira M, Seki N, Nagase T, Suzuki E, Nomura N, Ohara O, Hattori M, Sakaki Y, Eki T, Murakami Y, Saito T, Ichikawa H, Ohki M. Gene identification in 1.6-Mb region of the Down syndrome region on chromosome 21. *Genome Res* 1997; 7(1):47-58.

LEGENDS

Table 1: Comparison between HSA21 and HSA21 oligoarray content. Gene Database on NCBI was used to estimate HSA21 content. Only current sequences were considered except pseudogenes and hypothetical proteins. “nb of seq” is the number of HSA21 sequences represented at least one probe on the HSA21 oligoarray. “% of coverage” is the proportion of HSA21 sequences currently annotated in Gene Database that are represented on the microarray.

Table 2: Experimental design. List of comparisons between LCLs from DS patients and control individuals according to the mixed model used in the study. The table is divided into 4 blocks, each representing 10 experiments. “1” means that DS sample was labelled with cyanine (Cy) 5 and control sample was labelled with Cy3. “-1” means that DS sample was labelled with Cy3 and control sample was labelled with Cy5.

Table 3: List of significantly over-expressed genes (SOE) in DS. Listed HSA21 genes are described by their gene symbol and their GenBank ID. “A” corresponds to the mean of $(\log_2(\text{DS}) + \log_2(2N))/2$ for the corresponding gene across the 40 hybridizations; “M” corresponds to the mean of $\log_2(\text{DS}) - \log_2(2N)$ for the corresponding gene across the 40 hybridizations. “var(M)” is the variance of the M values; “DS/control ratio” is equal to 2^M ; “CLASS” categorize the gene according to its expression level in DS cell lines relative to control cell lines : Class I contains genes which are expressed according to the gene dosage default, Class II contains genes with amplified overexpression, Class III contains genes under gene dosage compensation and Class IV contains genes that are highly variable.

Table 4: List of non differentially expressed genes (NDE) between DS and controls LCLs.

Same legend as Table 3.

Table 5: QPCR Results compared to microarray data. The ten HSA21 tested genes are identified by their name and GenBank ID. “DS/Control RATIO” of “HSA21 Oligoarray Results” part gives the mean expression ratio for the corresponding gene between DS and Control cell lines; “FDR” is adjusted p-values calculated from HSA21 oligoarray data; “STATUS” is SOE for significantly over-expressed, NDE for non differentially expressed and NS for non significant. QPCR results are summarized as “DS/Control RATIO” which the ratio of mean expression values calculated from normalized Ct obtained for DS cell lines relative to control cell lines; “p-val” is the p-value deduced from t test.

Figure 1: Classification of HSA21 genes according to their expression ration between DS and control LCLs

Figure 2: Distribution of DS/control ratios for HSA21 genes significantly over-expressed , non differentially expressed and non HSA21 reference genes. (A) histogram of the DS/Control mean expression ratios for all HSA21 genes that are significantly over-expressed (SOE) listed in table 3; (B) histogram of the DS/Control mean expression ratios for all HSA21 genes that are non differentially expressed (NDE) listed in table 4; (C) histogram of the DS/Control mean expression ratios for all non HSA21 genes used as reference on the HSA21 oligoarray.

Figure 3: PCA on HSA21 and non HSA21 gene expression. Red symbols represent DS samples and blue symbols represent control samples. Squares represent samples extracted from female individuals and diamonds represent samples extracted from male individuals.

Figure 4: Correlation between variations of HSA21 gene expression in DS and the expression level of allelic variants. Effect on reaching the DS phenotype threshold.

Figure 5: Distribution of the variances of M values for SOE, NDE and NS genes.

Figure 6: Distributions of expressed, SOE and NDE genes along HSA21. Left axis measures the proportion of expressed genes among the genes located in each 5 Mb interval represented by the large blue bars; right axis measures the proportion of SOE or NDE among the expressed genes which are represented by the purple and yellow bars respectively.

putative expressed sequences	HSA21 content (NCBI 36.2)	HSA21 oligoarray content	
		nb of seq	% of HSA21 coverage
Genes	156	145	93.59%
KRATP	49	Excluded	NA
Open Reading Frames (orf)	45	36 (+22 sequences not precisely located on HSA21)	80.00%
Predictions	Not represented	118 (among which 20 are represented with thier reverse sequences)	NA

Table 1

		DS									
		Male						Female			
Controls		4	5	3	6	7	1	2	8	10	9
		Male	12	1	-1						1
	14		1						-1	1	-1
	13			1	-1			-1	1	-1	1
	11	-1			1		-1	1		1	-1
Female	17		1			-1			1	-1	
	16		-1								1
	18			-1	1				-1		1
	21				-1	1				1	-1
	19						1				-1
	20							-1			1
	15							1	-1		

