Clinical Chemistry 57:8
1188–1195 (2011)



# Accurate Single-Nucleotide Polymorphism Allele Assignment in Trisomic or Duplicated Regions by Using a Single Base–Extension Assay with MALDI-TOF Mass Spectrometry

Anne L. Trewick,[1] Julia S. El-Sayed Moustafa,[1] Adam J. de Smith,[1] Philippe Froguel,[1] Gottfried Greve,[2,3] Pål R. Njølstad,[2,3] Lachlan J.M. Coin,[4] and Alexandra I.F. Blakemore[1*]

BACKGROUND: The accurate assignment of alleles embedded within trisomic or duplicated regions is an essential prerequisite for assessing the combined effects of single-nucleotide polymorphisms (SNPs) and genomic copy number. Such an integrated analysis is challenging because heterozygotes for such a SNP may be one of 2 genotypes—AAB or ABB. Established methods for SNP genotyping, however, can have difficulty discriminating between the 2 heterozygous trisomic genotypes. We developed a method for assigning heterozygous trisomic genotypes that uses the ratio of the height of the 2 allele peaks obtained by mass spectrometry after a single-base extension assay.

METHODS: Eighteen *COL6A2* (collagen, type VI, alpha 2) SNPs were analyzed in euploid and trisomic individuals by means of a multiplexed single-base extension assay that generated allele-specific oligonucleotides of differing $M_r$ values for detection by MALDI-TOF mass spectrometry. Reference data (mean and SD) for the allele peak height ratios were determined from heterozygous euploid samples. The heterozygous trisomic genotypes were assigned by calculating the *z* score for each trisomic allele peak height ratio and by considering the sign (+/−) of the *z* score.

RESULTS: Heterozygous trisomic genotypes were assigned in 96.1% (range, 89.9%–100%) of the samples for each SNP analyzed. The genotypes obtained were reproduced in 95 (97.5%) of 97 loci retested in a second assay. Subsequently, the origin of nondisjunction was determined in 108 (82%) of 132 family trios with a Down syndrome child.

CONCLUSIONS: This approach enabled reliable genotyping of heterozygous trisomic samples and the determination of the origin of nondisjunction in Down syndrome family trios.



Down syndrome (DS)[5] is associated with a number of subphenotypes, some of which, such as mental retardation and the characteristic facial appearance, are present in all affected individuals. Other subphenotypes, such as leukemia and congenital heart defects, are present in only a proportion of DS cases. Thus, trisomy per se does not cause these phenotypes; a more complex etiology must be responsible.

Accurate genotyping of single-nucleotide polymorphisms (SNPs) within regions of variant genomic copy number, including aneuploidies such as trisomy 21 and common or rare genomic structural variants, is essential to understand how SNP alleles and genomic copy number interact to affect phenotype. Although the phenotypic implications of copy number variants (CNVs) have been studied extensively in recent years *(1)*, there has been little investigation into the combined effects of CNVs and SNPs.

In addition to its potential applications in genetic dissection of common complex traits, this approach may allow identification of loci segregating with subphenotypes within genomic or chromosomal disorders (e.g., for investigation of congenital heart defects in trisomy 21). A particular analytical challenge, however, concerns the assignment of genotypes to heterozygotes

[1] Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK; [2] Department of Clinical Medicine, University of Bergen, Bergen, Norway; [3] Department of Paediatrics, Haukeland University Hospital, Bergen, Norway; [4] Department of Epidemiology and Public Health, Imperial College London, London, UK.
* Address correspondence to this author at: Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital Campus, Du Cane Rd., London W12 0NN, UK. Fax +44-207-5946511; e-mail a.blakemore@imperial.ac.uk.

Received January 16, 2011; accepted June 1, 2011.
Previously published online at DOI: 10.1373/clinchem.2010.159558

[5] Nonstandard abbreviations: DS, Down syndrome; SNP, single-nucleotide polymorphism; CNV, copy number variant; PHR, peak height ratio; MAF, minor-allele frequency; HMW, high molecular weight; LMW, low molecular weight; $PHR_{Ref}$, mean heterozygote PHR for the euploid reference population; $SD_{Ref}$, heterozygote SD for the euploid reference population; SES, standard error of the skew; NDJ, nondisjunction; CHD, congenital heart defect.

for SNPs in duplicated regions. These individuals will have 2 copies of one of the alleles and 1 copy of the other, and be of genotype AAB or ABB. The allele-detection system must therefore have a quantitative element to be able to distinguish between the 2 genotypes, since qualitative methods will report only an AB genotype.

