A novel procedure for genotyping of single nucleotide polymorphisms in trisomy with genomic DNA and the invader assay

Kelly J. Duffy, Medical College of Wisconsin
Jack Littrell, Medical College of Wisconsin
Adam Locke, Emory University
Stephanie Sherman, Emory University
Michael Olivier, Medical College of Wisconsin

Journal Title: Nucleic Acids Research
Volume: Volume 36, Number 22
Publisher: Oxford University Press | 2008-12-01, Pages e145-e145
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1093/nar/gkn736
Permanent URL: https://pid.emory.edu/ark:/25593/r8h10

Final published version: http://dx.doi.org/10.1093/nar/gkn736

Copyright information:

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 2.0 Generic License (http://creativecommons.org/licenses/by-nc/2.0/), which permits making multiple copies, distribution, public display, and publicly performance, distribution of derivative works, provided the original work is properly cited. This license requires copyright and license notices be kept intact, credit be given to copyright holder and/or author. This license prohibits exercising rights for commercial purposes.

Accessed November 30, 2018 4:02 PM EST
A novel procedure for genotyping of single nucleotide polymorphisms in trisomy with genomic DNA and the invader assay

Kelly J. Duffy1, Jack Littrell1, Adam Locke2, Stephanie L. Sherman2 and Michael Olivier1,*

1Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI and 2Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

Received September 4, 2008; Revised and Accepted October 1, 2008

ABSTRACT

Individuals with trisomy 21 display complex phenotypes with differing degrees of severity. Numerous reliable methods have been established to diagnose the initial trisomy in these patients, but the identification and characterization of the genetic basis of the phenotypic variation in individuals with trisomy remains challenging. To date, methods that can accurately determine genotypes in trisomic DNA samples are expensive, require specialized equipment and complicated analyses. Here we report proof-of-concept results for an Invader® assay-based genotyping procedure that can determine SNP genotypes in trisomic genomic DNA samples in a simple and cost-effective manner. The procedure requires only two experimental steps: a real-time measurement of the fluorescent Invader® signal and analysis with a specifically designed clustering algorithm. The approach was tested using genomic DNA samples from 23 individuals with trisomy 21, and results were compared to genotypes previously determined with pyrosequencing. Additional assays for 15 SNPs were tested in a set of 21 DNA samples to assess assay performance. Our method successfully identified the correct SNP genotypes for the trisomic genomic DNA samples tested, and thus provides an alternative to determine SNP genotypes in trisomic DNA samples for subsequent association studies in patients with Down syndrome and other trisomies.

INTRODUCTION

Down syndrome is the phenotypic manifestation of trisomy of chromosome 21 and occurs in one in 800–1000 live births. Current diagnostic methods for detecting chromosomal abnormalities such as trisomy 21 are well established clinically. Such methods include karyotyping, fluorescent in situ hybridization (FISH), quantitative polymerase chain reaction (qPCR) (1) and, more recently, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (2). These techniques can reliably detect extra chromosome copy numbers, and thereby help in the clinical diagnosis of trisomies. However, this diagnosis fails to characterize and explain the genetic basis for the highly variable clinical presentation of patients with trisomy 21. It is well known that certain traits, such as cognitive impairment and dysmorphic features, occur in all affected individuals to variable degrees. Most other traits associated with Down syndrome are seen in only a portion of trisomy 21 individuals and occur with varying degrees of severity.

An extra copy of chromosome 21 should theoretically result in a 50% increase in gene expression levels for all genes on that chromosome. However, studies have shown that there is not always a direct correlation between chromosomal imbalance and gene expression levels (3–5). Recently, Prandini et al (6) measured gene expression levels for 100 and 106 expressed chromosome 21 genes in lymphocyte and fibroblast cell lines derived from individuals with trisomy 21, respectively. In their study, they uncovered considerable gene-expression variation for many chromosome 21 genes, suggesting that allelic differences in combination with gene dosage effects are likely contributing to the phenotypic variability. The ability to determine if this genotype-specific effect is contributing to the resulting variable phenotypes seen in individuals with trisomy 21, however, requires the ability to quantify the numbers of each allele present at any variable site on chromosome 21 in the trisomy samples.