Table 2

Gene SYMBOL	GenBank ID	HSA21 oligoarray results				CLASS
		DS/Control RATIO	A	M	var(M)	
SAMSN1	AF222927.1	2.27	10.47	1.18	0.72	II
SNF1LK	AB047786.1	2.14	9.47	1.10	1.15	II
STCH	U04735.1	1.97	9.86	0.98	0.42	II
BTG3	D64110.1	1.81	10.63	0.86	0.21	II
TTC3	D84296.1	1.79	10.59	0.84	0.20	II
C21orf57	AY040875.1	1.74	9.49	0.80	0.28	II
SLC37A1	AF311320.1	1.72	9.33	0.78	0.32	II
PDXK	AY303972.1	1.71	10.06	0.77	0.23	II
IFNAR2	BC013156.1	1.67	8.38	0.74	0.14	II
IL10RB	BT009777.1	1.66	10.18	0.73	0.21	II
MRPS6	AB049942.1	1.64	10.26	0.72	0.13	II
CCT8	BC095470.1	1.64	12.64	0.72	0.32	II
CBS	AF042836.1	1.62	8.13	0.69	0.48	II
ITGB2	BC005861.2	1.61	11.34	0.68	0.40	II
PIGP	AF216305.1	1.60	8.14	0.68	0.25	II
C21orf91	BC015468.2	1.59	9.50	0.67	0.26	I
USP25	AF170562.1	1.58	9.93	0.66	0.10	I
SLC5A3	L38500.2	1.57	8.92	0.66	0.18	I
as-TTC3	BF979681.2	1.56	9.07	0.64	0.64	I
C21orf33	BC003587.1	1.52	10.63	0.60	0.20	I
PTTG1IP	NM_004339.2	1.52	9.89	0.60	0.20	I
CRYZL1	BC033023.2	1.52	9.45	0.60	0.12	I
SON	AY026895.1	1.51	12.64	0.60	0.14	I
ZNF294	NM_015565.1	1.51	8.77	0.60	0.11	I
C21orf66	AY033903.1	1.51	9.50	0.59	0.12	I
PFKL	X15573.1	1.50	12.44	0.58	0.23	I
POFUT2	NM_015227.3	1.48	7.80	0.57	0.24	I
IFNAR1	AY654286.1	1.47	8.10	0.56	0.18	I
MRPL39	AF109357.1	1.47	9.82	0.55	0.15	I
USP16	AY333928.1	1.46	9.94	0.54	0.10	I
INFR2	AY644470.1	1.45	10.35	0.54	0.24	I
MCM3AP	AY590469.1	1.45	11.69	0.54	0.19	I
C21orf7	AY171599.2	1.43	8.00	0.52	0.49	I
DONSON	AF232673.1	1.42	9.73	0.50	0.11	I
SUMO3	BC008420.1	1.41	10.82	0.50	0.10	I
DYRK1A	D86550.1	1.41	10.29	0.49	0.15	I
GABPA	BC035031.2	1.40	8.99	0.49	0.07	I
DSCR3	D87343.1	1.40	10.09	0.48	0.12	I
SFRS15	AF023142.1	1.39	10.42	0.47	0.14	III
TMEM50B	AF045606.2	1.38	10.07	0.46	0.20	III
HMGNI	M21339.1	1.38	11.82	0.46	0.10	III
C21orf59	AF282851.1	1.35	9.61	0.44	0.21	III
ATP5J	BC001178.1	1.35	7.99	0.44	0.11	III
MORC3	BC094779.1	1.34	8.51	0.42	0.15	III
PWP2H	U56085.1	1.34	9.43	0.42	0.11	III
PRMT2	U80213.1	1.33	8.67	0.42	0.14	III
D21S2056E	U79775.1	1.33	10.55	0.41	0.18	III
HLCS	AB063285.1	1.32	8.48	0.40	0.18	III
C21orf2	NM_004928.1	1.31	8.80	0.38	0.14	III
C21orf108 (exon 39)	AF231919.1	1.30	7.84	0.38	0.18	III
CHAF1B	U20980.1	1.29	8.35	0.37	0.13	III
SYNJ1	AF009039.1	1.29	8.07	0.37	0.13	III
C21orf49	BC117399.1	1.29	7.36	0.36	0.37	III
U2AF1	M96982.1	1.27	12.15	0.35	0.06	III
C21orf51	AY081144.1	1.26	9.47	0.33	0.09	III
DSCR2	AY463963.1	1.26	11.14	0.33	0.12	III
C21orf25	AB047784.1	1.25	8.24	0.32	0.17	III
C21orf45	AF387845.1	1.25	9.69	0.32	0.12	III