We present a simple postassay analysis of Sequenom® iPLEX™ data for accurately assigning genotypes in such cases. The iPLEX assay is a multiplexed single-base extension assay that generates allele-specific products for detection by MALDI-TOF mass spectrometry. Although originally designed for SNP analysis, the assay has also been used for detecting trisomy *(2–4)*, as well as high-throughput validation of CNVs *(5, 6)*. Therefore, the iPLEX assay has potential for integrated high-throughput joint analyses of SNPs and CNVs. SNP array platforms may also be used to generate signal-intensity information, thereby permitting assignment of trisomic genotypes via CNV-prediction algorithms *(7)*. Genotyping with SNP array platforms can be prohibitively expensive, however, compared with the Sequenom technology, which allows cost-effective targeted genotyping of selected SNPs.

The iPLEX genotype-calling algorithm assigns genotypes by calculating the relative yield of each allele peak. Genotype assignment with this algorithm, however, is inaccurate for SNPs embedded within regions of increased copy number. On initial review of the assay data for trisomic heterozygotes (Fig. 1), it became clear to us that the software could not assign genotypes correctly to these samples because they gave unusual relative-yield results. Trisomic samples are either assigned a euploid genotype or reported as a "no-call." We therefore developed an alternative method of data interpretation for assigning genotype that uses the ratio of allele peak heights—the peak height ratio (PHR). As an example, we present data for SNPs within the *COL6A2* (collagen, type VI, alpha 2) gene in individuals with trisomy 21.

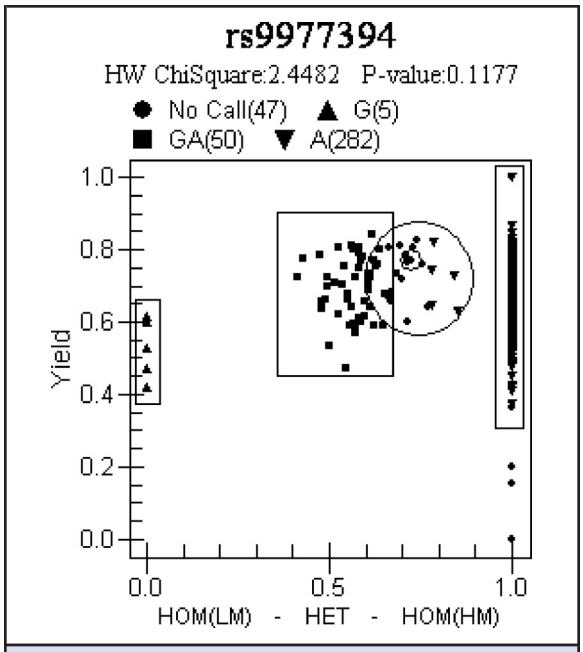


**Fig. 1. Sequenom cluster plot.**

The data points in boxes show results for euploid samples; the 3 genotypic states are clearly distinguishable. The data points in the circle are mostly trisomic heterozygotes that were assigned inappropriate genotypes or for which no genotype was called. This example illustrates the difficulty the built-in Sequenom software has in interpreting nondisomic data. HOM(LM), LMW homozygote; HET, heterozygote; HOM(HM), HMW homozygote. ©Figure reproduced with permission from Sequenom, Inc.

## Materials and Methods

### SAMPLES

Genomic DNA was extracted from peripheral-blood samples that had been collected from family "trios" (mother, father, and affected child) for research into congenital heart defects at the Brompton Hospital, London, UK, and the Haukeland University Hospital, Bergen, Norway. Samples were available from 132 trios in which the child had free trisomy 21 (confirmed by karyotyping) and 33 trios in which the child was euploid. An additional 20 samples from parents of trisomic children and 55 samples from apparently healthy adult controls were also available. A total of 405 adult samples were available for analysis. Research Ethics Committee approval was obtained for the study, and written informed consent was obtained from each participant.

### SNP SELECTION AND GENOTYPING

We used the Tagger Pairwise feature within the International HapMap Project Website to select *COL6A2* SNPs for analysis (http://hapmap.ncbi.nlm.nih.gov/). An $r^2$ cutoff of 0.8 and a minor-allele frequency (MAF) of 0.05 were used to select SNPs. For analysis, we selected 21 tag SNPs, which covered the entire coding region plus 10 kb upstream and downstream. The iPLEX Gold MALDI-TOF mass spectrometry system (Sequenom) was used to genotype the samples. We designed PCR and extension primers by using the "moderate multiplex" setting in the Assay Design feature of the Sequenom MassARRAY software. The details for all primers are given in Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue8. The multiplexed PCR reactions were carried out with 2 $\mu$L

DNA solution, which contained 5.0–10.0 ng DNA per multiplexed reaction. The PCR primers were 19 or 20 bp in length and generated amplicons between 80 and 120 bp in length. The extension reactions yielded a nested set of allele-specific products of increasing $M_r$ values from 5300 to 7800. Primers were synthesized by Sigma-Aldrich. The PCR and primer-extension reactions were carried out in 384-well polypropylene plates.

