Methodology article

Simultaneous quantitative and allele-specific expression analysis with real competitive PCR

Chunming Ding*[1], Esther Maier[2], Adelbert A Roscher[2], Andreas Braun[3] and Charles R Cantor*[1,3]

Address: [1]Bioinformatics Program and Center for Advanced Biotechnology, Boston University, Boston, MA 02215 USA, [2]Children's Hospital, University of Munich, Lindwurmstrasse 4, 80337 Munich, Germany and [3]SEQUENOM, Inc., San Diego, CA 92121 USA

Email: Chunming Ding* - cmding@bu.edu; Esther Maier - emaier@kk-i.med.uni-muenchen.de; Adelbert A Roscher - Adelbert.Roscher@med.uni-muenchen.de; Andreas Braun - abraun@sequenom.com; Charles R Cantor* - ccantor@sequenom.com

* Corresponding authors

Abstract

Background: For a diploid organism such as human, the two alleles of a particular gene can be expressed at different levels due to X chromosome inactivation, gene imprinting, different local promoter activity, or mRNA stability. Recently, imbalanced allelic expression was found to be common in human and can follow Mendelian inheritance. Here we present a method that employs real competitive PCR for allele-specific expression analysis.

Results: A transcribed mutation such as a single nucleotide polymorphism (SNP) is used as the marker for allele-specific expression analysis. A synthetic mutation created in the competitor is close to a natural mutation site in the cDNA sequence. PCR is used to amplify the two cDNA sequences from the two alleles and the competitor. A base extension reaction with a mixture of ddNTPs/dNTP is used to generate three oligonucleotides for the two cDNAs and the competitor. The three products are identified and their ratios are calculated based on their peak areas in the MALDI-TOF mass spectrum. Several examples are given to illustrate how allele-specific gene expression can be applied in different biological studies.

Conclusions: This technique can quantify the absolute expression level of each individual allele of a gene with high precision and throughput.

Background

Mutations in the human genome can cause biochemical changes in protein products, or mRNA expression levels [1]. Epigenetic modifications such as DNA methylation can also change gene expression. Abnormal methylation is frequently detected in cancer (for review, see [2]). Allele-specific expression of heterozygotic genes (either a mutant allele or a wild type allele is expressed at a higher level than the other allele) might be one reason that carriers have different disease manifestations. Recently, Lo and colleagues showed that allelic variation in gene expression is common in the human genome [3]. Out of the 602 heterozygous genes surveyed in the kidney and lung tissues from seven individuals, 326 (54%) showed preferential

expression of one allele in at least one individual. Bray and colleagues showed that skewed allelic expression in the human brain occurs due to cis-acting variations [4]. In addition, Yan and colleagues found that allelic variation of gene expression can follow Mendelian inheritance [5].

Fluorescent dideoxy terminators and capillary gel electrophoresis [5], microarrays (Affymetrix HuSNP oligo arrays), real time PCR [3] and polymerase colonies [6] have been used for allele-specific expression. However, direct analysis of allelic expression has been limited for technological reasons. For example, capillary gel electrophoresis sometimes has difficulty detecting and quantifying extension products from the two different alleles. Real time PCR, when used for allele-specific detections, requires substantial optimization so that the two alleles can have sufficient annealing temperature difference. The technical difficulties have caused experimental inconsistencies such as debates over whether APOE ε 3/4 brain mRNA shows allele-specific expression [7,8].

Here we present a method that employs real competitive PCR [9] for allele-specific expression analysis. Real competitive PCR combines competitive PCR, primer extension reaction and matrix-assisted laser desorption/ionization – time of flight mass spectrometry (MALDI-TOF MS). Real competitive PCR can be used for high-throughput (>7,000 reactions/day using one mass spectrometer), absolute quantification of RNA transcripts with high precision (coefficient of variation < 10% with no assay optimization). To distinguish the two transcripts of a gene from the two alleles, transcribed single nucleotide polymorphisms (SNPs) or rare disease mutations in carriers are used as markers. MALDI-TOF MS can distinguish two oligonucleotides (in the 4000–9000 Da range) with molecular weight difference as small as a few Da, and thus can unambiguously detect the two alleles and the competitor. This technique can be used to quantify simultaneously the absolute expression level of each individual allele of a gene with high precision and throughput.