Table 3

Gene SYMBOL	GenBank ID	HSA21 oligoarray results				CLASS
		DS/Control RATIO	A	M	var(M)	
C21orf6	BC017912.1	1.22	9.75	0.29	0.08	III
KIAA0179	D80001.1	1.22	10.15	0.28	0.08	III
C21orf21	AA969880	1.16	7.33	0.22	0.05	III
C21orf49	BC117399.1	1.14	7.80	0.19	0.03	III
C21orf42	AY035383.1	1.13	9.65	0.18	0.03	III
C21orf108 (exon 26)	AF231919.1	1.09	8.86	0.13	0.02	III
DSCR9	BC066653.1	1.07	7.93	0.10	0.01	III
C21orf58	BC028934.1	1.07	7.99	0.09	0.01	III
as-C21orf56	BC084577.1	1.05	12.79	0.08	0.01	III
PKNOX1	AY196965.1	1.03	7.98	0.04	0.00	III
DSCR10	AB066291.1	0.95	8.60	-0.07	0.00	III
C21orf34	AF486622.1	0.95	8.15	-0.07	0.00	III
DCR1-20.0-reverse	AJ001893.1	0.95	7.14	-0.07	0.00	III
CXADR	AF200465.1	0.90	8.16	-0.15	0.02	III
AIRE	AB006682.1	0.82	7.28	-0.28	0.08	III
BRWD1	AB080586.1	1.42	9.31	0.51	0.26	IV
PRDM15	AY078498.2	1.39	10.04	0.47	0.22	IV
DSCR4	DQ179113.1	1.32	7.41	0.40	0.16	IV
ADARB1	AY135659.1	1.30	7.21	0.38	0.14	IV
DCR1-12.0	AJ001868.1	1.29	7.37	0.36	0.13	IV
DCR1-15.0	AJ001872.1	1.28	7.08	0.36	0.13	IV
DCR1-13.0	AJ001869.1	1.25	9.20	0.32	0.11	IV
ABCG1	AY048757.1	1.25	8.11	0.32	0.10	IV
DSCAM_Intronic_Model	BG221591.1	1.22	8.39	0.28	0.08	IV
DCR1-13.0-reverse	AJ001869.1	1.14	8.98	0.19	0.04	IV
H2BFS	AB041017.1	1.13	13.19	0.17	0.03	IV
KCNE1	BC046224.1	1.12	7.18	0.16	0.03	IV
DCR1-17.0	AJ001875.1	1.12	7.16	0.17	0.03	IV
SH3BGR	X93498.1	1.12	7.39	0.16	0.03	IV
W90635	W90635.1	1.09	7.39	0.12	0.01	IV
C21orf8	AA843704.1	1.09	7.55	0.12	0.01	IV
AL041783	AL041783.1	1.06	7.10	0.09	0.01	IV
B184	AL109967.2	1.06	9.10	0.08	0.01	IV
C21orf54	AA934973.1	1.01	7.26	0.01	0.00	IV
JAM2	AY016009.1	1.01	8.10	0.01	0.00	IV
PRED24	AP001695.1	1.00	7.41	0.00	0.00	IV
as-HLCS	AB063285.1	0.98	7.58	-0.03	0.00	IV
C21orf12	AP001705.1	0.96	7.49	-0.05	0.00	IV
PRED21	AP001693.1	0.95	7.75	-0.07	0.00	IV
DSCR1	AY325903.1	0.93	7.13	-0.10	0.01	IV
PRED41	AP001726.1	0.93	7.15	-0.11	0.01	IV
RUNX1	D43968.1	0.80	7.37	-0.32	0.10	IV
CLIC6	AF448438.1	0.73	9.43	-0.45	0.20	IV

Table 4

Gene	GenBank ID	HSA21 Oligoarray Results			QPCR Results	
		DS/Control RATIO	FDR	STATUS	DS/Control RATIO	p-val
SNF1LK	AB047786.1	2.15	0.0263	SOE	3.36	0.0007
STCH	U04735.1	1.97	0.0004	SOE	2.06	0.0016
MX1	AF135187.1	1.50	0.1315	NS	1.78	0.0114
DYRK1A	D86550.1	1.41	0.0057	SOE	1.77	0.0055
CSTB	AF208234.1	1.35	0.1244	NS	1.49	0.1347
CHAF1B	U20980.1	1.30	0.0082	SOE	1.38	0.0876
TMEM1	BC101728.1	1.27	0.1006	NS	1.27	0.1297
SOD1	AY835629.1	1.16	0.1785	NS	1.96	0.0017
H2BFS	AB041017.1	1.13	0.9872	NDE	1.05	0.4438
DSCR1	AY325903.1	0.92	0.8158	NDE	1.11	0.3011