Mass spectrometry analysis was carried out with an Autoflex Mass Spectrometer (Bruker Daltonics) on a 384-element bioarray. Fifteen nanoliters of extension-reaction product was dispensed onto the bioarray. The MassARRAY Workstation software (version 3.3; Sequenom) was used to process and analyze the bioarrays. Assay quality metrics were calculated for the parental samples only, because these samples were known to be euploid. Data were rejected if the genotyping success rate was <90%, the yield was <0.5, the MAF was <0.05, or the loci were not in Hardy–Weinberg equilibrium.

#### GENOTYPING TRISOMIC HETEROZYGOTES

Further analysis of the Sequenom data was carried out by importing raw data files into Microsoft Excel 2000, and an alternative method of analyzing the data was developed to improve the success rate for genotyping trisomic heterozygotes.

The ratio of the 2 allele peaks in each heterozygous sample (the PHR, Eq. 1) is calculated as follows:

$$\text{PHR} = \frac{\text{Height of high molecular weight (HMW) allele}}{\text{Height of low molecular weight (LMW) allele}} \quad (1)$$

The PHR yielded a measure of the relative quantities of each allele product independently of assay recovery and responded linearly to an increase in the height of either peak (Fig. 2). Thus, the PHR method has substantial benefits over the relative-yield method because it is more sensitive to differences in peak height and allows better discrimination between trisomic heterozygous genotypes. Ideally, euploid samples should produce a PHR result close to 1.0, whereas trisomic samples should yield a PHR close to either 2.0 or 0.5, depending on the genotype (i.e., AAB or ABB).

By way of convention to standardize the procedure, we calculated the PHR as the HMW allele divided by the LMW allele. Fig. 3 shows the distribution of PHRs obtained for one of the analyzed *COL6A2* SNPs (rs7279347). The mean PHR for the parental samples for this SNP (Fig. 3A) was 0.99, although that was not the case for every SNP (see Table 2 in the online Data Supplement). The PHRs obtained for the euploid children and those obtained for trisomic children also overlapped to a degree (Fig. 3B). Therefore, to interpret a trisomic heterozygote PHR, we adopted the following approach:

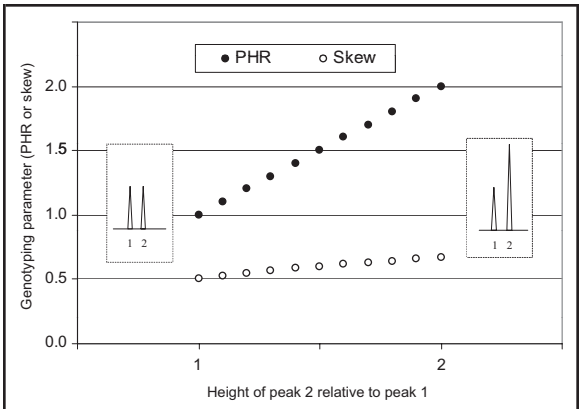


**Fig. 2. Genotyping using the Sequenom iPLEX assay and the skew value, compared with using the PHR.**

As the height of allele peak 2 increases relative to allele peak 1 (expected in a heterozygous trisomic sample), the PHR produces a linear response, whereas the skew method used by the iPLEX system has a nonlinear response.

- The range of PHRs expected in a euploid reference population was determined, and the mean heterozygote PHR ($PHR_{Ref}$) and the SD ($SD_{Ref}$) for the reference population were calculated for each SNP for samples in the reference group. We used parametric statistics to analyze the PHR data, because we assumed that any analytical error would be randomly distributed. We tested this assumption by calculating the skew coefficient for each SNP. The data were considered normally distributed (gaussian) if the skew coefficient was <2 times the standard error of the skew (SES).
- The expected euploid PHR interval (reference interval) was determined by calculating the mean PHR ± 2 SDs. This interval would yield a measure of the total analytical error expected under normal circumstances (when the 2 allele peaks are expected to be of the same height).
- To standardize and simplify the comparison of different SNPs when the $PHR_{Ref}$ does not always equal 1, we calculated the *z* score (Eq. 2):

$$z = (PHR_{child} - PHR_{Ref})/SD_{Ref} \quad (2)$$

where $PHR_{child}$ is the PHR of the child in the trio. This procedure also allowed us to gain a measure of uncertainty as to the trisomic genotypes, particularly those