Results and Discussion

Recently, matrix-assisted laser desorption ionization – time of flight mass spectrometry (MALDI-TOF MS) was adapted for quantitative gene expression analysis [9]. This technique, dubbed as real competitive PCR, combines competitive PCR, primer extension reaction and MALDI-TOF MS. After isolation of RNA and reverse transcription, cDNA is spiked with a synthetic oligonucleotide (the competitor) with an identical sequence except one single base roughly in the middle of the sequence to the cDNA of interest. The competitor and the cDNA of interest are co-amplified by PCR. Excess dNTPs are removed by shrimp alkaline phosphatase treatment after PCR. Then, a base extension reaction is carried out with an extension primer,

a combination of three different ddNTPs and one dNTP and a ThermoSequenase. The base extension primer hybridizes right next to the mutation site and either one or two bases are added for the competitor and the cDNA, yielding two oligonucleotide products with different molecular weights (typically around 300 Da difference). In a typical molecular weight window of 4,000 to 9,000 Da, MALDI-TOF MS can easily distinguish two oligonucleotides if they differ by more than 10 Da. These two extension products are thus readily identified, and the ratio of their concentrations is quantified by MALDI-TOF MS.

As shown in Figure 1, when the synthetic mutation created in the competitor is close to a natural mutation site in the cDNA sequence, real competitive PCR can be used for accurate allele-specific gene expression analysis. PCR is used to amplify the two cDNA sequences from the two alleles and the competitor. A base extension reaction with a mixture of three different ddNTPs and one dNTP is used to generate three (instead of two in a typical real competitive PCR experiment) oligonucleotides for the two cDNAs and the competitor. The three products are identified and their ratios are calculated based on their peak areas in the mass spectrum.

Since the amount of competitor spiked in is known, the absolute concentration of each of the two cDNAs can be easily calculated. Thus, it is possible to simultaneously quantify the gene expression levels from the two alleles of one gene. The competitor and the two cDNAs are virtually identical in sequence and are amplified with the same kinetics. The allele specificity is superior due to the high precision of MALDI-TOF MS in molecular weight determination.

