

## Methods

# Accurate quantitation of allele-specific expression patterns by analysis of DNA melting

Sangkyun Jeong,<sup>1</sup> Yoonsoo Hahn,<sup>2</sup> Qi Rong,<sup>1</sup> and Karl Pfeifer<sup>1,3</sup>

<sup>1</sup>Laboratory of Mammalian Genes and Development, National Institute of Child Health and Development, National Institutes of Health, Bethesda, Maryland 20892, USA; <sup>2</sup>Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Epigenetic and genetic mechanisms can result in large differences in expression levels of the two alleles in a diploid organism. Furthermore, these differences may be critical to phenotypic variations among individuals. In this study, we present a novel procedure and algorithm to precisely and accurately quantitate the relative expression of each allele. This method uses the differential melting properties of DNAs differing at even a single base pair. By referring to the melting characteristics of the two pure alleles, the fractional contribution of the two alleles to any unknown mixture can be mathematically resolved. These methods are highly accurate and precise because each single melting reaction yields multiple data points for analysis. Finally, we discuss how this approach can be used more generally to accurately quantitate gene expression relative to known standards.

[Supplemental material is available online at [www.genome.org](http://www.genome.org) and at <http://pfeiferlab.nichd.nih.gov/>.]

The two alleles of a gene in diploid organisms are not always expressed equally. This unequal expression can be due to epigenetic or to genetic processes. Epigenetic mechanisms that lead to unequal allele expression include genomic imprinting and X-inactivation. In addition, an increasing number of genes are being identified that are subject to allelic exclusion but where the choice of allele appears to be stochastic. Some examples of genes thus regulated include immunoglobulin and odorant receptors in mammals. Genetic mechanisms can also account for differences in allele expression levels. Polymorphism in *cis* regulatory sequences that control RNA synthesis and/or stability can result in differential expression of two alleles (Pastinen and Hudson 2004). All these differences in mRNA expression levels between alleles have the potential to give rise to differences in the total biochemical or biophysical activity of the expressed molecules (Yan et al. 2002a; Ueda et al. 2003) and, therefore, confer variable fitness to their host organism. Thus understanding polymorphic alleles with respect to their relative expression level may provide insights into the mechanisms of phenotypic variation of biomedical significance.

Analysis of allelic expression variation depends on identification of a single nucleotide polymorphism (SNP) within the RNA coding sequence. Based on the SNP, the relative expression levels have been assessed by several methods including RNase Protection Assay (Winter et al. 1985) and Single Nucleotide Primer Extension assayed by radioactive nucleotide incorporation (SNUPE) (Kuppuswamy et al. 1991), or by mass spectrophotometry (rcPCR) (Knight et al. 2003). These methods are all technically challenging and, more importantly, limited in their ability to precisely quantitate variations in allelic expression.

We have developed a novel theorem for the quantification of a mixture of two different cDNAs by exploiting the unique melting properties of the cDNA variants. In this study, we successfully apply this analysis to establish a very rapid and accurate

procedure for the quantitative determination of the allelic variation between two polymorphic alleles. In addition, we discuss the general applicability of our procedure in quantitating gene expression.

## Results

A double-stranded DNA (dsDNA) molecule melts to two molecules of single-stranded DNA (ssDNA) under conditions that abrogate the interacting forces between bases. Melting of a dsDNA by continual increase of temperature is easily achieved with the aid of a conventional real-time thermocycler and yields a sigmoidal melting curve when the amount of dsDNA is plotted against temperature.

Even DNAs carrying a single nucleotide polymorphism (SNP) can be distinguished based on their unique melting curves (Fig. 1). We have identified an SNP in the 3'-UTR of murine alpha-fetoprotein (*Afp*) cDNA that distinguishes 129 and FVB strains (Fig. 1A). A probe pair, Afp-F and Afp-R, was designed to hybridize to the SNP-containing region. The Afp-R probe emits red light with maximum intensity at 640 nm from its Red640 dye by fluorescence resonance energy transfer (FRET) from the activated fluorescein dye of Afp-F. This energy transfer can occur only when the two dyes are in close proximity as when they each hybridize to their adjacent cognate sequences on the DNA. Thus the intensity of emitted fluorescent light from Red640 reflects the amount of probe-DNA hybridization.

