

Video Article

An Allele-specific Gene Expression Assay to Test the Functional Basis of Genetic Associations

Silvia Paracchini¹, Anthony P. Monaco¹, Julian C. Knight¹¹Wellcome Trust Centre for Human Genetics, University of OxfordCorrespondence to: Silvia Paracchini at silviap@well.ox.ac.ukURL: <https://www.jove.com/video/2279>DOI: [doi:10.3791/2279](https://doi.org/10.3791/2279)

Keywords: Cellular Biology, Issue 45, Gene expression, regulatory variant, haplotype, association study, primer extension, MALDI-TOF mass spectrometry, single nucleotide polymorphism, allele-specific

Date Published: 11/3/2010

Citation: Paracchini, S., Monaco, A.P., Knight, J.C. An Allele-specific Gene Expression Assay to Test the Functional Basis of Genetic Associations. *J. Vis. Exp.* (45), e2279, doi:10.3791/2279 (2010).

Abstract

The number of significant genetic associations with common complex traits is constantly increasing. However, most of these associations have not been understood at molecular level. One of the mechanisms mediating the effect of DNA variants on phenotypes is gene expression, which has been shown to be particularly relevant for complex traits¹.

This method tests in a cellular context the effect of specific DNA sequences on gene expression. The principle is to measure the relative abundance of transcripts arising from the two alleles of a gene, analysing cells which carry one copy of the DNA sequences associated with disease (the risk variants)^{2,3}. Therefore, the cells used for this method should meet two fundamental genotypic requirements: they have to be heterozygous both for DNA risk variants and for DNA markers, typically coding polymorphisms, which can distinguish transcripts based on their chromosomal origin (Figure 1). DNA risk variants and DNA markers do not need to have the same allele frequency but the phase (haplotypic) relationship of the genetic markers needs to be understood. It is also important to choose cell types which express the gene of interest. This protocol refers specifically to the procedure adopted to extract nucleic acids from fibroblasts but the method is equally applicable to other cell types including primary cells.

DNA and RNA are extracted from the selected cell lines and cDNA is generated. DNA and cDNA are analysed with a primer extension assay, designed to target the coding DNA markers⁴. The primer extension assay is carried out using the MassARRAY (Sequenom)⁵ platform according to the manufacturer's specifications. Primer extension products are then analysed by matrix-assisted laser desorption/ionization time of-flight mass spectrometry (MALDI-TOF/MS). Because the selected markers are heterozygous they will generate two peaks on the MS profiles. The area of each peak is proportional to the transcript abundance and can be measured with a function of the MassARRAY Typer software to generate an allelic ratio (allele 1: allele 2) calculation. The allelic ratio obtained for cDNA is normalized using that measured from genomic DNA, where the allelic ratio is expected to be 1:1 to correct for technical artifacts. Markers with a normalised allelic ratio significantly different to 1 indicate that the amount of transcript generated from the two chromosomes in the same cell is different, suggesting that the DNA variants associated with the phenotype have an effect on gene expression. Experimental controls should be used to confirm the results.

Video Link

The video component of this article can be found at <https://www.jove.com/video/2279/>

Protocol

1) Cell Culture

1. Select cell types expressing the gene of interest.
2. Select cell lines heterozygous both for DNA risk variants and transcribed coding polymorphism within the gene of interest (Figure 1).
3. Culture the selected cell lines according to the supplier's specifications and prepare 2 x 10 cm petri dishes for DNA and RNA extraction respectively.
4. When cells reach 80% confluence start the procedure to harvest cells and isolate DNA and RNA either following the procedure described below or using commercially available kits.

2) DNA Extraction

1. Remove media from the dish, wash with PBS and add to the dish 500 μ L of a lysis solution (0.6% SDS, 100 mM NaCl, 50 mM Tris, 20 mM EDTA and 50 μ g/mL RNase A).
2. Rock the plate gently for 20 minutes at room temperature.
3. Scrape the cell lysate into a microcentrifuge tube and incubate at 37 °C overnight.

4. Add proteinase K to a final concentration of 100 µg/mL and incubate at 37 °C overnight.
5. Extract the DNA with an equal volume of phenol-chloroform pH 8, repeat, then extract with an equal volume of chloroform/isoamyl alcohol.
6. Precipitate the DNA with 1/10th volume 3 M NaCl and 2.5 volumes of 100% ethanol.
7. Leave on ice for 5 min and then spin at 13,000 rpm for 30 min at 4 °C.
8. Wash the DNA with 70% ethanol, allow to air-dry and resuspend in 200 µL of TE.
9. Incubate at 65 °C for 10 min and then leave for 2 days at room temperature.
10. Store at 4 °C or long term at -20 °C.