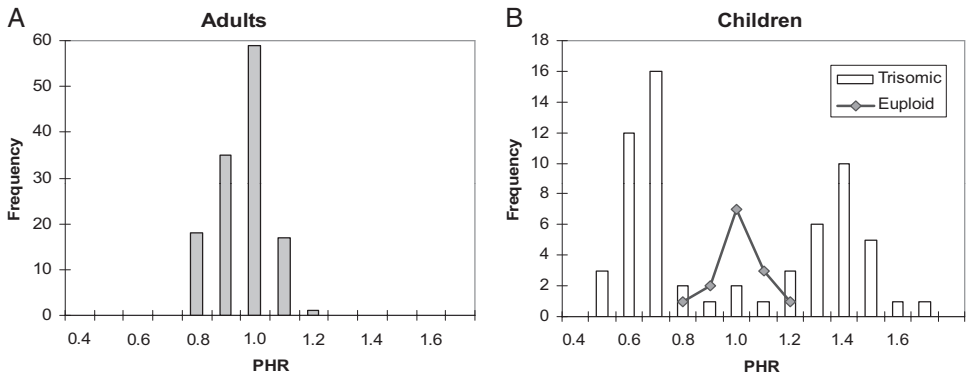


**Fig. 3. Histograms showing the distribution of heterozygote PHRs for the *COL6A2* SNP rs7279347.**

The trisomic genotype was determined by establishing a reference interval of PHRs from the parental samples (A). The PHR for the trisomic samples produced a bimodal distribution because of the co-occurrence of 2 distinct genotypes (AAB and ABB). Because some trisomic PHRs and euploid PHRs showed a region of overlap (B), trisomic genotypes in this region could be assigned only at a degree of certainty that was lower than that outside of the euploid reference interval.

falling within the region of overlap. When the trisomic PHR fell within $\pm 0.5$ SDs of the $PHR_{Ref}$ we considered the genotype not determinable, owing to the extent of overlap with the euploid samples. Conversely, when the trisomic PHR fell outside of the reference interval (i.e., $PHR_{Ref} \pm 2$ SDs), the trisomic genotype could be assigned with the highest degree of certainty.

#### ASSIGNMENT OF THE ORIGIN OF NONDISJUNCTION

After determination of the genotypes of the trisomic children and their parents, there are 2 steps involved in determining the origin of nondisjunction (NDJ) that must be considered in sequence: (*a*) identifying the nondisjoining parent, and (*b*) identifying the meiotic division (meiosis I or II) in which the NDJ occurred.

Determining the nondisjoining parent requires the presence of a SNP in which one of the parents is homozygous for one allele and the other parent is homozygous for a different allele (i.e., AA and BB). The trisomic offspring will therefore be AAB or ABB, thereby indicating which parent was the source of the extra chromosome.

Determining whether the NDJ occurred in the first or second meiotic division requires the presence of a different SNP in which the NDJ parent is heterozygous (e.g., AB) and the correctly disjoining parent is homozygous (e.g., BB). Thus, the child will be either ABB for NDJ in meiosis I, or AAB or BBB for NDJ in meiosis II.

A third method of determining the origin of NDJ is possible in a few instances of meiosis II NDJ. If, for example, the parental genotypes are AB and BB and the child's genotype is AAB, then the two A alleles must have come from the heterozygous parent via a meiosis II NDJ.

#### R SCRIPT

To facilitate the genotyping of trisomic samples and determining the origin of NDJ, we developed a script, which we wrote for use in the statistical programming environment R (version 2.9.2) *(8)*, to automatically calculate *z* scores for heterozygous trisomic samples and to subsequently assign genotypes and the parent and origin of NDJ as previously described.

Where calculated *z* scores permitted, the trisomic genotype was assigned for heterozygous SNPs, with samples being assigned a genotype of "HMW trisomic" for samples with a *z* score $>0.5$ or "LMW trisomic" for samples with a *z* score $<-0.5$ for the SNP in question.

When the appropriate SNP could be identified, assignment of the nondisjoining parent and the meiotic division in which the NDJ occurred was also automated, as described above.