Table 5

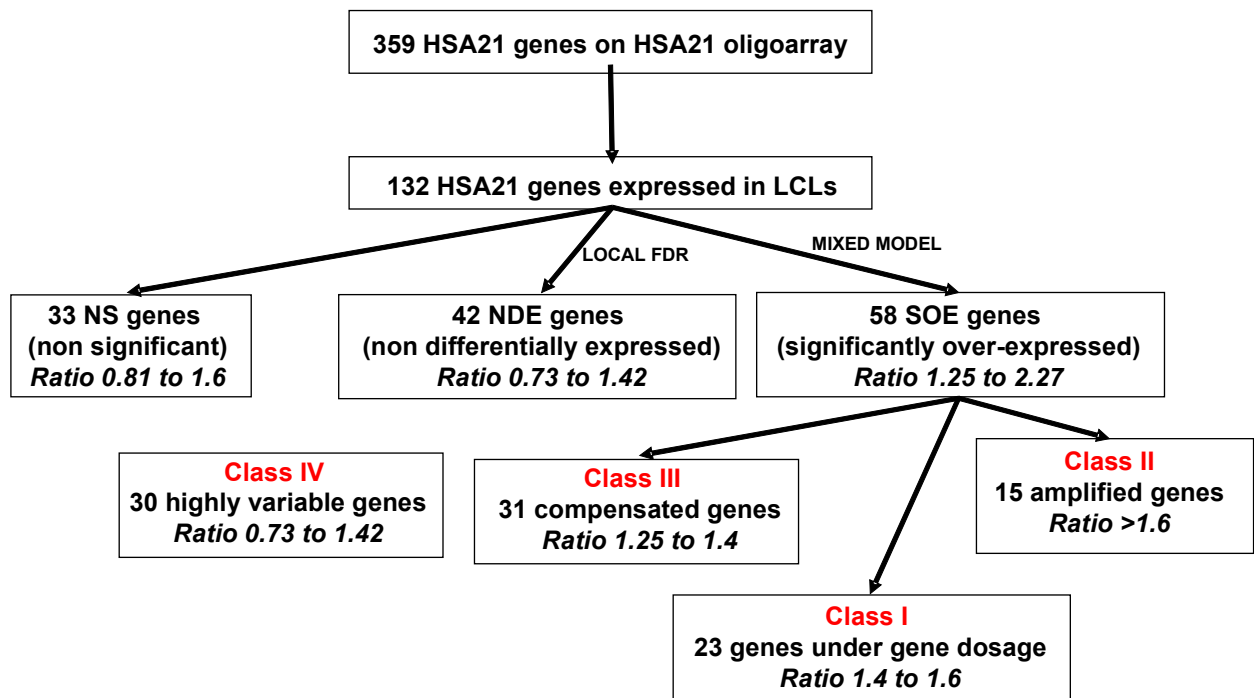


Figure 1

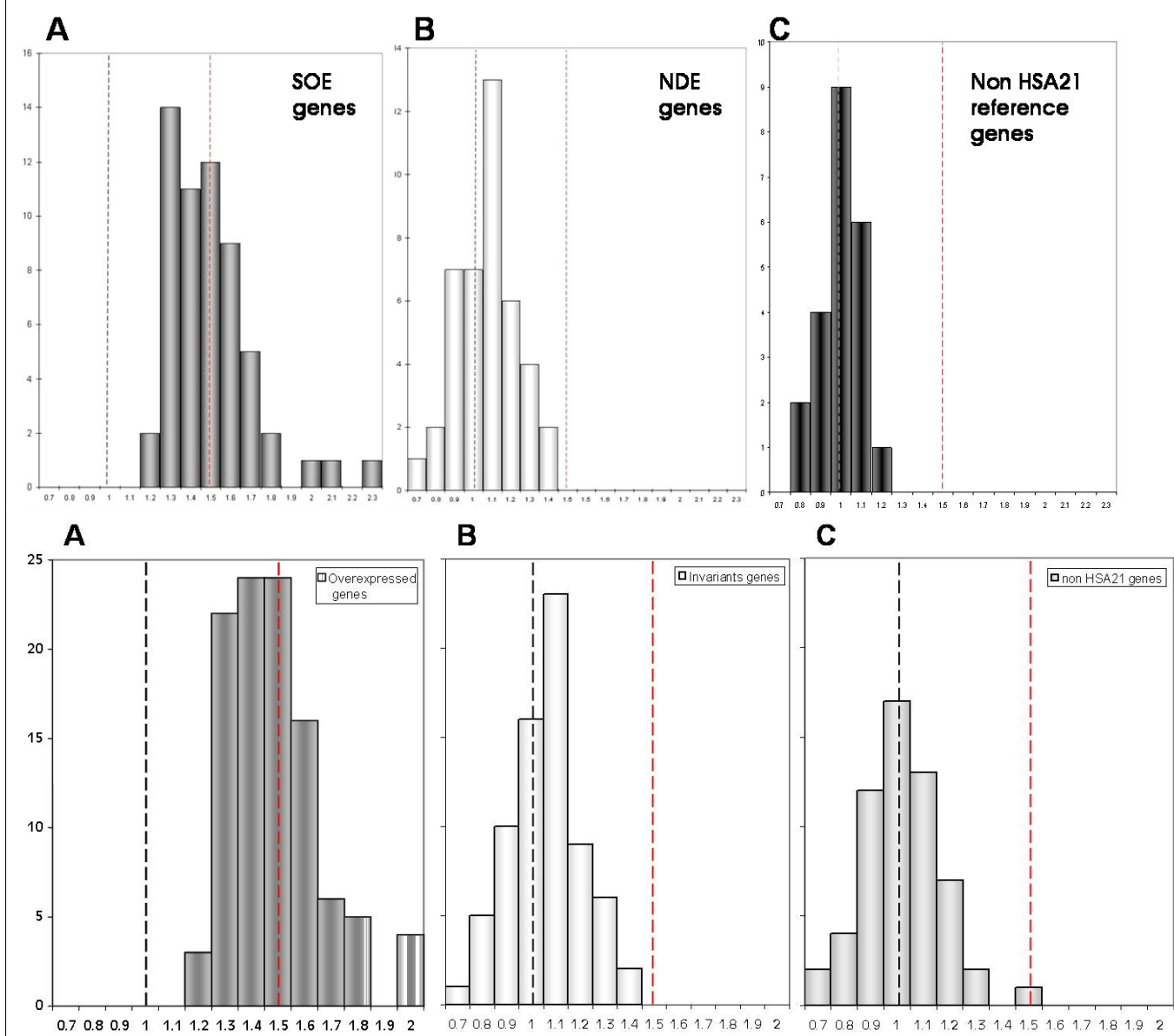