3) RNA Extraction

1. Remove media from the dish, wash with PBS, add 1 mL of TRIzol and incubate for 10 min at room temperature.
2. Scrape cells, pipette a few times, transfer in a microcentrifuge tube and incubate for 5 min.
3. Spin for 10 min at 10,000 rpm at 4°C.
4. Transfer the supernatant in a fresh tube and add 200 µL of chloroform.
5. Shake and spin for 10 min at 10,000 rpm at 4°C.
6. Transfer the aqueous phase in a fresh tube and add an equal volume of 70% ethanol.
7. Load the solution to 1 or 2 (depending on the volume) RNeasy column and spin for 30 sec at maximum speed.
8. From here follow instruction of the RNeasy kit (Qiagen).

4) Nucleic Acid Sample Preparation

1. Quantify DNA and RNA concentration using a NanoDrop Spectrophotometer (Thermo Scientific).
2. Prepare cDNA from 500-1000 ng of RNA using random decamers and SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions.

5) PCR and Primer Extension Assay

1. Design two pairs of PCR primers for each DNA marker, one pair for genomic DNA amplification and the other pair for cDNA amplification (**Fig 2**). Place forward and reverse primer for cDNA amplification in different exons to avoid genomic contamination. Design primer to obtain a PCR product in the range of 100-500 bp length.
2. Prepare a 10 µL PCR reaction including 0.5 U Immolase Taq (Bioline) 0.8 mM dNTPs, 2 mM MgCl₂, 0.2 M each primer and 20 ng of either cDNA or DNA. Amplify each sample in four independent reactions.
3. Run the PCR reaction under the conditions optimized for each primer pair keeping cycle number in linear phase of amplification.
4. Remove PCR primer and dNTPs excess by incubating the PCR product with exonuclease I and Shrimp Alkaline Phosphatase (SAP) at 37 °C for 20 minutes.
5. Spot each PCR product twice on a 384-well plate, so that 8 measurements will be available for each sample.
6. Design the extension primer with the Assay Design (Sequenom) software so that the same primer can be used for both genomic and cDNA PCR products. Specify primer length in the range of 14-18 bp and extension mix including combinations of 3ddNTPs (**Fig 2**).
7. Carry out the primer extension reaction and MALDI-TOF/MS analysis using the MassARRAY system (Sequenom) according to the manufacturer's instruction.
8. A typical primer extension reaction include a starting step at 94 °C for 2 min, followed by 40 cycles of 94 °C for 5 s, 52 °C for 5 s and 72 °C for 5 s.
9. The primer extension products are cleaned with SpectroCLEAN resin and transferred onto microarray (chip) by SpectroPOINT nanolitre dispenser.
10. The extended oligonucleotides are quantified by MALDI-TOF using a SpectroREADER mass spectrometer.

6) Statistical Analysis

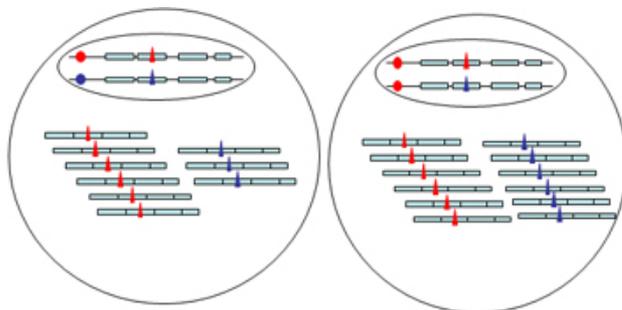
1. Check MALDI-TOF/MS results with the MassARRAY Typer software for general quality of the experiment.
2. Extract the Allelotyping report which includes the peak area data.
3. For each individual spectrum calculate the ratio between the peak area of allele 1 and allele 2 (peak area ratio).
4. For each cDNA sample derive a peak area ratio mean and standard deviation, using the corresponding function in Excel, across the 8 measurements.
5. Derive the same mean peak area ratio across the 8 measurements for genomic DNA.
6. Divide the cDNA mean ratio by the genomic mean ratio to assess the deviation to the expected 1:1 ratio.