The R script DS_Sequenom.R is available at http://www1.imperial.ac.uk/medicine/people/a.blakemore/.

## Results

#### GENOTYPING EUPLOID SAMPLES

Genotyping reactions (see Table 1 in the online Data Supplement) were carried out on 405 adult euploid samples (parents and normal adult controls). Six of these samples failed to amplify or produce consistent results and were excluded from further analysis. The MAF obtained for each SNP was, on average, within 0.07 (range, 0.0–0.096) of that published in dbSNP on

the NCBI website *(9)*. Two SNPs (rs2236490 and rs11908960) were monomorphic in our assays despite reported MAFs >0.10 in the CEU population (Utah residents with ancestry from northern and western Europe). The call success rate for these 2 SNPs was <90%, suggesting that possible analytical problems were encountered with these assays. Another SNP (rs9982817) both showed a low genotyping success rate and was not in Hardy–Weinberg equilibrium. Data for all 3 of these SNPs were excluded from further analysis (see Table 3 in the online Data Supplement).

Calculating the PHR for the reference samples revealed a gaussian distribution of the data in the majority of cases (see Table 2 in the online Data Supplement); however, when the mean PHR was <1.0, the distribution tended not to be gaussian (i.e., a skew coefficient >2 times the SES). Transforming the data by taking the reciprocal of the PHR so that the mean PHR was >1.0 yielded a skew coefficient <2 times the SES (see Fig. 2 in the online Data Supplement). Therefore, as a general principle, we subsequently calculated the PHR so that the mean PHR was >1.0.

#### GENOTYPING TRISOMIC SAMPLES

The euploid PHR reference interval was determined for each SNP by calculating $PHR_{Ref} \pm 2\ SD_{Ref}$'s. One would expect 95% of all euploid samples to fall within these limits. To test the validity of the reference interval, we calculated the PHRs for the samples from the euploid children ($PHR_{child}$) and compared the PHRs to each reference interval. For some of the SNPs, there were relatively few heterozygous euploid children (median, 9.5; range, 1–21) who did not give all the expected results; therefore, we calculated a mean (excluding rs11554669 as an outlier), which showed that 92% of these samples fell within the reference intervals for all of the SNPs tested (see Table 4 in the online Data Supplement) and which was close to that expected theoretically.

We predicted that because the PHR for heterozygous trisomic samples is determined from allele peaks expected to be of unequal height, these PHRs would fall outside of the euploid reference intervals. Fig. 4 shows the distribution of *z* scores for 2 such SNPs. When the *z* scores for all genotypes (including heterozygotes and homozygotes) were plotted for all children, the trisomic heterozygote genotypes were clearly separated from the euploid heterozygotes, which were clustered around the *x* axis (i.e., $z = 0$). The trisomic heterozygote genotype could be assigned as either TCC or TTC by the direction of the *z* score (positive or negative) relative to the homozygous samples (CC or TT) and the euploid heterozygotes. Fig. 4A is an ideal plot that shows clear demarcation between the different genotypes. Fig. 4B shows a commonly encountered situation in which trisomic heterozygotes occur within the euploid interval. We hypothesize that these atypical PHRs may represent the occurrence of CNVs superimposed on the state of trisomy, thereby causing a normalization of the PHR.

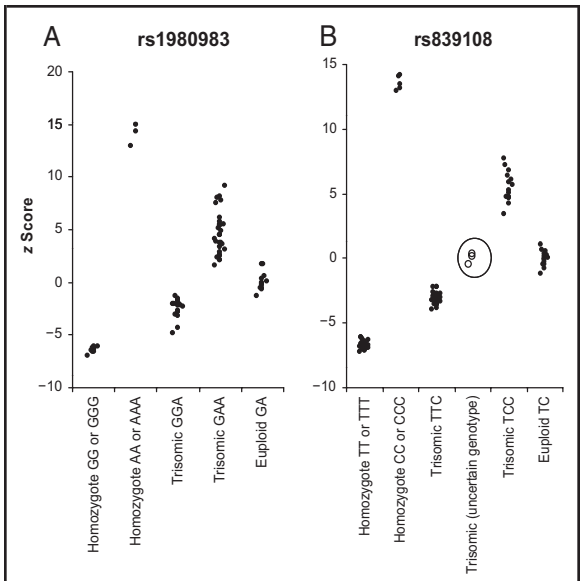


**Fig. 4. The distribution of *z* scores for rs1980983 (A) and rs839108 (B).**

*z* Scores for all of the children's samples (including homozygotes and heterozygotes) are plotted. (A), Shown is how the genotype of the heterozygous trisomic samples is determined by comparing the sign of the *z* score (+ or −) with the sign of the *z* score for the homozygous samples. (B), Shown is an additional group of samples with atypical *z* scores for trisomic samples (circled). These scores may indicate the existence of CNVs superimposed on a trisomic state. A deletion CNV, for example, would restore the local copy number to 2, whereas a duplication CNV would produce a local copy number of 4. In either case, the PHR would be 1:1, as for a euploid sample.