The assessment of allelic differences in humans is most often performed using single nucleotide polymorphisms (SNPs) that are the most common form of genetic variation in the human genome. Due to their abundance and their biallelic nature, SNPs are commonly being used as markers in human disease studies. Numerous SNP genotyping methods have been developed that allow the efficient
interrogation of the two alleles in chromosomally normal DNA samples. In contrast to this well-established genotyping methodology for SNPs in chromosomally normal individuals, genotyping DNA samples from individuals with trisomy poses a greater challenge since two types of heterozygotes must be distinguished. The determination of SNP genotypes in a trisomic population requires specific techniques based on allele quantification, not just detection. The most commonly used procedures for determining SNP genotypes in trisomy include quantitative fluorescence-PCR (7,8), pyrosequencing (9), restriction fragment length polymorphism (RFLP) analysis (9,10), and, more recently, high-density oligonucleotide arrays (11,12). While accurate, these methods require special equipment, expensive reagents and complicated analyses, establishing the need for a new technique that is simple, efficient and cost-effective, and can be used to interrogate a small number of specific SNPs in a study cohort to examine allelic association with trisomy-related traits.

In 1999, Kwiatkowski et al. (13) developed the Invader® assay for SNP genotyping. This technique uses a structure-specific flap endonuclease (FEN) to cleave a three-dimensional complex formed by hybridization of allele-specific overlapping oligonucleotides. When the oligonucleotide complementary to the SNP allele in the target molecule anneals, it triggers the cleavage of the oligonucleotide by a thermostable FEN called cleavase XI enzyme. The cleavage product subsequently triggers a secondary cleavage reaction on a fluorescence resonance energy transfer (FRET) probe to generate a fluorescent signal. This signal amplification system can be used to quantify DNA targets, from both PCR product and genomic DNA, as well as mRNA targets for gene expression monitoring (13–18). Genotype data analysis tools have been developed for the efficient automated computational assignment of genotype results (18).

In the current study, we demonstrate that a real-time analysis of the Invader® assay and a modified data analysis clustering algorithm can be used to genotype SNPs from a trisomic population in a quantitative manner allowing for allele copy number determination. Further, we prove the success and accuracy of our technique using human genomic DNA from individuals with trisomy 21 without prior amplification. This novel application of the Invader® assay requires few experimental steps: real-time measurement of the fluorescent Invader® signal followed by analysis with a clustering algorithm developed in our laboratory. The specificity of the Invader® assay, the ability to genotype directly from genomic DNA and the accurate quantitative determination of allele copy number all make this approach an appealing alternative for SNP genotyping in trisomic DNA samples for the analysis of the genetic basis of trisomy-related phenotypes and their variability in individuals with Down syndrome.

MATERIALS AND METHODS

DNA

Genomic DNA samples were obtained from peripheral blood samples collected from 23 individuals with trisomy 21, as described in Kerstann et al. (9). Additionally, genomic DNA was isolated from 21 commercially available lymphoblast cell lines derived from individuals with trisomy 21 (Coriell Cell Repository, Camden, NJ, USA) and used in the optimization of our protocol. The study was approved by the Institutional Review Boards of the Medical College of Wisconsin and Emory University.

Genotyping assays

Invader® assays were designed using the InvaderCreator software (Third Wave Technologies, Madison, Wisconsin, USA). All assays were designed to be run at the same incubation temperature (63°C). Primary probe oligonucleotides ranged in size from 25 to 30 bases. Prior to use in the Invader® genotyping assay, oligonucleotides were purified by HPLC. Each 200μM probe of 100μl were added to a Falcon 96-well U-bottomed plate (Becton Dickinson, Franklin Lakes, New Jersey, USA) and placed in the autosampler of the HPLC system (Hitachi, Pleasanton, California, USA). The samples were run through a Tricorn Source 15Q 4.6/100 PE column (Amersham Biosciences, Pittsburg, Pennsylvania, USA) with mobile phase buffers (pH 7.35) at a flow rate of 1.0 ml/min. Fractions of each sample eluting from the column were collected in 30-s intervals for 1 h, while the absorbance values for each fraction were measured by UV detector and recorded. Fractions corresponding to the largest peak on the resulting chromatogram were further purified by de-salting using a 10 μM sodium phosphate buffer (pH 6.8) and a NAP-10 column (Pharmacia Biotech, Piscataway, New Jersey, USA), and were eluted with 1.5 ml of milli-Q water.

Genomic DNA samples were diluted to a final concentration of 50 ng/μl, denatured at 95°C for 10 min and placed immediately on ice. Invader® reactions were set up with the following final concentrations: 200 μM of each primary probe, 200 μM Invader probe, 10 ng/μl Cleavase XI enzyme, 230 mM MgCl₂, 2.6 M betaine, 1.4 μM FRET 1 probe and TE buffer (pH 8.0). Invader® assays for each allele of a given SNP were run in separate plate wells. Plates were run on an ABI 7900 real-time machine at 95°C for 5 min followed by incubation at 63°C for 7 h. During this time readings at 485 nm excitation and 530 nm emission were recorded for 10 s of every minute for 7 h.