Figure 2

Figure 3

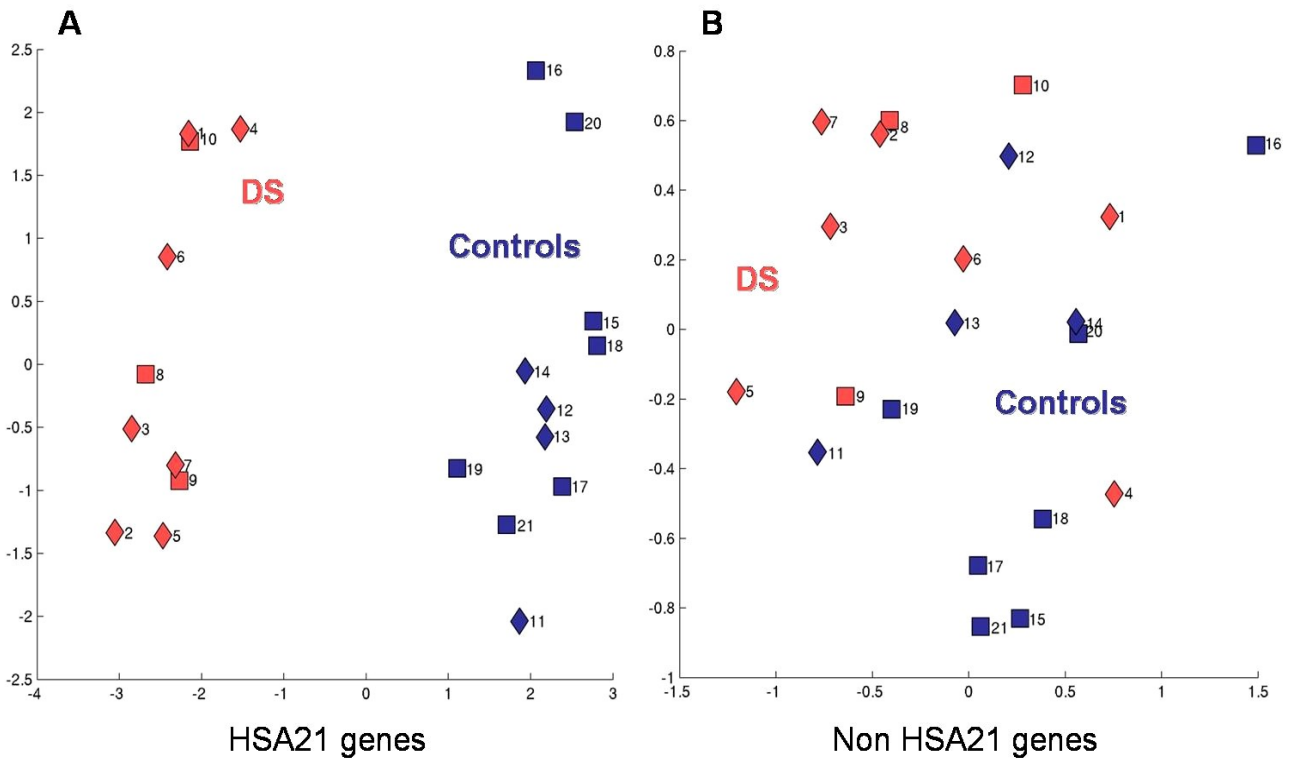