This method of interpreting the Sequenom data allowed genotyping of >96% of heterozygous trisomic samples, with the magnitude of the *z* score bestowing a level of confidence on the genotype called (Fig. 5). On average, >90% of trisomic samples were greater than $\pm 1\ SD_{Ref}$ from the disomic reference PHR. These genotypes could be called with the highest degree of confidence. When the trisomic *z* score was less than ±0.5, genotypes could not be called with a high degree of confidence; this situation accounted for a mean of <4% of the samples. Overall, the PHR method of assigning genotypes achieved a success rate (96.1%) similar to that of the original Sequenom method (97.6%) (see Table 5 in the online Data Supplement).

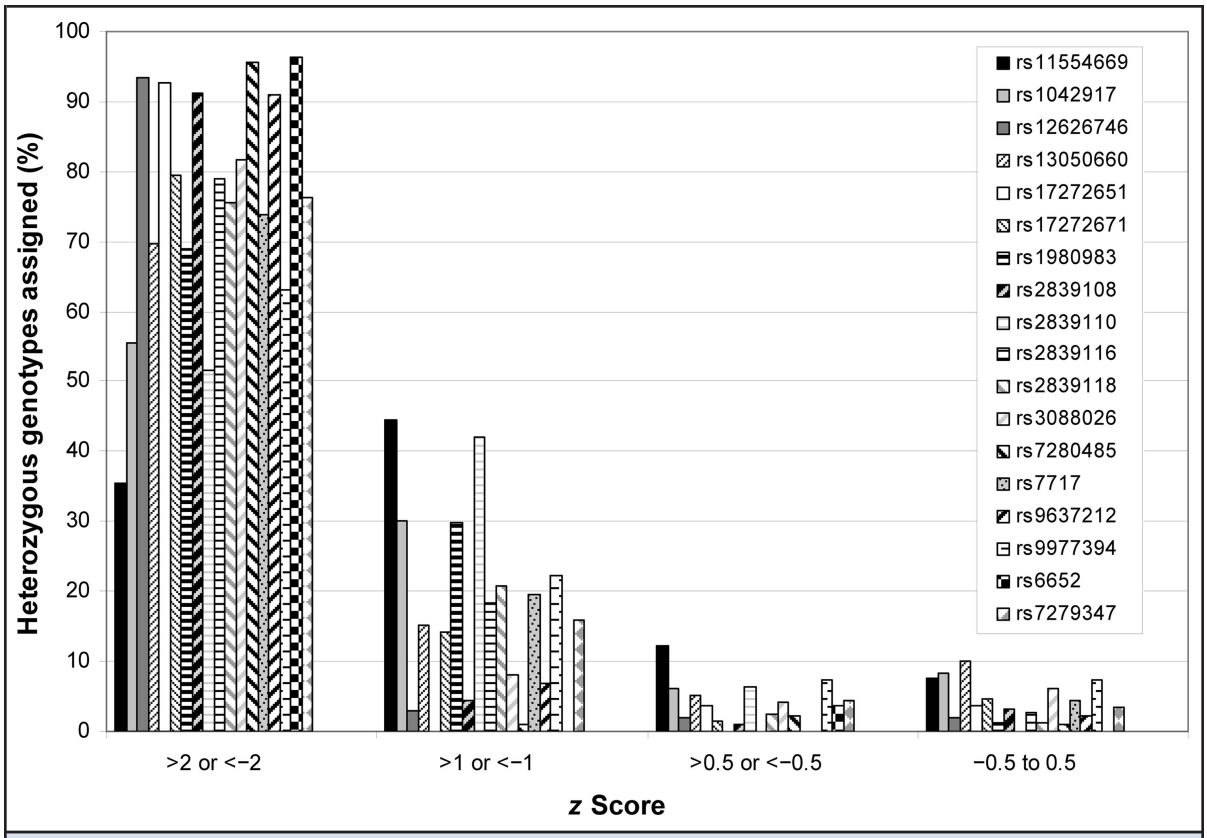


**Fig. 5. Distribution of *z* scores for the PHR method of genotyping trisomic heterozygote samples for 18 *COL6A2* SNPs.**

Genotypes were called with the highest degree of confidence when the PHR was >2. Only when the *z* score was between −0.5 and 0.5 could the genotype not be called. The SNPs are listed in the inset in the order that they occur in the figure proper.