Clustering algorithm

The clustering algorithm used in this study represents a modified version of the CA clustering algorithm described previously (18). Raw fluorescence values obtained from the real-time Invader® assay experiments were scaled to values between 0.05 and 0.95. The CA clustering algorithm was modified to accommodate a three-allele system by geometrically separating the cluster space using three lines instead of two lines to distinguish genotype clusters, one line to separate the two heterozygous clusters and two to separate the homozygotes from the heterozygotes. The lines were created in a manner such that the line distinguishing the two heterozygous clusters equally divides an imaginary line that runs from the upper left
quadrant center to the lower right quadrant center. The resulting halves were then divided into quarters, with one-quarter dedicated to the homozygous cluster and three-quarters to the heterozygous cluster. Data points were assigned to initial clusters based on the lines created by finding the probability ellipses for the points in each cluster (Figure 1). A probability was calculated for each point in a cluster using the probability ellipse and the point was assigned to the cluster where it had the highest probability.

RESULTS

In this study, we used the Invader® assay in combination with a trisomy-specific clustering algorithm to accurately genotype genomic DNA samples with trisomy 21. In this novel approach, genomic DNA was used directly without prior amplification to interrogate the two alleles of an SNP separately in a monoplex assay format with the same fluorescent dye to identify the cleavage products for each allele. Real-time measurements of the cleavage products allow for the quantitative determination of allele copy number, and thus the genotype of the trisomic DNA. Here, we present data for SNP genotyping of trisomic DNA, including heterozygous and homozygous allele combinations, obtained using this new approach for three representative SNPs on chromosome 21. Results from these Invader® assays were compared with results obtained previously by pyrosequencing using the same genomic DNA samples and SNP assays.

Real-time analysis

In order to alleviate any potential bias in allele amplification during PCR, we conducted our genotyping of trisomy 21 DNA experiments using genomic DNA. All genomic DNA samples were manually loaded in triplicate into a 384-well plate, denatured on a GeneAmp 9700 thermocycler (Applied Biosystems) for 10 min at 95°C and immediately placed on ice to stop the denaturation process. Invader® assay reagents for three SNPs (Table 1) were added to the denatured DNA samples in a monoplex assay format and incubated in an ABI 7900. For each genomic DNA sample and SNP Invader® assay reaction fluorescence data were averaged and recorded across 10 s of every minute for 7 h. These data were then averaged across replicates and plotted against time in minutes, with the resulting curves representing the quantity of each SNP allele detected over time. Figure 2 illustrates the fluorescence measures for allele ‘A’ of SNP rs874221 in six different trisomic DNA samples as fluorescence of FAM.

Table 1. Assay Primer and Probe Sequences

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>PCR primers</th>
<th>Sequence primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2837043</td>
<td>5’-GTCAGGAAGGGAATGGAGGGTCT-3’</td>
<td>5’-GGCATGAGGCCAG-3’</td>
</tr>
<tr>
<td>rs2837042</td>
<td>5’-CTCTGAGCGCAAGTGACTGT-3’</td>
<td>5’-AGGTAAGATCTCGATTC-3’</td>
</tr>
<tr>
<td>rs874221</td>
<td>5’-CTGATAAGGAACACTGAGGAC-3’</td>
<td>5’-TGACACTTGTGAGGC-3’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Primary probes</th>
<th>Invader probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2837043</td>
<td>5’-TTGAATGAACTCTTATATAGACCGCCTTACC-3’</td>
<td>5’-GAATGTGAACTCTTATATAGACCGCCTTACC-3’</td>
</tr>
<tr>
<td>rs2837042</td>
<td>5’-GGCGCGAGGTGAGCTGACATGGCCCTCA-3’</td>
<td>5’-AGATGAGACCGTACTGGGAAGGTTTTGATAAAATCTGGATA</td>
</tr>
</tbody>
</table>
probe 2, which was specifically designed for allele ‘A’. In the figure, the relative fluorescence signal obtained from the Invader® cleavage reaction was plotted along the y-axis, while time was plotted along the x-axis. The fluorescence values as measured at 485 nm excitation and 530 nm emission represent a quantitative measure of SNP rs874221 allele ‘A’ in each of the trisomic DNA samples.