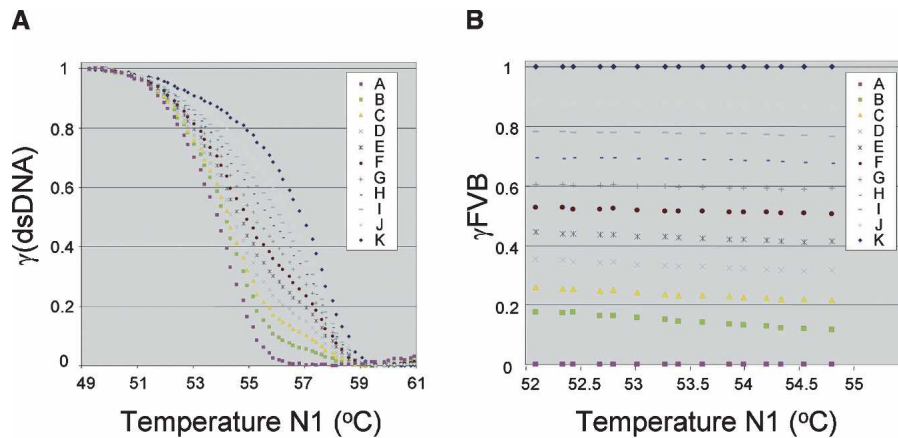
The PCR-amplified *Afp* cDNA amplicons of FVB (Fig. 1B, purple squares) and of 129 (Fig. 1B, blue diamonds) origins were each separately annealed with FRET probes and their melting behaviors analyzed (Fig. 1B). The resulting melting curves were then normalized by converting the maximum fluorescence value for each amplicon to 1 and the minimum fluorescence value to 0, with all other values adjusted proportionally. This normalization converts the Y-axis into the fraction of dsDNA [ $\gamma(\text{dsDNA})$ ] at a given temperature in each melting, and enables different samples and runs to be compared and analyzed in a standardized manner. The Afp-R probe matches perfectly with the FVB amplicon but

### <sup>3</sup>Corresponding author.

E-mail [kpfeifer@mail.nih.gov](mailto:kpfeifer@mail.nih.gov); fax (301) 402-0543.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.6028507>.





**Figure 2.** Melting curves for composite mixtures reflect the relative contribution of each of the two alleles. (A) Melting of FRET probe Afp-R from PCR amplicons of pure 129 (curve A), pure FVB (curve K), and of 129 + FVB mixtures (curves B–J). Mixes were generated prior to PCR amplification by pooling FVB and 129 cDNAs in proportions defined in Table 1. Note the ordered leftward shift (decreased  $T_m$  values) as the fractional value of 129 cDNA increases. (B) The calculated fractional values for FVB ( $\gamma_{FVB}$ ) at given temperatures were averaged and plotted.  $\gamma_{FVB}$ s were calculated using Equation 1, as described in the text, comparing every temperature pair that falls within the parameters described in Methods. Specifically, we used temperatures in which <90% but >10% of 129 amplicons were still double-stranded. For example, in this experiment, we used melting data from between 51.9°C and 55.5°C.

melting curve analysis is both highly reproducible and precise and, moreover, is highly accurate in calculating the allelic composition over a wide range of relative fractions.

Note that in these experiments, we use a 1:1 volume mixture of A and B cDNAs as our bias correction reference. This reference is appropriate in these initial studies, where we are seeking to demonstrate that our assay accurately ascertains the relative fractions of A and B alleles added. As described below, an appropriate control for most studies is not a 1:1 volume mixture but a 1:1 molar mixture.

We next wished to determine the ability of our method to resolve the mixing ratio when one allele is overwhelmingly dominant as is often the case with genes that are subject to genomic imprinting, X chromosome inactivation, and allelic exclusion. We therefore prepared incremental cDNA mixtures in which 129 alleles represented only a small fraction of the starting material. After PCR amplification and melting curve analysis, using as references only pure 129, pure FVB, and a 10  $\mu$ L:10  $\mu$ L mixture, we calculated the values shown in Table 2. Again, the calculated  $\gamma_{FVB_C}$  values support the notion that this is a highly precise assay with which to quantitate allele usage even when the samples are highly skewed toward one allele. Together, the re-

sults in Tables 1 and 2 demonstrate that the melting analysis faithfully determines the relative contribution of each allele in a two-allele mixture.