The reproducibility of the technique was tested by reassaying 21 trisomic samples in a second assay. Ten of these samples were chosen because they were heterozygous at ≥7 SNP loci and there was sufficient sample to allow reassay. The other 11 samples were reassayed because one or more SNP assays had failed in the first analysis. A total of 97 heterozygous SNP genotypes were tested in the repeat assay. The heterozygous trisomic genotype obtained in the second assay was identical to that obtained in the first assay in 95 cases (97.9%).

### ASSIGNMENT OF THE ORIGIN OF NDJ

For studies of the transmission of loci segregating with DS subphenotypes, it is essential to determine the origin of the nondisjoining chromosome in trisomy. To be able to determine the origin of NDJ, it is critical that the trisomic genotype be determined correctly. Using the PHR method to genotype the 18 SNPs in *COL6A2* in the trisomic samples in this study, we were able to unambiguously assign the origin of NDJ in 76 (57.6%) of the 132 trisomic children. Maternal NDJ was seen in 67 (88.2%) of those samples genotyped, and paternal NDJ was seen in 9 samples (11.8%). Meiosis I NDJ was seen in 45 (67.2%) of the trios that showed maternal NDJ and in 6 (66.7%) of the trios that showed paternal NDJ. This result was consistent with findings reported in the literature *(10)*, thus supporting the validity of this method in assigning trisomic genotypes.

For the remaining 56 trios, the parental genotypes were such that it was not possible to determine the origin of NDJ. For example, if both parents had the same SNP haplotype or if parental SNP genotype data were missing, it was impossible to determine which parent transmitted the extra chromosome to the trisomic child. Because one of the quality metrics of the Sequenom assay was a genotyping success rate of at least 90% (which was calculated from the parental euploid samples), some parental SNP genotypes were unassigned because of assay failure. By assigning these missing genotypes by phasing with fastPHASE 1.2 *(11)*, we were able to determine the origin of NDJ in another 32 trios (24%). These additional data did not markedly affect the propor-

tions of maternal and paternal NDJ in the sample studied (88.9% and 11.1%, respectively).

## Discussion

The plethora of publicly available data regarding genetic variation across the genome is an exciting development in the quest to identify genes segregating with phenotype, and the accurate genotyping of samples is essential to this end. This consideration is particularly important when investigating CNVs and the effects of gene dosage imbalance.

The final effect of gene dosage imbalance will ultimately depend on the gene product and the interactions between it and other gene products. If the product of an affected gene is a transcription factor, the expression of downstream genes might be affected. Alternatively, if the gene product is a subunit of a multimeric protein, then it may adversely affect the interactions with the related gene products and lead to structural and/or functional abnormalities. The effect of gene dosage imbalance, whether due to trisomy or subchromosomal duplications, may be compounded by the presence of a functional SNP variant or a deleterious allele, which could lead to increased or decreased activity of the relevant gene.

When genotyping trisomic samples (or those with regions of genomic duplication), one of the main challenges to overcome is in discriminating between AAB and ABB heterozygous genotypes. The genotyping technique adopted for this study used a well-defined protocol capable of accurately genotyping samples under normal circumstances. We used this method to analyze 18 tag SNPs in the *COL6A2* gene for trisomic and euploid samples. The assay successfully genotyped the euploid samples but produced a high failure rate with samples from trisomic individuals. Visual inspection of the data allowed us to assign some of the trisomic heterozygote genotypes, but the 2 heterozygote states (AAB and ABB) were poorly discriminated. The data-analysis algorithms required for genotype assignment have been optimized for disomic samples and do not assign correct genotypes to heterozygous trisomic samples. We therefore devised a novel method for interpreting the data acquired from the mass spectrogram specifically for assigning genotypes to trisomic samples.