One example of allele-specific expression analysis by real competitive PCR is shown in Figure 2A. A single nucleotide polymorphism (refSNP ID: rs2069849) located in exon 2 of the interleukin 6 gene is selected as the marker for allele-specific expression. Complementary DNA (0.025 ng) prepared from the IMR-90 cell line (ATCC) was co-amplified with 5×10^{-22} Mol (301 copies) of the competitor. The oligonucleotide products from the base extension reaction were analyzed by MALDI-TOF MS. The peak area ratios represent accurately the concentration ratios of the two cDNAs and the competitor. Coefficient of variations (CV is defined as standard deviation divided by the mean) for the relative frequencies of the three peaks were 9.2%, 4.1% and 4.4% for four real competitive PCR replicates, indicating excellent precision. The interleukin 6 gene also shows modest skewing in allelic expression (98 copies of C allele was expressed, and 136 copies of T allele was expressed, see Figure 2A).
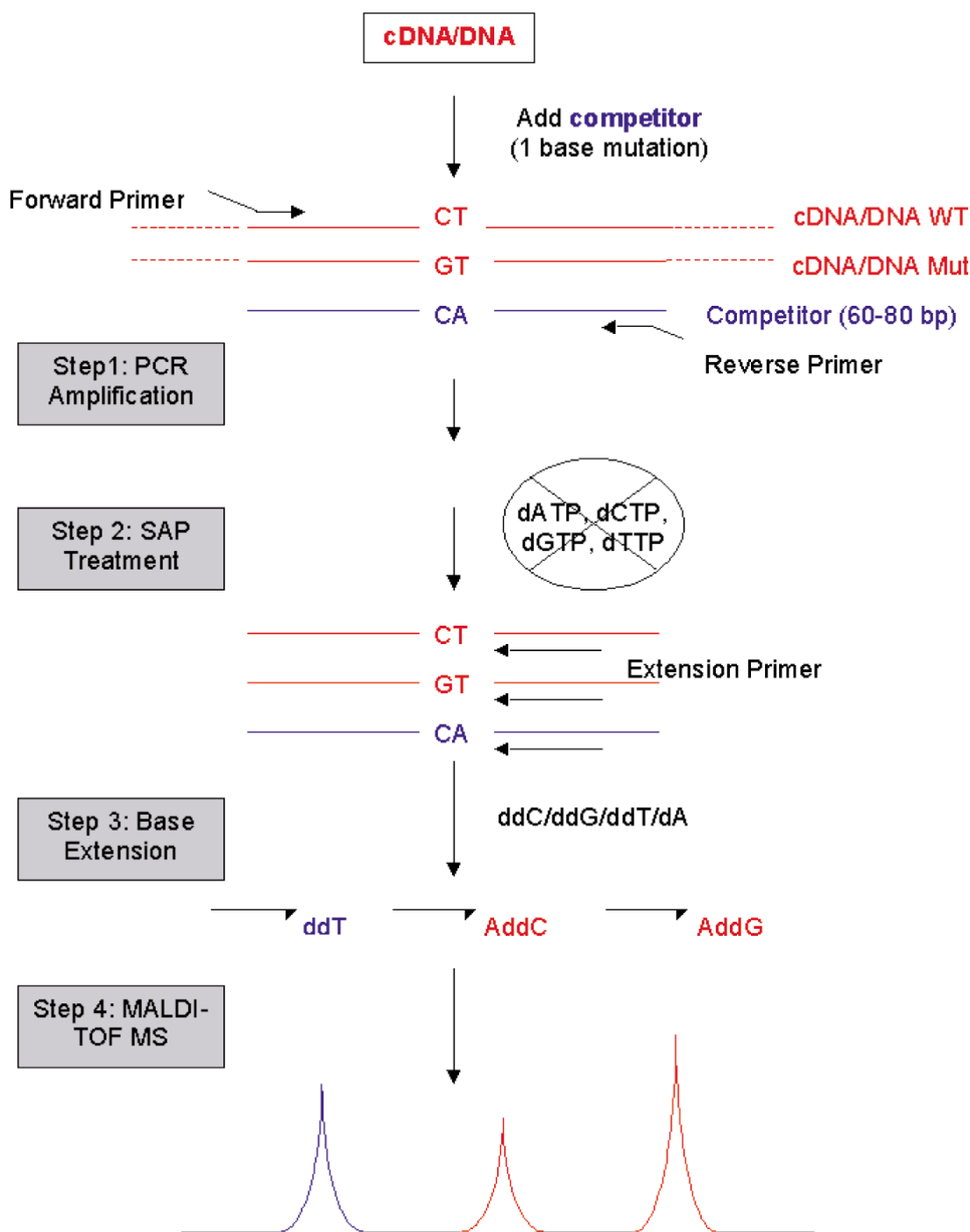
Figure 1
Schematic view of quantitative and allele-specific expression analysis with real competitive PCR. A point mutation in the cDNA sequence is used as the marker for allele-specific gene expression analysis. The competitor is designed to have a synthetic mutation next to the natural mutation and is used for quantitative gene expression analysis. Three extension products from the two cDNA sequences and the competitor have different molecular weights, and are detected by MALDI-TOF MS. The peak area ratios of these products represent accurately the concentration ratios of the two cDNAs and the competitor. Since the absolute quantity of the competitor is known, the absolute quantities of the two cDNA sequences can be readily calculated.