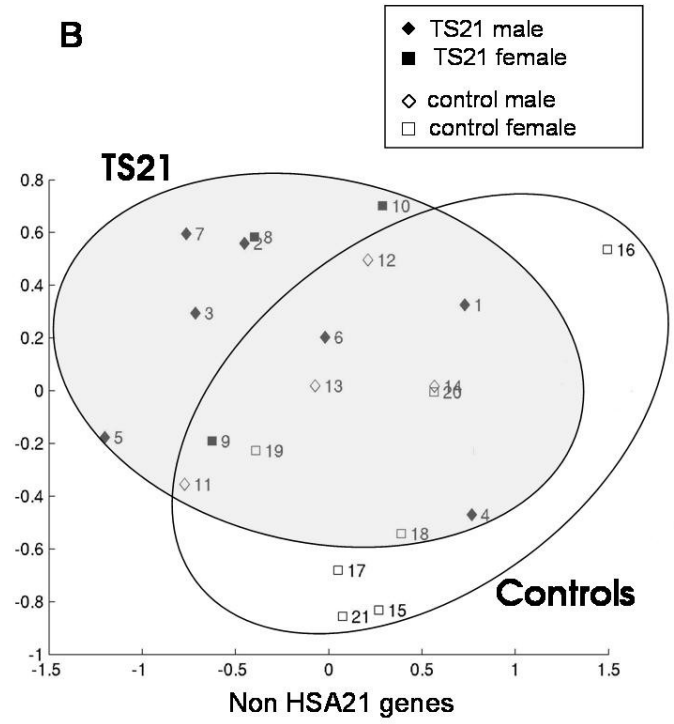
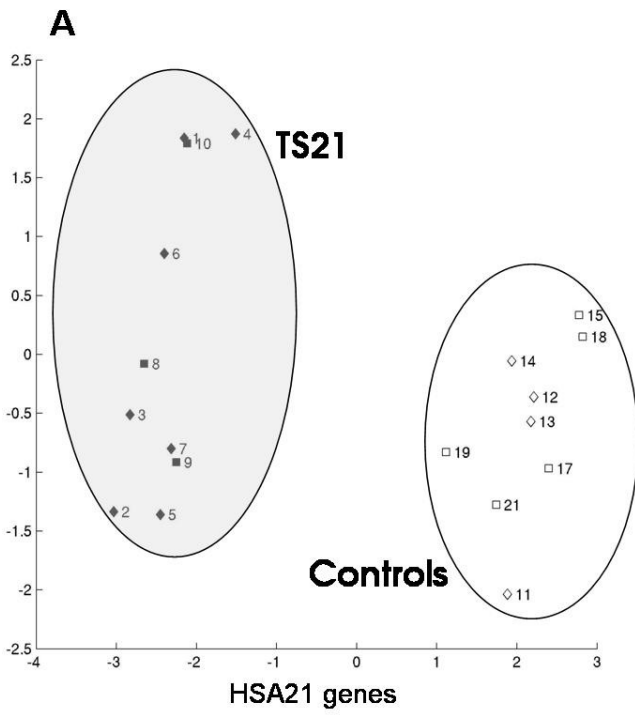