7) Experimental Controls

1. To rule out experimental artifacts, repeat the experiment at least three times.
2. Whenever possible, design assays for additional DNA markers.
3. Design multiple PCR primer pairs and extension primers annealing to both forward and reverse strands.
4. Whenever possible test as negative controls cell lines that are heterozygous for DNA markers but do not carry the DNA risk variants associated with the phenotype (**Fig. 1** and **Fig. 3**)

8) Representative Results

An example of allele-specific expression analysis is shown in Figure 3 with examples of mass spectrometry traces. The figure shows the results from the analysis of 4 allele-specific primer extension products carried out on 4 different templates. The top graphs represent the results obtained from the analysis of genomic DNA of two different cell lines, which are both heterozygous for a transcribed coding polymorphism. However, only cell line A is heterozygous for the risk DNA variant (Figure 1). The lower graphs represent the results obtained from the analysis of the cDNA generated from the same cell lines. As a result of experimental artifact the first peak is always higher than the second peak even in the genomic analysis. Therefore, the results from the genomic analysis are used to normalize the cDNA data. After normalization, the allele ratio is significantly different from 1 in cell line A (where the second peak appears lower compared to the other spectra), while it is very close to 1 in cell line B. The data suggest that cell line A carries a DNA variant, in phase with the allele measured by the second peak, which reduces the expression of the gene under analysis. Cell line B provides a very convenient negative control.



Heterozygous at	Cell line A	Cell line B
Risk variant	✓	✗
Coding marker	✓	✓

Figure 1. Cell line selection strategy. Cell lines A and B are heterozygous for a transcribed coding marker (triangle) but only cell line A is heterozygous for risk DNA variant (circle). The blue allele of the risk variant is associated (either with a direct effect or because in close correlation with the actual functional variant) with a lower level of transcription.

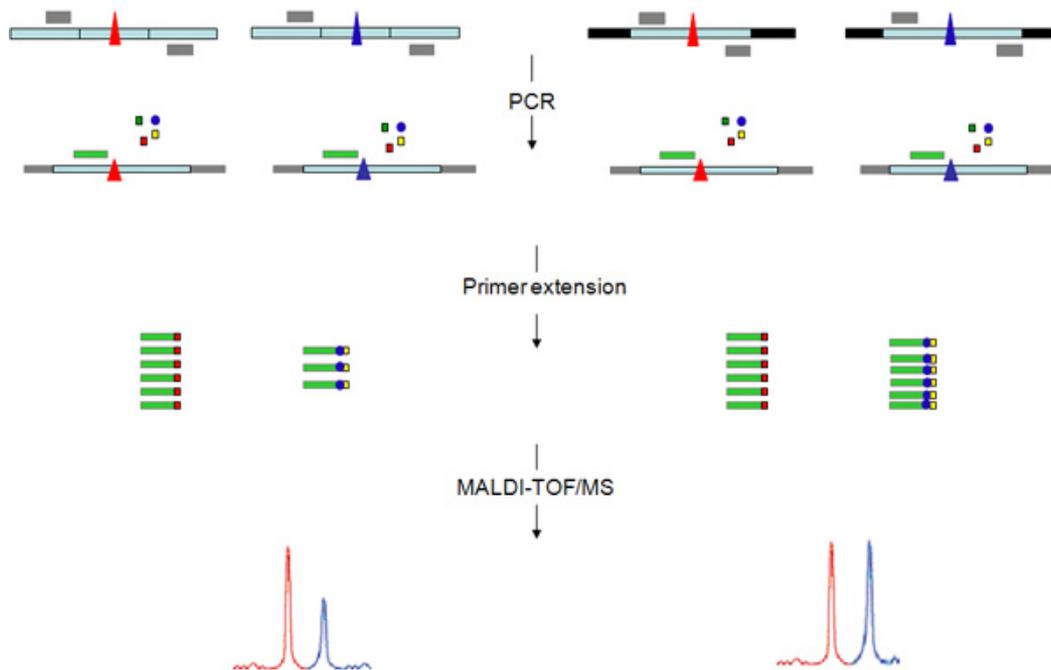


Figure 2. Allele-specific primer extension assay. This figure shows the workflow of the experimental procedure carried for both cDNA (left) and genomic DNA (right) templates. PCR primers (grey rectangles) are designed to amplify a heterozygous coding polymorphism (triangle). To avoid genomic contamination PCR primers anneal sequences placed in different exons (light blue bars) when amplification is carried out on cDNA template. The PCR product is then used as template for a primer extension reaction carried out with an extension primer, annealing one base next to the polymorphism, and the appropriate mix of three dideoxynucleotide (ddNTPs) terminators (squares) and one deoxynucleotide (circle) to extend the primer of one or two bases respectively. The resulting primer extension products have different masses which allow

separation and quantification by MALDI-TOF mass spectrometry. The quantity of product corresponding to the blue allele of the DNA marker is relatively lower to the red allele.