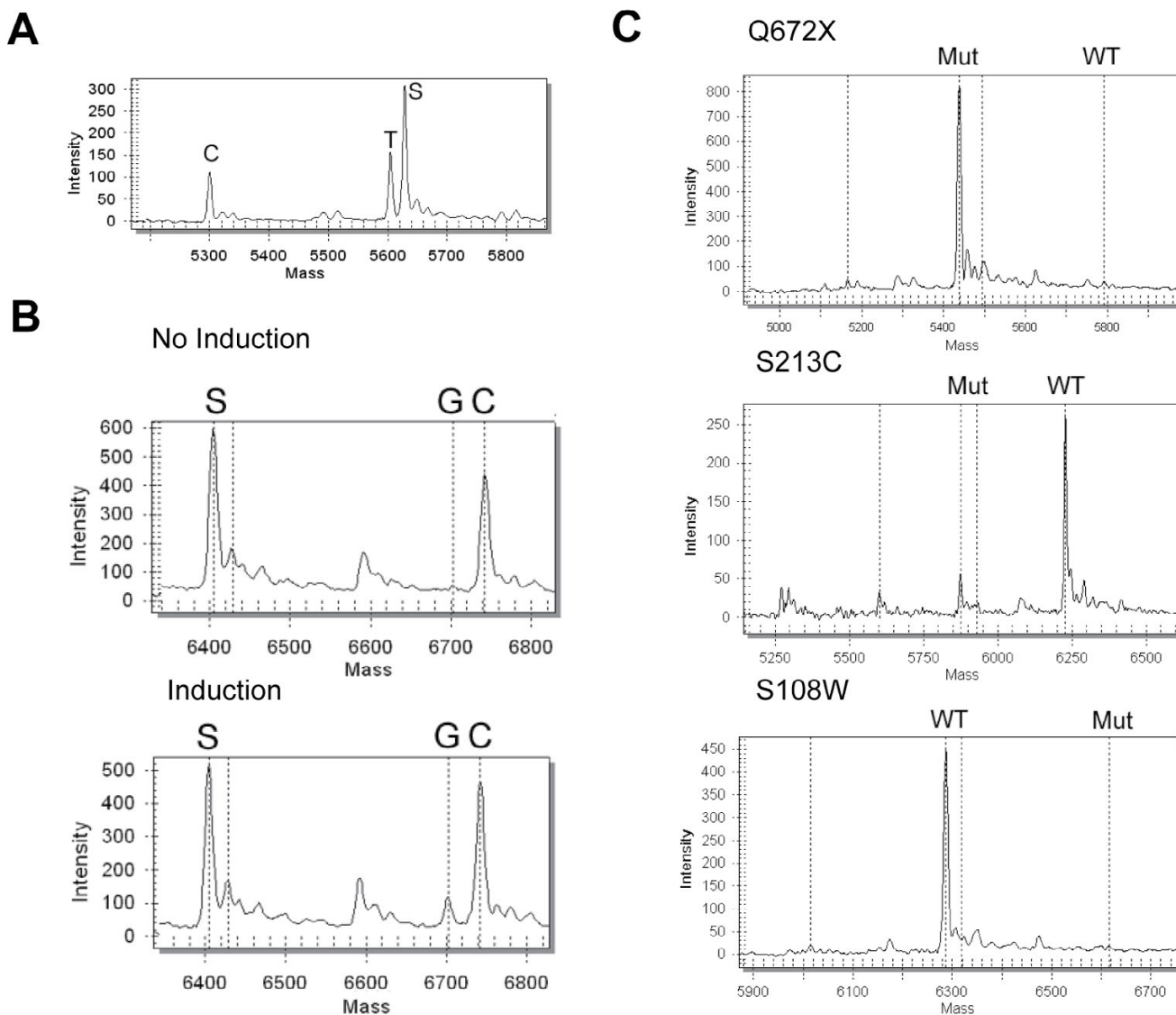
Figure 2
Mass spectra for allele-specific expression analysis. (**A**) Interleukin 6 gene. Peaks are identified by C, T and S. C represents the allele where the polymorphic site has a C residue. T represents the allele where the polymorphic site has a T residue. S represents the competitor. The peak areas of C, T and S peaks are automatically computed by the RT software package (SEQUENOM). The peak area ratios represent the concentration ratios of the starting cDNA sequences and the competitor. The peak frequencies are 0.209, 0.263 and 0.528 for peak C, T and S, respectively. (**B**) *lexA* gene. Peak S, G and C represent the competitor, the exogenous and the endogenous *lexA* gene, respectively. Without arabinose induction, only endogenous *lexA* gene expression was seen. With modest arabinose induction, both the endogenous and exogenous *lexA* gene expression were seen. Without induction, the peak frequencies are 0.601, 0.004 and 0.395 for peak S, G and C, respectively. With induction, the peak frequencies are 0.509, 0.075 and 0.416 for peak S, G and C, respectively. (**C**) *ABCD-1* gene. Mut and WT represent mutant and wild type alleles, respectively. For Q672X, the peak frequencies are 0.984 and 0.016 for peak Mut and WT, respectively. For S213C, the peak frequencies are 0.187 and 0.813 for peak Mut and WT, respectively. For S108W, the peak frequencies are 0.995 and 0.005 for peak WT and Mut, respectively.