**End-point determination and optimization**

To determine the time point that most clearly distinguished between fluorescence curves, and therefore allele quantity, we created scatter plots of fluorescence values taken every 30 min across the 7-h recording period for the three SNPs of interest in this study. In these plots, the fluorescence of allele ‘A’ for a given SNP was plotted against the fluorescence of allele ‘B’ for the same SNP at a given time point for 10 trisomic DNA samples run in duplicate. This was completed for all three SNPs of interest. For each of the SNPs, all time points from 210 to 300 min could accurately be used to create scatter plots that showed clear separation of clusters and allowed for genotype determination of DNA samples (Figure 3), whereas earlier time points could not clearly distinguish clusters. Any of these time points could then be used as end-point measurements for future genotyping studies using this SNP assay, thereby eliminating the need to run additional real-time studies and allowing the use of only a thermocycler and fluorescence plate reader for trisomic DNA genotyping.

**Clustering**

We confirmed our genotype determinations in these trisomic DNA samples by applying a clustering algorithm to the data from an end-point time identified previously (any time point between 210 and 300 min). The clustering algorithm used was a modified version of the CA clustering algorithm, which assigns individual data points to a predetermined number of clusters based on standard

---

**Figure 2.** Trisomic genomic DNA genotyping using real-time Invader assay. Real-time fluorescence curves for six different trisomy 21 genomic DNA samples generated during real-time Invader cleavage reaction. Only the fluorescence from one FAM probe (FAM probe 2, specific for allele ‘A’) is shown. The amount of allele-specific fluorescence corresponds to the copy number of that allele in each DNA sample.

**Figure 3.** Multiple time point scatterplots for rs874221. End-point genotype determinations for 60 and 270 min with trisomy 21 DNA cell lines are shown. Early time points cannot accurately identify genotypes. However, time points from 210 to 300 min can be reliably used to determine genotypes for this SNP using the modified clustering algorithm. Shown are genotype clusters with points in blue representing samples with the GGG genotype, green representing GGA samples, teal representing AAG and purple representing samples with the AAA genotype.
deviations of the distribution of data points around the centroid of each cluster. To account for the extra genotype cluster of trisomy samples, the original algorithm was modified by adding an extra line to distinguish the two heterozygote clusters in addition to the two previously existing lines that separate the heterozygotes from the homozygote clusters.

This algorithm was used to verify the genotype determinations of DNA samples from all 23 individuals with trisomy 21 and Invader® assays for three SNPs on chromosome 21 to determine genotype calls, using 90% probability as our cutoff. Because of these stringent specifications, any data point with <90% probability of residing within a given cluster would be considered a ‘no call’ by the algorithm. This assures accuracy in our genotype calls and increases the likelihood of distinguishing between the two heterozygous clusters.

**Comparison to pyrosequencing genotype calls**

Only those genotype results that were confidently called by the clustering algorithm (cluster probability of $\geq 90\%$) were compared with the genotypes determined by pyrosequencing. For SNPs rs2837043 and rs874221, the genotype calls determined by real-time Invader® for all samples tested correctly matched the genotype determination by pyrosequencing demonstrating 100% concordance. For SNP rs2837042, one discrepant call was obtained when calls from Invader genotyping were compared to pyrosequencing. While only one sample yielded a discrepant call, 17% of genotyping calls failed altogether, i.e. no genotype could be assigned due to low levels of detected fluorescence. The relatively high failure rate can be attributed to the DNA quality used for these experiments. In fact, an additional five DNA samples tested were excluded from the data analysis since none of the SNPs yielded any successful genotype calls.

**Analysis of invader® assay performance**

In order to better assess the genotype call rate using this Invader® assay approach, we genotyped 21 additional trisomy 21 DNA samples for 15 additional chromosome 21 SNPs. All SNPs were genotyped successfully, and analyzed as described above; however, we did not have independent verification of our genotype calls for these samples. Across all 15 SNPs, we observed a genotype call failure rate of 15.2% (0–80%) for the DNA samples used. Seven DNA samples had failure rates of >15% across the 15 SNPs. When these DNA samples are not included in the analysis, the average failure rate across the remaining 14 DNA samples for all 15 SNPs is 7.5%, a failure rate comparable to other genotyping platforms, such as Invader assay in chromosomally normal samples [2.3–3.2%, (19,20)], MassArray [6.4%, (21)] or FP-TDI [32.3%, (21)]. This observation suggests that a significant portion of the high failure rate observed in our genotyping can be attributed to specific DNA samples, such as DNA purity, concentration or fragmentation.