The PHR method we have described enabled accurate assignment of one or the other of the 2 possible trisomic heterozygote genotypes (AAB or ABB) by comparison to a reference interval established for disomic AB genotypes. The PHR method was successful in genotyping >96% of trisomic heterozygous samples for all SNPs analyzed, and these results were reproducible in a repeat assay. When we used the standard Sequenom genotyping algorithm, we were able to assign trisomic heterozygote genotypes correctly only by visually inspecting the data. The PHR method has 2 advantages over the Sequenom method: It allows better discrimination between the 2 possible trisomic heterozygous genotypes, and it assigns a numerical value to the genotype, which confers some degree of certainty to the result. Subsequently, we were able to determine the origin of NDJ in 81.8% of the family trios. Analyzing more SNPs will allow the determination of the origin of NDJ for the entire sample cohort. This method therefore has potential value for carrying out subsequent association studies, because it enables SNP genotyping and the determination of NDJ origin within the same high-throughput assay.

This information could have many applications in the future. Analysis of polymorphic variation on chromosome 21 might lead to the identification of loci predisposing to the subphenotypes in trisomy and provide markers for predicting the occurrence of clinically serious phenotypes, such as congenital heart defects (CHDs) *(12)*. This potential application could add an extra dimension to antenatal screening programs by contributing to the process of "informed choice" by the potential parents of a DS child, as well as well as aiding the long-term management of these patients. In addition, genes involved in CHDs in trisomy 21 may also be implicated in CHDs in euploid individuals. Knowledge of which genes are involved and how they interact with other gene products could have future implications for managing CHDs in the nontrisomic population. New therapeutic approaches that target pathways controlling cardiogenesis and that use stem cell–guided cardiac repair are now being developed [reviewed by Passier et al. *(13)*] and will be realistic therapeutic options for these patients in the future.

From a wider perspective, there is a high level of interest in the phenotypic implications of genomic structural variations *(12, 14)*. Although new methodologies are being developed for the computational imputation of CNVs *(7, 15)*, the fact that techniques for high-throughput laboratory analysis are still lacking has hampered efforts to assess the contribution of CNVs to human disease *(16)*. The approach we have described adds another tool to our armory, one that allows integrated SNP and copy number analysis of complex human disease.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

**Authors' Disclosures or Potential Conflicts of Interest:** *No authors declared any potential conflicts of interest.*

**Role of Sponsor:** The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

**Acknowledgments:** The authors thank the Genetics Service, Newcastle upon Tyne Hospitals NHS Foundation Trust, and the Haukeland University Hospital, Bergen, Norway, for providing the samples used in this study.

## References

1. de Smith AJ, Walters RG, Froguel P, Blakemore AI. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. Cytogenet Genome Res 2008;123:17–26.
2. Tsui NBY, Chiu RWK, Ding C, El-Sheikhah A, Leung TN, Lau TK, et al. Detection of trisomy 21 by quantitative mass spectrometric analysis of single-nucleotide polymorphisms. Clin Chem 2005;51:2358–62.
3. Tong YK, Ding C, Chiu RWK, Gerovassili A, Chim SSC, Leung TY, et al. Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: theoretical and empirical considerations. Clin Chem 2006;52: 2194–202.
4. Huang DJ, Nelson MR, Zimmerman B, Dudarewicz L, Wenzel F, Spiegel R, et al. Reliable detection of trisomy 21 using MALDI-TOF mass spectrometry. Genet Med 2006;8:728–34.
5. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006;444: 444–54.
6. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature 2010;464:704–12.
7. Coin LJM, Asher JE, Walters RG, El-Sayed Moustafa JS, de Smith AJ, Sladek R, et al. cnvHap: an integrative population and haplotype-based multiplatform model of SNPs and CNVs. Nat Methods 2010;7:541–6.
8. R Development Core Team. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2009. http://www.R-project.org (Accessed June 2011).
9. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.
10. Howard CM, Davies GE, Farrer MJ, Cullen LM, Coleman MM, Williamson R, et al. Meiotic crossing-over in nondisjoined chromosomes of children with trisomy 21 and a congenital heart defect. Am J Hum Genet 1993;53:462–71.
11. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78: 629–44.
12. de Smith AJ, Trewick AL, Blakemore AI. Implications of copy number variation in people with chromosomal abnormalities: potential for greater variation in copy number state may contribute to variability of phenotype. Hugo J 2010;4:1–9. http://www.springerlink.com/content/0m32872880427230/ (Accessed June 2011).
13. Passier R, van Laake LW, Mummery CL. Stem-cell-based therapy and lessons from the heart. Nature 2008;453:322–9.
14. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature 2010;463:671–5.
15. Su SY, Asher JE, Jarvelin MR, Froguel P, Blakemore AIF, Balding DJ, Coin LJM. Inferring combined CNV/SNP haplotypes from genotype data. Bioinformatics 2010;26:1437–45.
16. The Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010;464: 713–20.