We next tested allele-specific expression of the *lexA* gene in *Escherichia coli*. Gene expression perturbation in *E. coli* was used for gene network studies [10]. Expression perturbation was achieved by introducing an exogenous copy for

each gene of interest in an inducible expression plasmid. The expression of each gene potentially in a gene regulatory network was perturbed via the induction of the exogenous gene expression, and the expression changes of other genes were analyzed. These perturbed gene expression levels were then fed into a multiple linear regression algorithm to estimate the network interactions. This approach appears to be a powerful tool for functional genomics analysis. However, self-regulatory interactions such as positive and negative self-feedbacks can only be resolved by measuring the exogenous and endogenous gene expression separately. In the original study on the *E. coli* network, a reporter gene (*luciferase*), expressed under identical conditions as the gene of interest, was used to estimate the exogenous gene expression. However, this estimate is likely to be inaccurate since the expression level of the *luciferase* gene is likely to be different from the exogenous genes, even when they are under the control of the same promoter. If we can directly and separately quantify the expressions of the exogenous and the endogenous gene, we will be able to obtain significantly more accurate estimates of self-regulatory interactions in gene networks. To this end, an exogenous *lexA* was introduced into *E. coli* via the pBADX53 vector. The exogenous *lexA* gene is distinguishable from the endogenous *lexA* gene by a silent mutation (TCC to TCG silent mutation at codon 103). The exogenous *lexA* expression was induced with arabinose. Without arabinose, only endogenous *lexA* transcript was detected (Figure 2B). With an intermediate arabinose induction, exogenous *lexA* was expressed at about 20% level compared with the endogenous *lexA* (Figure 2B).

In the third example, we tested allele-specific expression of the *ABCD-1* gene (located on the X chromosome) involved in X-linked adrenoleukodystrophy (XALD). The manifestation of symptoms in XALD carriers was previously shown to be associated with a higher degree of non-random X chromosome inactivation [11]. A non-random X chromosome inactivation is likely to cause a preferential expression down-regulation of one of the *ABCD-1* allele. If the wild type allele is inactivated, the mutant allele will be predominantly expressed. Thus, the individual might show symptoms similar to a homozygous mutant. X chromosome inactivation studies can only provide a genome-wide, indirect picture while direct allele-specific gene expression can provide the direct link between gene expression and disease manifestation. We thus carried out allele-specific gene expression for three carriers with three different *ABCD-1* mutations (S108W, S213C and Q672X). The S108W carrier showed predominant (>99%) mutant allele expression while the S213C and Q672X showed predominant wild type allele (89% and >99%, respectively) expression (Figure 2C). This result is in complete concordance with results obtained previously [11].

Conclusions

We present here a straightforward method for quantitative and allele-specific gene expression analysis with real competitive PCR. The allele specificity for gene expression analysis is based on the superior molecular weight determination ability of the MALDI-TOF MS technology. Highly precise (CV 4% – 9%) and absolute gene expression analysis is achieved. In addition, the real competitive PCR is based on the highly automated MassARRAY system (SEQUENOM), and is ideal for high-throughput (7000 reactions/day/instrument) analysis. The high-throughput and low cost features of this technique can easily be exploited in large-scale allele-specific expression studies.