It was also critical to address directly the accuracy of our method using independently derived standards. That is, we needed to demonstrate that the assay not only reflects the relative changes in fractional representation of A and B but also that the absolute values derived for  $\gamma_A$  and  $\gamma_B$  can be correctly ascertained. We therefore constructed several plasmids that each carries a defined number of FVB and 129 cDNA inserts. In other words, we derived several independent samples in which the starting proportions of FVB and 129 cDNAs were known with certainty. We amplified these plasmids and then assayed the amplicon mixture as described above. In this experiment, we used a plasmid (pTG1) that has a single-copy insertion of each cDNA as our bias control. This plasmid provides, by its nature, the equimolar control for the bias correction

and means that the assay will yield fractional values reflecting the real allelic distribution. Our results (Table 3) demonstrate that melting curve analysis accurately and precisely quantitates the relative contributions of each allele in a two-allele mixture.

Lastly, we wished to confirm that these approaches would be generally applicable. We therefore tested the ability of these assays to quantitate allele-specific expression of the mouse *H19* and *Kcnq1* genes, two imprinted genes whose regulation is routinely analyzed in our laboratory. PCR primers and FRET probes spanning naturally occurring SNPs that distinguish *castaneus* and *domesticus* RNAs are described in Methods. Defined mixes of *castaneus* and *domesticus* cDNAs were analyzed as already described for the *Afp* assay. In Table 4, we compare the experimentally calculated frequency for the *domesticus* alleles ( $\gamma_{Dom_C}$ ) with the expected values [ $\gamma_{Dom}(Ex)$ ]. The strong agreement between expected and experimental values demonstrates the ability of this approach to quantitate allele-specific expression of these two genes.

We next wanted to examine allele-specific expression of *H19* and *Kcnq1* in real tissues equivalent to those previously characterized in earlier published studies. Previous reports have demonstrated that *H19* is expressed almost exclusively from the ma-

**Table 1.** DNA melting curves can be used to calculate the fractional value of each allele in a two-allele mixture

	A	B	C	D	E	F	G	H	I	J	K
FVB ( $\mu$ L)	0	2	4	6	8	10	12	14	16	18	20
129 ( $\mu$ L)	20	18	16	14	12	10	8	6	4	2	0
$\gamma_{FVB}(Ex)$	N/A	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	N/A
$\gamma_{FVB}(Ob)$	N/A	0.0942	0.2213	0.3188	0.4170	0.4982	0.5942	0.6922	0.7858	0.8831	N/A
$\gamma_{FVB_C}$	N/A	0.1210	0.2231	0.3212	0.4138	N/A	0.5940	0.6839	0.7767	0.8760	N/A
SD	N/A	0.0080	0.0092	0.0093	0.0087	N/A	0.0049	0.0055	0.0078	0.0047	N/A

FVB and 129 cDNAs were combined in defined proportions to generate mixes A through K. These cDNAs were amplified and the amplicons subjected to melting analysis as described in Figure 2. [ $\gamma_{FVB}(Ex)$ ] The fractional value of FVB cDNA expected based on mixtures generated; [ $\gamma_{FVB}(Ob)$ ] fractional values calculated using Equation 1; ( $\gamma_{FVB_C}$ ) fractional values corrected for biases using Equation 2; (SD) standard deviations ( $N = 4$ ); (N/A) not applicable because these samples represent the reference standards.

**Table 2.** DNA melting curves can be used to calculate the fractional value of each allele in a two-allele mixture where one allele is grossly under-represented

	A	B	C	D	E	F	G	H	I	J	K
FVB ( $\mu$ L)	50	50	50	50	50	10	50	50	50	50	0
129 ( $\mu$ L)	0	1	2	3	4	10	5	6	7	8	50
$\gamma$ FVB(Ex)	0	0.9803	0.9615	0.9433	0.9259	0.5000	0.9090	0.8928	0.8771	0.8620	1
$\gamma$ FVB <sub>c</sub>	N/A	0.9853	0.9627	0.9412	0.9267	N/A	0.8990	0.8819	0.8651	0.8502	N/A
SD	N/A	0.0063	0.0124	0.0091	0.0006	N/A	0.0053	0.0063	0.0062	0.0072	N/A

FVB and 129 cDNAs were combined in defined proportions to generate mixes A through K. These cDNAs were amplified and the amplicons subjected to melting analysis as described. [ $\gamma$ FVB(Ex)] The fractional value of FVB cDNA expected based on mixtures generated; ( $\gamma$ FVB<sub>c</sub>) fractional values corrected for probe bias; (SD) standard deviations ( $N = 3$ ); (N/A) not applicable because these samples represent the reference standards.