Figure 3



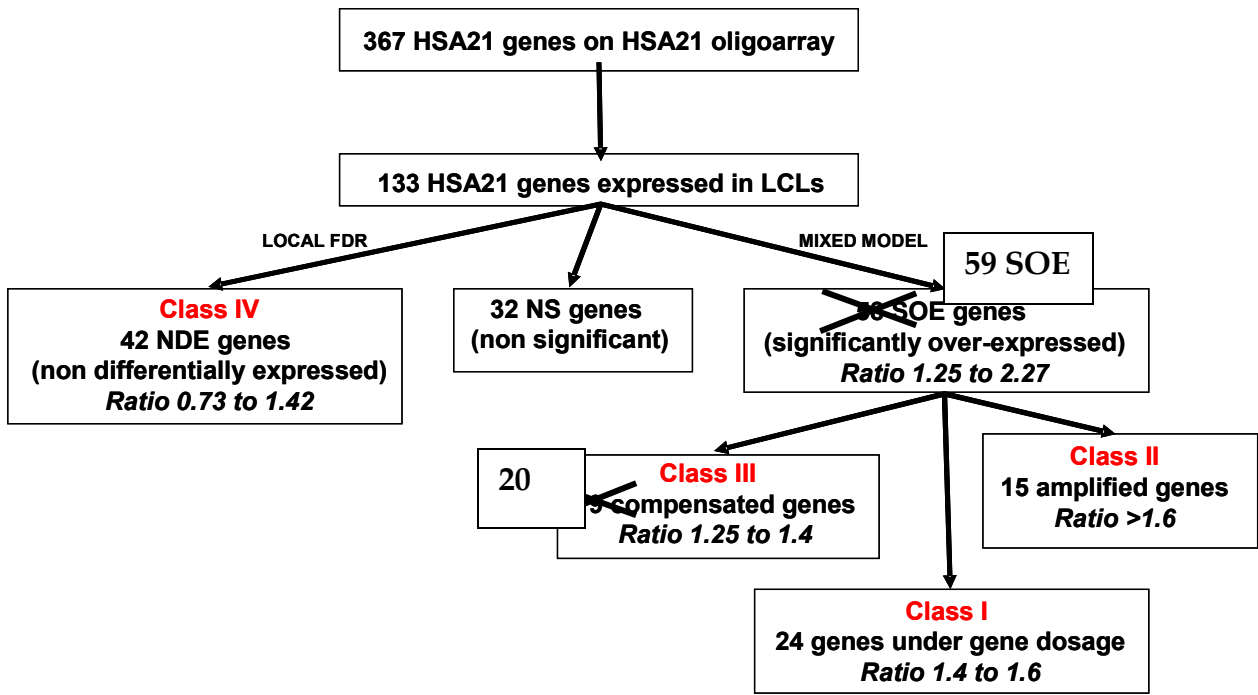


Figure 1-

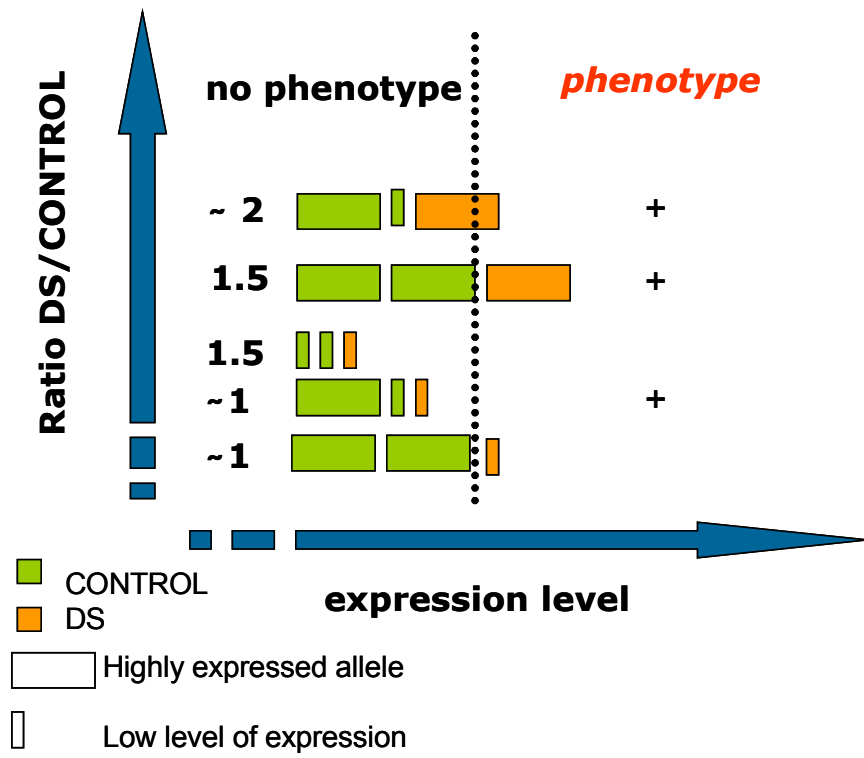