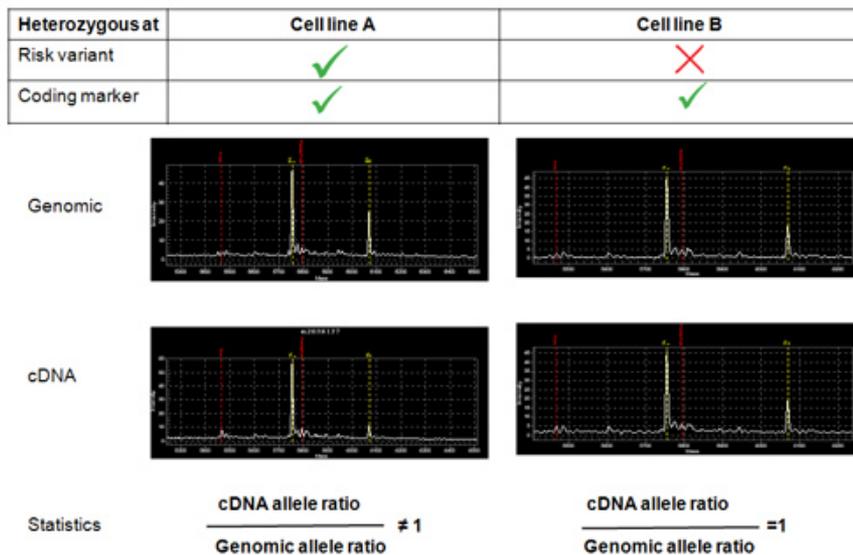


Figure 3. Results of an allele-specific expression assay. The figure shows the mass spectrometry results from the analysis carried out on the genomic DNA and the cDNA of cell line A and cell line B (Figure 1).

Discussion

This method enables evaluation of the effect of disease-associated DNA variants on gene expression by assessing *in vivo* allele-specific difference in transcript level. In this protocol relative transcript abundance is measured by MALDI-TOF but other technologies allowing allele-specific quantification, such as TaqMan^{6,7}, can be used. The major limitation of this approach is the availability of transcribed coding markers. Methods based on the principle described here, but taking advantage of different classes of polymorphisms, have been described. The haploChip assay measures allele-specific expression using markers located within 1 kb of the transcriptional start or end site of a gene which would be present in chromatin immunoprecipitated material isolated with antibodies specific to RNA polymerase II⁸. Alternatively, intronic SNPs can be used when the template is heteronuclear RNA⁹.

The specific protocol described here, in combination with the haploChip assay, has been used successfully to identify a candidate gene for dyslexia (or reading disability). Initially a genetic association study identified a DNA sequence associated with dyslexia and which was spanning three genes¹⁰. The allele-specific gene expression assay showed that in the presence of the dyslexia-associated DNA variants expression variation was observed for only one of the three genes but not the other two¹¹.

Disclosures

No conflicts of interest declared.

Acknowledgements

This work was funded by the Wellcome Trust through grants to A.P.M [076566/Z/05/Z], J.C.K. [074318] and a core award to the Wellcome Trust Centre for Human Genetics [075491/Z/04].

References

1. Cookson, W., Liang, L., Abecasis, G., Moffatt, M., & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat Rev Genet* 10 (3), 184-194 (2009).
2. Singer-Sam, J., LeBon, J.M., Dai, A., & Riggs, A.D. A sensitive, quantitative assay for measurement of allele-specific transcripts differing by a single nucleotide. *PCR Methods Appl* 1 (3), 160-163 (1992).
3. Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., & Kinzler, K.W. Allelic variation in human gene expression. *Science* 297 (5584), 1143 (2002).
4. Ding, C. & Cantor, C.R. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proc Natl Acad Sci U S A* 100 (6), 3059-3064 (2003).
5. Gabriel, S. & Ziaugra, L. SNP genotyping using Sequenom MassARRAY 7K platform. *Curr Protoc Hum Genet* Chapter 2, Unit 2 12 (2004).
6. Lo, H.S. *et al.* Allelic variation in gene expression is common in the human genome. *Genome Res* 13 (8), 1855-1862 (2003).
7. Liu, X. *et al.* Expression-based discovery of variation in the human glutathione S-transferase M3 promoter and functional analysis in a glioma cell line using allele-specific chromatin immunoprecipitation. *Cancer Res* 65 (1), 99-104 (2005).

8. Knight, J.C., Keating, B.J., Rockett, K.A., & Kwiatkowski, D.P. *In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet* 33 (4), 469-475 (2003).
9. Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 16 (2), 184-193 (2004).
10. Francks, C. *et al.* A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* 75 (6), 1046-1058 (2004).
11. Paracchini, S. *et al.* The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 15 (10), 1659-1666 (2006).