ternal allele and that paternal repression is dependent on the presence of the *H19ICR* (also called the *H19DMR*) on the paternal chromosome. In liver RNAs prepared from wild-type mice where the maternal allele is *castaneus* and the paternal allele is FVB *domesticus*, we saw that the paternal chromosome contributes only  $1.5\% \pm 1.1\%$  ( $N = 4$  mice) of the total RNA. However, upon paternal deletion of the *H19ICR*, the paternal contribution is  $39.8\% \pm 0.8\%$  ( $N = 3$  mice), consistent with results using RNase Protection (Thorvaldsen et al. 1998) and by SNUPE (Kaffer et al. 2000; Srivastava et al. 2000) but with considerably greater precision. Expression of *Kcnq1* is also biased toward the maternal chromosome, but imprinting of *Kcnq1* is gradually lost during embryonic development (Gould and Pfeifer 1998). Using RNAs isolated from wild-type *castaneus/domesticus* mice, we found that the maternal chromosome accounted for the following percentages of total *Kcnq1* RNA: e13.5,  $95.8\% \pm 0.3\%$ ; e14.5,  $90.4\% \pm 0.8\%$ ; e15.5,  $77.9\% \pm 1.2\%$ ; e16.5,  $75.1\% \pm 4.1\%$ ; p2,  $61.3\% \pm 3.0\%$ ; and p21,  $58.7\% \pm 0.9\%$  ( $N = 3$  mice). Again, these results are entirely consistent with previous studies that used differential sensitivity to NlaIII digestion to distinguish the two alleles, which were then each quantitated by densitometry scans (Gould and Pfeifer 1998). Note that the precision of results associated with DNA melting analysis is much improved, allowing the investigator the ability to distinguish subtle differences otherwise not discernible, even while the hands-on labor required is drastically reduced. In fact, once the cDNAs are acquired, analysis is complete within  $\sim 1$  h.




## Discussion

It is now clear that numerous loci show strong allele-specific biases in gene expression. Some genes show disparate expression

of the two alleles due to epigenetic mechanisms. Examples include not only genes regulated by genomic imprinting and X-inactivation, but also many biologically important cases where allelic exclusion is random in terms of the choice of allele. Some examples of genes subject to allelic exclusion include those encoding odorant receptors, immunoglobulins, T-cell receptors, interleukins, Natural Killer cell receptors, and p120 catenin (Held et al. 1995; Goldmit and Bergman 2004; Gimelbrant et al. 2005; Shykind 2005). Moreover, it is becoming clear that even among genes that are biallelically expressed, there may be large allelic variations in the RNA levels based on genetic variations between alleles (Yan et al. 2002b; Lo et al. 2003). These variations are presumably associated with natural polymorphisms in critical *cis*-acting elements at these loci. Such differences can affect phenotype and may account for some of the enormous natural variation seen between individuals in outbred populations as well as some experimentally interesting variations seen when crossing inbred lines of laboratory animals.

In this study, we present a methodology for accurate, precise, and reproducible quantitation of allele-specific expression. Our primary model system for this protocol development was the mouse *Afp* gene. In order to study imprinting mechanisms, we have recently established several mouse models in which the Imprinting Control Region (ICR) from the *Igf2/H19* locus on distal chromosome 7 was moved to the nonimprinted *Afp* locus on mouse chromosome 5 (Park et al. 2004). We have already shown that these insertions result in paternal-specific methylation of the mutant chromosomes. However, to determine whether these differential methylation patterns also resulted in parent-of-origin functional differences in *Afp* expression, we needed the ability to accurately quantitate the expression of *Afp* from mutant chromosomes relative to that of the wild-type allele. Because some of the insertions carried subtle mutations in the ICR, we wanted to

**Table 3.** Melting curve analysis can accurately ascertain the fractional representation of each allele in a two-allele mixture

Construct									
	129-FVB			FVB-129-FVB			129-129-129-FVB		
Sample	pTG1	pTG2	pTG3	pGTG1	pGTG2	pGTG3	pTTTG1	pTTTG2	pTTTG3
$\gamma$ FVB (Ex)		0.5000			0.6667			0.2500	
$\gamma$ FVB <sub>c</sub>	N/A	0.5093	0.4945	0.6695	0.6564	0.6542	0.2417	0.2517	0.2627
SD	N/A	0.0012	0.0107	0.0078	0.0036	0.0020	0.0191	0.0059	0.0085