**DISCUSSION**

We have developed a novel genotyping procedure that allows accurate trisomy genotype calls to be determined from real-time Invader® assay fluorescence data in combination with a trisomy-specific clustering algorithm, modified from the CA clustering algorithm described by Olivier et al. (18). The new method has been developed to eliminate the need for amplification of DNA prior to genotyping with the Invader® assay, as has been shown in several other studies (15,17,22,23). Eliminating the PCR step alleviates any potential allele-specific amplification that may significantly impact the genotyping results, which are based on allele quantification. Thus, it decreases the number of false genotype calls that could result from PCR.

Our technique is significantly different from the other available options for trisomy genotyping in several additional respects: (i) it requires only a thermocycler or real-time quantitative PCR machine with no need for other more advanced specialized equipments; (ii) the analysis of results is simple, straightforward and can be performed using Microsoft Excel; and the trisomy-specific clustering algorithm that uses probability functions to assign data points to four separate genotype clusters; (iii) most importantly, the real-time Invader® reactions need only be run once for each SNP assay to determine the time points that are optimal for cluster differentiation. After the ideal time points have been determined for a given SNP they can be used as end-point measurements for any future genotyping of that SNP on any sample. Subsequent genotyping requires only a thermocycler to run the Invader® reaction and a fluorometer to measure the end-point fluorescence values that can be used for the clustering algorithm.

To assess the accuracy of our genotyping method, we used pyrosequencing results that had previously been determined and published (9) for these DNA samples and SNPs as an independent verification of our results. When we compared the genotype calls from both methods, we observed a 100% concordance with pyrosequencing calls for SNPs rs2837043 and rs874221. The third SNP we investigated, rs2837042, was a predominantly homoygous SNP. As a consequence, the samples we genotyped only represented two genotype clusters, one homozygous and one heterozygous cluster. This significantly reduces the accuracy and the confidence of genotyping assignments based on geometric clustering. Despite this limitation, when the data was analyzed by the clustering algorithm, 88% of the genotype calls still matched those of pyrosequencing. The average concordance of our genotyping method with pyrosequencing was $\sim 96\%$. While this degree of agreement is encouraging, the overall failure rate of genotype calls is a concern. The reasons for this are unclear but could potentially be DNA sample related. Alternatively, assay probes may not have been purified sufficiently, resulting in background fluorescence in the assay. This SNP also showed one discordant call, suggesting that the assay conditions or data analysis needs to be optimized. However, it is important to note that this discordant call could also be the result of a miscall by the pyrosequencing method. Although pyrosequencing is a
reliable genotyping method, there are potential sources of error that exist with this technology. Previous studies by our group comparing Taqman genotyping and Invader® assay genotyping demonstrated that discordant genotyping calls are equally likely in both methods (18). Based on these results, it is conceivable that the same error rates extend to pyrosequencing.

Our detailed analysis of the performance of 15 additional SNP assays using 21 DNA samples isolated from commercially available trisomy 21 cell lines suggests that the overall failure rate for genotyping of trisomic samples is as low as 7.5%. In fact, four of these SNPs were successfully genotyped for all DNA samples, indicating that the method may be susceptible to SNP-specific complications and DNA quality that could be overcome with additional assay optimization.

The development of SNP genotyping methods for trisomic populations is an important step in our effort to better understand the role that SNPs play in determining the phenotypic variation in individuals with Down syndrome. Our detailed analysis of the performance of 15 additional SNP assays using 21 DNA samples isolated from commercially available trisomy 21 cell lines suggests that the overall failure rate for genotyping of trisomic samples is as low as 7.5%. In fact, four of these SNPs were successfully genotyped for all DNA samples, indicating that the method may be susceptible to SNP-specific complications and DNA quality that could be overcome with additional assay optimization.

In addition to the described application to SNP genotyping, the technology is also potentially adaptable and suitable for other clinically relevant applications. Optimized assays with improved accuracy and reduced failure rate could potentially be cost-efficient alternatives for clinical genotyping and trisomy detection. In addition, the ability of the assay to quantitatively measure allele copy numbers could be exploited to develop alternative approaches to detect copy number variants in human DNA samples. However, further testing and optimization of individual assays will be required to validate the technology for these additional applications, and it remains to be seen whether Invader® technology would be superior to other already existing platforms for these applications.

The CA tool is available from the corresponding author upon request.

FUNDING
National Down Syndrome Society Charles Epstein Research Award (to M.O.). Funding for open access charge: Medical College of Wisconsin (Human and Molecular Genetics Center, institutional support.

Conflict of interest statement. None declared.

REFERENCES