Methods

cDNA and oligonucleotides

Interleukin 6 gene expression analysis

Complementary DNA for interleukin 6 gene expression analysis was prepared from cell line IMR-90 (ATCC). The PCR primer sequences for the interleukin 6 gene expression analysis are, 5'-ACGTTGGATGGCAGGACAT-GACAACTCATC-3' and 5'-ACGTTGGATGCCATGCTACATTTGCCGAAG-3'. The extension primer sequence is 5'-CGCAGCTTTAAG-GAGTT-3'. The synthetic competitor sequence is 5'-GCCCATGCTACATTTGCCGAAGAGCCCTCAGGCTGGA CTGCATAAACTCCTTAAAGCTGCGCAGAATGAGAT-GAGTTGTCATGTCCTGCAG-3'. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). The synthetic competitor was PAGE purified by the vendor and absorbance at 260 nm was measured in our laboratory.

lexA gene expression analysis

RNA samples for *lexA* gene expression analysis were provided by Dr. Timothy Gardner (Boston University). The exogenous *lexA* gene has a TCC to TCG silent mutation at codon 103 so that it can be distinguished from the endogenous *lexA* gene. The exogenous *lexA* gene was cloned in the vector pBADX53. Bacterial culture and RNA extraction were carried out as previously described [10]. The PCR primer sequences for the *lexA* gene expression analysis are, 5'-ACGTTGGATGGCGCAACAGCATATTGAAGG-3' and 5'-ACGTTGGATGACATCCCGCTGACGCGCAGC-3'. The extension primer sequence is 5'-ATCAGCATTCGGCTT-GAATA-3'. The synthetic competitor sequence is 5'-ACATCCCGCTGACGCGCAGCAGGAAATCAGCATTCGG CTTGAATATGGAAGGATCGACCTGATAATGACCT-TCAATATGCTGTTGCGC-3'. The synthetic competitor was PAGE purified by the vendor and absorbance at 260 nm was measured in our laboratory.

ABCD-1 gene expression analysis

Complementary DNA and genomic DNA samples for *ABCD-1* gene expression analysis were prepared as

previously described [11]. Three *ABCD-1* carriers, S108W, S213C and Q672X, were used in this study. PCR primers for the three mutations are: 5'-ACGTTGGATGAGCAGCT-GCCAGCCAAAAGC-3' and 5'-ACGTTGGATGACTCG-GCCGCCTTGGTGAG-3' for S108W, 5'-ACGTTGGATGTAGGAAGTCACAGCCACGTC-3' and 5'-ACGTTGGATGAACCCTGACCAGTCTCTGAC-3' for S213C, and 5'-ACGTTGGATGTCCCTGTGGAAATAC-CACAC-3' and 5'-ACGTTGGATGAGTCCAGCTTCTC-GAACTTC-3' for Q672X. The extension primers are: 5'-GGCGGGCCACATACACC-3' for S108W, 5'-AGTGGCTT-GGTCAGGTTG-3' for S213C and 5'-AATAC-CACACACACTTGCTA-3' for Q672X.

Real competitive PCR
Real competitive PCR was carried out as was previously described [9].

Step 1: PCR amplification
Each PCR reaction contains 1 µL diluted cDNA (0.025 ng/µL), 0.5 µL 10× HotStar *Taq* PCR buffer, 0.2 µL MgCl$_2$ (25 mM), 0.04 µL dNTP mix (25 mM each), 0.02 µL HotStar *Taq* Polymerase (50 U/µL, Qiagen), 0.1 µL competitor oligonucleotide (5 × 10^{-9} µM), 1 µL forward and reverse primer (1 µM each) and 2.14 µL ddH$_2$O. The PCR condition was: 95°C for 15 min for hot start, followed by denaturing at 94°C for 20 sec, annealing at 56°C for 30 sec and extension at 72°C for 1 min for 45 cycles, and finally incubated at 72°C for 3 min.

Step 2: Shrimp alkaline phosphatase treatment
PCR products were treated with shrimp alkaline phosphatase to remove excess dNTPs. A mixture of 0.17 µL hME buffer (SEQUENOM), 0.3 µL shrimp alkaline phosphatase (SEQUENOM) and 1.53 µL ddH$_2$O was added to each PCR reaction. The reaction solutions (now 7 µL each) were incubated at 37°C for 20 min, followed by 85°C for 5 min to inactive the enzyme.