Nine plasmids carrying 129 (filled) and FVB (open) *Afp* cDNA inserts were assayed. The structure of each plasmid is depicted in cartoon form, and the expected fractional value of the FVB allele,  $\gamma$ FVB(Ex), is indicated. After restriction enzyme digestion and PCR amplification, melting analysis was performed using FRET probes Afp-F and Afp-R. Using three reference standards, pure FVB cDNA, pure 129 cDNA, and plasmid pTG1 (a 1:1 mass mixture of FVB and 129 DNAs), the corrected fractional values for FVB DNA,  $\gamma$ FVB<sub>c</sub>, were calculated. (SD) Standard deviation ( $N = 3$ ); (N/A) not applicable.

**Table 4.** DNA melting curves can be used to calculate the fractional value of *domesticus* and *castaneus* alleles of *H19* and *Kcnq1*

	A	B	C	D	E	F	G	H	I	J	K
<i>H19</i>											
Dom ( $\mu$ L)	10	9	8	7	6	5	4	3	2	1	0
Cas ( $\mu$ L)	0	1	2	3	4	5	6	7	8	9	10
$\gamma$ Dom(Ex)	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
$\gamma$ Dom <sub>C</sub>	N/A	0.912	0.788	0.682	0.582	N/A	0.405	0.285	0.203	0.100	N/A
SD	N/A	0.009	0.006	0.011	0.009	N/A	0.009	0.007	0.005	0.002	N/A
<i>Kcnq1</i>											
Dom ( $\mu$ L)	10	9	8	7	6	5	4	3	2	1	0
Cas ( $\mu$ L)	0	1	2	3	4	5	6	7	8	9	10
$\gamma$ Dom(Ex)	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
$\gamma$ Dom <sub>C</sub>	N/A	0.933	0.819	0.707	0.589	N/A	0.410	0.311	0.204	0.103	N/A
SD	N/A	0.006	0.003	0.031	0.006	N/A	0.007	0.002	0.003	0.002	N/A

*Domesticus* (FVB) and *castaneus* cDNAs were combined in defined proportions to generate mixes A through K. These cDNAs were amplified and the amplicons subjected to melting analysis using FRET probes as described in Methods. [ $\gamma$ Dom(Ex)] The fractional value of *domesticus* cDNA expected based on mixtures generated; ( $\gamma$ Dom<sub>C</sub>) experimentally determined fractional values corrected for probe bias; (SD) standard deviations ( $N = 3$ ); (N/A) not applicable as these samples were used as reference controls.

be sure our quantitation could detect subtle changes in allele preferences. Thus, we developed the methodology described in this report.

This new method is simple, fast, and readily adaptable by any molecular biology lab that has access to a real-time PCR system capable of using FRET probes. This method is based on the differential melting behaviors of DNAs that vary at even a single nucleotide. In this technique, the melting of FRET probes is assayed and normalized, and these data are used to calculate the allelic composition through the use of three reference curves. Two of the reference standards are the pure samples of each of the two alleles. The third standard is a 1:1 mixture of the two alleles and corrects for potential bias of the FRET sensor probe for its perfect match. The actual bias we observed varied from probe to probe. For the AFP assay, a true 50:50 mix would appear to be ~60:40 without correction. For other probes tested, the mix would appear something like 52:48 without correction. Due to slight shifts in melting curves between runs, for the most accurate calculations, all three standards need to be included with each sample run.

If the third standard is a 1:1 volume mixture, then one can readily correct for biases in FRET probe hybridization and therefore accurately determine the relative fractional composition of various samples. For example, one can accurately measure that sample 1 has twice the fractional input of allele A as compared to sample 2. This is actually what is being accomplished in most allele-specific RNA assays (albeit usually without quantitation). More useful, however, if the third reference standard is a 1:1 molar mixture, then one is able to correct for probe bias and also accurately determine the absolute fractional value for each of two alleles in the mixture.

Any laboratory can readily obtain a 1:1 molar mixture reference standard. In the experiments presented here, we obtained this control by subcloning the amplicon for each allele into a single plasmid vector. Restriction enzyme digestion released the inserts as separate but equimolar fragments. Alternatively, genomic DNA can be used as the 1:1 molar reference standard. In this case, the PCR primers used to generate the amplicons for melting analysis should be designed so that they are internal to the exon carrying the relevant SNP.