Figure 4

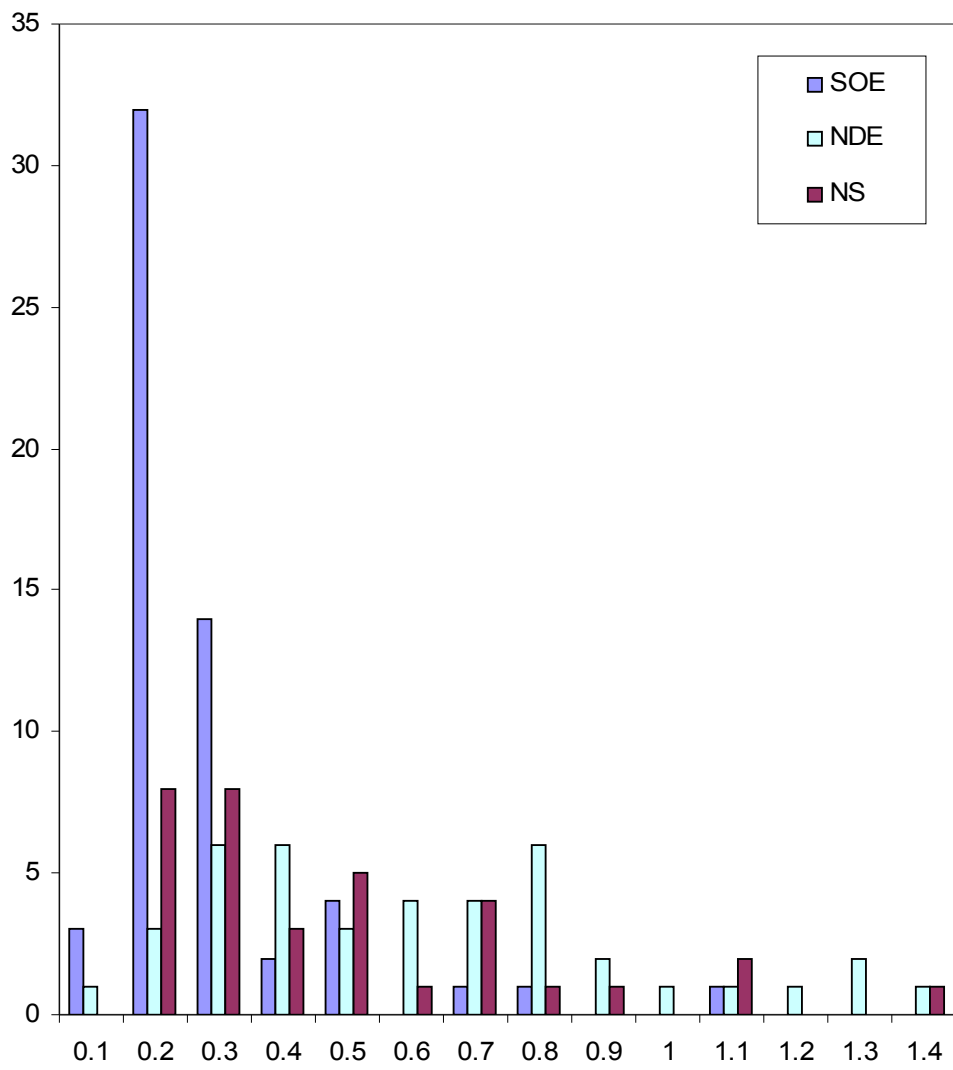


Fig.5

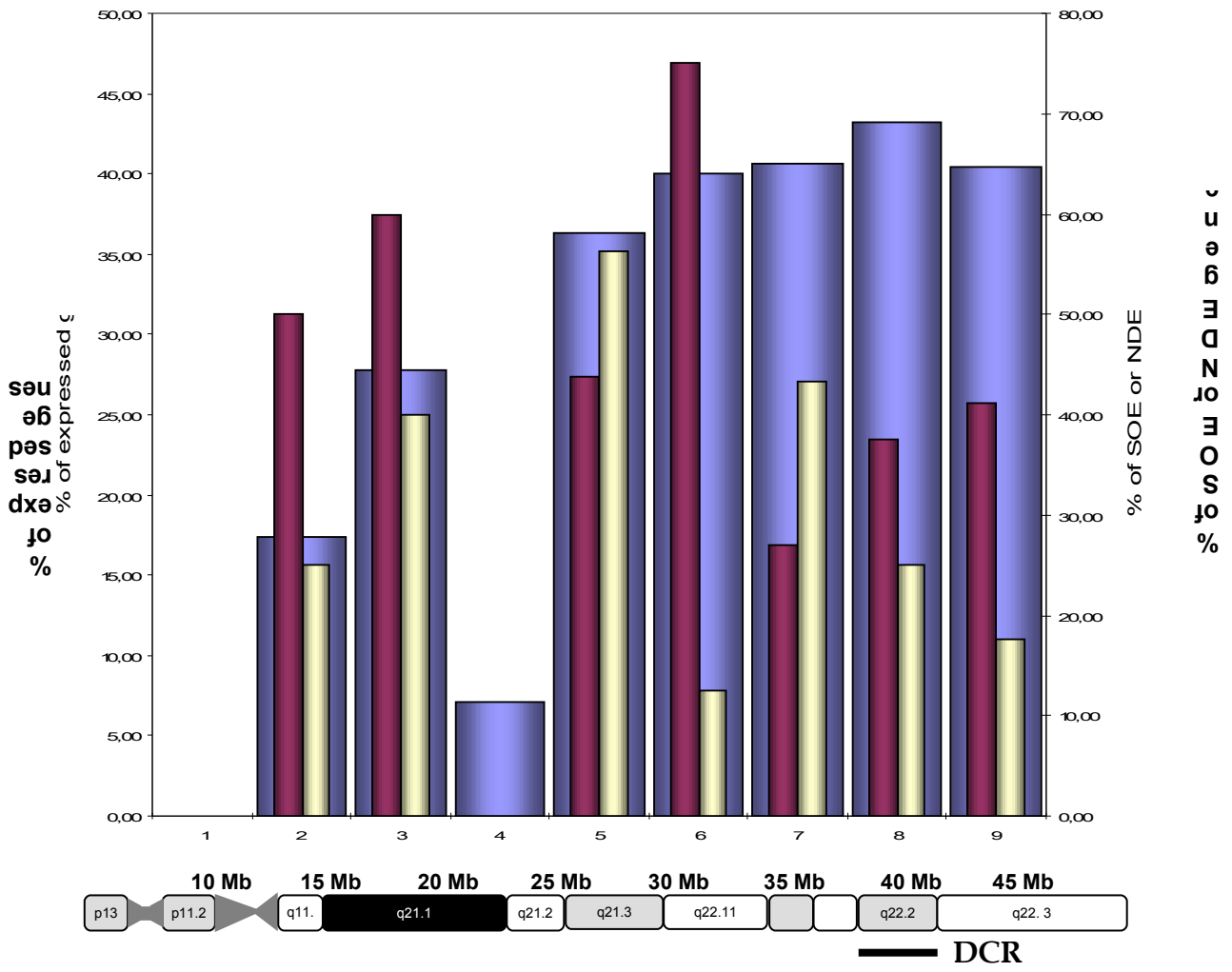


Figure 6