Step 3: Single base extension reaction
For each base extension reaction, 0.2 µL of selected ddNTPs/dNTP mixture (SEQUENOM), 0.108 µL of selected extension primer, 0.018 µL of ThermoSequenase (32 U/µL, SEQUENOM) and 1.674 µL ddH$_2$O were added. The base extension condition was, 94°C for 2 min, followed by 94°C for 5 sec, 52°C for 5 sec and 72°C for 5 sec for 40 cycles. The ddNTPs/dNTP mixtures are: ddATP/ddCTP/ddGTP/dTTP for interleukin 6 and *ABCD-1* Q672X, ddTTP/ddCTP/ddGTP/dATP for *lexA*, and ddATP/ddCTP/ddTTP/dGTP for *ABCD-1* S108W and S213C.

Step 4: Liquid dispensing and MALDI-TOF MS
The final base extension products were treated with SpectroCLEAN (SEQUENOM) resin to remove salts in the reaction buffer. This step was carried out with a Multimek (Beckman) 96 channel auto-pipette and 16 µL resin/water solution was added into each base extension reaction, making the total volume 25 µL. After a quick centrifugation (2,500 rpm, 3 min) in a Sorvall legend RT centrifuge, approximately 10 nL of reaction solution was dispensed onto a 384 format SpectroCHIP (SEQUENOM) pre-spotted with a matrix of 3-hydroxypicolinic acid (3-HPA) by using a MassARRAY Nanodispenser (SEQUENOM). A modified Bruker Biflex MALDI-TOF mass spectrometer was used for data acquisitions from the SpectroCHIP. Mass spectrometric data were automatically imported into the SpectroTYPER (SEQUENOM) database for automatic analysis such as noise normalization and peak area analysis.

Authors' contributions
CD conceived the project, performed the experiments in this manuscript and wrote the manuscript. EM, AAR, AB helped the XALD experiment and prepared the cDNA samples. CRC provided guidance and funding for this project. All authors read and approved the manuscript.

References
1. Cowles CR, Joel NH, Altshuler D, Lander ES: **Detection of regulatory variation in mouse genes.** *Nat Genet* 2002, **32:**432-437.
2. Momparler RL: **Cancer epigenetics.** *Oncogene* 2003, **22:**6479-6483.
3. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP: **Allelic variation in gene expression is common in the human genome.** *Genome Res* 2003, **13:**1855-1862.
4. Bray NJ, Buckland PR, Owen MJ, O'Donovan MC: **Cis-acting variation in the expression of a high proportion of genes in human brain.** *Hum Genet* 2003, **113:**149-153.
5. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW: **Allelic variation in human gene expression.** *Science* 2002, **297:**1143.
6. Butz JA, Yan H, Mikkilineni V, Edwards JS: **Detection of allelic variations of human gene expression by polymerase colonies.** *BMC Genet* 2004, **5(3):**.
7. Growdon WB, Cheung BS, Hyman BT, Rebeck GW: **Lack of allelic imbalance in APOE epsilon3/4 brain mRNA expression in Alzheimer's disease.** *Neurosci Lett* 1999, **272:**83-86.
8. Lambert JC, Perez-Tur J, Dupire MJ, Galasko D, Mann D, Amouyel P, Hardy J, Delacourte A, Chartier-Harlin MC: **Distortion of allelic expression of apolipoprotein E in Alzheimer's disease.** *Hum Mol Genet* 1997, **6:**2151-2154.
9. Ding C, Cantor CR: **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* 2003, **100:**3059-3064.
10. Gardner TS, di Bernardo D, Lorenz D, Collins JJ: **Inferring genetic networks and identifying compound mode of action via expression profiling.** *Science* 2003, **301:**102-105.
11. Maier EM, Kammerer S, Muntau AC, Wichers M, Braun A, Roscher AA: **Symptoms in carriers of adrenoleukodystrophy relate to skewed X inactivation.** *Ann Neurol* 2002, **52:**683-688.