We have tested various types of samples including plasmid DNA and RT-PCR-generated cDNA, and we have varied the initial substrate concentrations over orders of magnitude and in all

cases have obtained reliable and accurate results with the protocol described here (data not shown). However, we have noted that it is quite important to complete the analysis as a two-step process in order to obtain optimal results. After PCR amplification, a part of the sample is removed to an EDTA solution carrying the FRET probes for melting analysis. In contrast, inclusion of the FRET probes in the original PCR reaction gave results with significantly larger standard errors and that deviated more from the expected values.

We have developed a highly functional assay for determining the fractional composition of a two-allele mixture. The accuracy of this methodology is clearly evident in a comparison of expected [ $\gamma$ (Ex)] and calculated ( $\gamma$ <sub>C</sub>) values in Tables 1–4. The reproducibility and precision of the assay is evident in the small standard deviations associated with these calculated  $\gamma$ FVB values. We expect that the extraordinary precision is due both to the inherent advantages of the real-time PCR technology and also to the nature of the algorithm for calculating  $\gamma$ . Because there are multiple N1 and N2 temperature data sets for each melting experiment, multiple values for  $\gamma$ A are calculated and then averaged in each melting experiment. Thus, the final  $\gamma$ A value is a function of multiple measurements taken at various times during the melting experiment and does not rely on any one potentially faulty measurement.

Compared with other methods available for quantitating allele-specific expression, our method becomes most advantageous when one is either assaying many samples or, alternatively, assaying small numbers of samples but in many batches over a period of time. This is because our method does not require the initial investment of money to purchase the two FRET probes. There is also an initial investment of effort to subclone the cDNA fragments that will serve as the reference standards: pure allele A, pure allele B, and a 1:1 molar standard. (Note, however, equivalent standards are required for any quantitative assay.) Once these reagents are generated, our method is particularly inexpensive in terms of labor, money, and time as new RNA samples can be easily analyzed in less than half a day. The approach is technically straightforward and can be readily followed by any molecular biology lab with access to a real-time PCR system. The mathematical analysis for quantitation is easy to use. Our own software is specifically designed to interface with the Roche Light Cycler and is available to the research community at <http://pfeiferlab.nichd.nih.gov>.

In addition to the *Afp*, *H19*, and *Kcnq1* genes described in this report, we have additionally generated assays to quantitate three housekeeping genes (*Ubc*, *G6pdx*, and *Rps27*) and the *H19ICR*, as well as to measure promoter usage at the *Igf2* locus (S.Y. Jeong and K. Pfeifer, unpubl.). In developing these eight assays, we have noted that there is a large flexibility in the design of the primers used to generate the substrates for melting analysis. However, careful design of the FRET probes is critical. The rules we observed in designing FRET probes are straightforward. First, the calculated melting temperature ( $T_m$ ) for the donor probe should exceed the  $T_m$  for the sensor probe by 3°–5°C. Second, the difference in the melting temperatures of the sensor probe annealed to the perfect and the 1-bp mismatch DNAs should be as great as possible. This is accomplished by designing the sensor probe to be an exact match to whichever allele will give the higher calculated melting temperature. Most importantly, to obtain an adequate signal, the FRET probes should not be likely to form any stable hairpins. There has been no apparent restriction on the nature of the SNPs that will work with this assay as we have distinguished A/C, T/C, G/T, C/G, A/T, and A/G polymorphisms.

Our methodology represents a novel use of DNA melting to analyze gene expression. Our methodology allows molecular biologists to accurately measure allele bias in wild-type and mutant cells in such cases as genomic imprinting, X-inactivation, and allelic exclusion. Likewise, this approach offers a simple way to screen multiple samples for allelic bias in order to study the potential effects of such bias in determining phenotype. For example, the role of allele bias in QTL candidate gene expression can be readily ascertained using these approaches (Farrall 2004). In addition, our approach can be extended to quantitate the absolute expression levels of any genes of interest. That is, the level of any nucleic acid of interest can be determined by introducing an artificially generated SNP-carrying competitor of known concentration to a cDNA mix of interest prior to PCR amplification. Given its simplicity, quickness, and accuracy, this DNA melting analysis approach will provide a straightforward paradigm to compare gene expression levels across cell types and developmental stages.

## Methods

### Derivation of Equation 1 to calculate the fractional representation of each allele in a two-allele mixture

In a mixture of dsDNAs A and B, the fraction of the total DNA corresponding to DNA A,  $\gamma_A$ , can be defined as  $\gamma_A = [A]/([A] + [B])$ . Likewise,  $\gamma_B = [B]/([A] + [B])$ . In a mixture of only two DNAs, then  $\gamma_A + \gamma_B = 1$ .

During the melting of this DNA mixture, the fraction of the total DNA that is double-stranded at any temperature N [ $\gamma(\text{dsMix})_N$ ] will be a function of the amount of each of the two components, A and B, that remain double-stranded at this temperature, weighted by the fractional contributions of A and B to the mixture:

$$\gamma(\text{dsMix})_N = \gamma(\text{dsA})_N \times \gamma_A + \gamma(\text{dsB})_N \times \gamma_B \quad (3)$$

The values of dsDNA fractions of the mixture and of the pure A and B DNAs at a given temperature can easily be obtained from the normalized melting curve. Therefore,  $\gamma_A$ , the fractional contribution of DNA A to the mixture, can be readily calculated using the dsDNA fraction values at any two temperature points, N1 and N2, as below,

$$\gamma_A = \frac{\gamma(\text{dsMix})_{N1} \times \gamma(\text{dsB})_{N2} - \gamma(\text{dsMix})_{N2} \times \gamma(\text{dsB})_{N1}}{\gamma(\text{dsA})_{N1} \times \gamma(\text{dsB})_{N2} - \gamma(\text{dsA})_{N2} \times \gamma(\text{dsB})_{N1}} \quad (4)$$

Because there are multiple N1 and N2 data sets in each melting experiment, multiple values for  $\gamma_A$  can be calculated and averaged in each melting analysis, increasing the precision of the calculated  $\gamma_A$ .

Alternatively, since  $\gamma_A + \gamma_B = 1$ ,  $\gamma_A$  can be calculated by substituting  $1 - \gamma_A$  for  $\gamma_B$  in Equation 3:

$$\gamma_A = \frac{\gamma(\text{dsMix})_N - \gamma(\text{dsB})_N}{\gamma(\text{dsA})_N - \gamma(\text{dsB})_N} \quad (5)$$

In sum,  $\gamma_A$ , the fraction of DNA A in the two-part mixture, represents the ratio between the distances from the pure DNA B to the mixture and from the pure DNA A to B on the normalized melting curve at any temperature and can be calculated from the experimentally obtainable values during the DNA melting. When Equation 5 is used,  $\gamma_A$  can be calculated for multiple temperature values (N), corrected for probe bias (as described below), and then averaged to obtain a  $\gamma_A(\text{ob})$  for that sample. Thus, as with Equation 4 above, the final  $\gamma_A$  value for each sample represents multiple experimental data points, thus increasing the precision and reproducibility of the melting curve analysis assay. In several experiments, analysis of melting curves using Equations 4 and 5 have given essentially identical results.

### Derivation of the bias-correction equation (Equation 2)

As described above,  $\gamma_A$  is the fractional component of DNA A in a mixture of DNAs A and B and can be mathematically defined:  $\gamma_A = [A]/([A] + [B])$ . Equation 1 describes algorithms for using the results of DNA melting analysis to calculate  $\gamma_A$  [i.e., to obtain  $\gamma_A(\text{Ob})$ ]. The melting analysis depends on hybridization of two FRET probes to DNAs A and B.  $\gamma_A(\text{Ob})$  will represent the real  $\gamma_A$  insofar as the DNAs A and B hybridize equally well to these FRET probes. Because one of the FRET probes matches perfectly with DNA A but has a 1-bp mismatch with DNA B, we presume that the FRET probe might hybridize better to DNA A and therefore that  $\gamma_A(\text{ob})$  will incorrectly overstate  $\gamma_A$ . In fact, it is more accurate not to represent  $\gamma_A(\text{ob})$  as equal to  $[A]/([A] + [B])$  but, rather, to state that  $\gamma_A(\text{Ob}) = [A^*]/([A^*] + [B^*])$ , where  $[A^*]$  and  $[B^*]$  represent the concentrations of DNAs A and B, respectively, hybridized to FRET probes.

Given the potential for this bias, we use the term  $\gamma_{A_C}$  to represent  $\gamma_A(\text{Ob})$  that has been corrected for probe bias. To correct for probe bias, we can use the 1:1 or half (H) mixture because we can calculate the degree of bias by observed  $\gamma_A(\text{ob})$  for this 1:1 mixture (hereafter called  $\gamma_{A_H}$ ) and compare this value to 0.5, the  $\gamma_A$  of a 1:1 mixture if there is no probe bias. Let us define  $\gamma_{A_H}$  and  $\gamma_{B_H}$  as the observed  $\gamma_A$  and  $\gamma_B$  values, respectively, in a true 1:1 mixture of A and B DNAs. Then the observed  $\gamma_A$  value [ $\gamma_A(\text{Ob})$ ] is related to the observed  $\gamma_{A_H}$  as described:

$$\gamma_A(\text{Ob}) = \frac{[A] \times \frac{\gamma_{A_H}}{0.5}}{[A] \times \frac{\gamma_{A_H}}{0.5} + [B] \times \frac{\gamma_{B_H}}{0.5}} \quad (6)$$

Through algebraic manipulations, Equation 6 can be used to define  $[B]/[A]$ :

$$\frac{[B]}{[A]} = \frac{\gamma_{A_H} - \gamma_A(\text{Ob}) \times \gamma_{A_H}}{\gamma_A(\text{Ob}) \times \gamma_{B_H}} = \frac{\gamma_{A_H} \times (1 - \gamma_A(\text{Ob}))}{\gamma_A(\text{Ob}) \times \gamma_{B_H}} = \frac{\gamma_B(\text{Ob}) \times \gamma_{A_H}}{\gamma_A(\text{Ob}) \times \gamma_{B_H}} \quad (7)$$

Since  $\gamma_{A_C} = [A]/([A] + [B])$  or, alternatively,  $\gamma_{A_C} = 1/(1 + [B]/[A])$ , then substitution for  $[B]/[A]$  using Equation 7 yields the following function, which can be used to correct for bias:

$$\begin{aligned} \gamma_{A_C} &= \frac{[A]}{[A] + [B]} \\ &= \frac{1}{1 + \frac{[B]}{[A]}} \\ &= \frac{1}{1 + \frac{\gamma_{A_H} \times \gamma_{B(Ob)}}{\gamma_{A(Ob)} \times \gamma_{B_H}}} \\ &= \frac{\gamma_{A(Ob)} \times \gamma_{B_H}}{\gamma_{A(Ob)} \times \gamma_{B_H} + \gamma_{A_H} \times \gamma_{B(Ob)}} \end{aligned} \quad (8)$$

Note that the critical experimental decision is what to use as the 1:1 mixture for measuring  $\gamma_{A_H}$ .

### Plasmids

Reference plasmids pcAFP-T (129 cDNA) and pcAFP-G (FVB cDNA) each carry a 236-bp insert including the 5'-UTR of mouse Afp generated by RT-PCR of total liver RNA using primers Afp-4F, 5'-AGATTCTCGGGCCTTTT-3'; and Afp-2R, 5'-GGTATAGA AATCTCATGG-3'. Additional plasmids, carrying multiple insertions of this cDNA fragment (see Table 3) were generated by subcloning.

### Reverse transcription and PCR

Reverse transcription was carried out on 2  $\mu$ g of total liver RNAs using Superscript II (Invitrogen) cDNA synthesis kits following the manufacturer's instructions. Amplicons for melting analysis of Afp were generated starting with ~0.1 ng of DNA template and using 10 pmol of primers Afp-5F (5'-AAGTTGATTTCCAA AACTCG-3') and Afp-2R to generate a 163-bp product. Amplicons for analysis of H19 were generated using primers 5'-GCCTCAAGCACACGGCCACA and 5'-GCACTAAGTCGATTGCACT GG to generate a 164-bp product. Amplicons for analysis of Kcnq1 were generated using primers 5'-CATCGGTGCCCGTCTG AACAG and 5'-TGCTGGGTAGGAAGAGCTCAG to generate a 188-bp product.

### DNA melting analysis

Ten microliters of PCR amplicons was mixed with 10  $\mu$ L of 10 mM EDTA and with 4 pmol of FRET probes, Afp-R and Afp-F (see Fig. 1B). DNA melting was performed using a Roche LightCycler under the following conditions: 30 sec at 95°C, 8 min at 49°C, and 0.2°C stepwise increases in temperature until 65°C was achieved. Assay results are essentially insensitive to the rate of temperature increase. However, it is important to measure fluorescence after stepwise increases and not during a continuous increase. The background-adjusted fluorescence values (F2/F1) from the Red640 dye were extracted using the LightCycler3 Data Analysis (Version 3.5.28) and then submitted to our analysis program. Due to the likelihood that the very small differences in the normalized values between samples and references at the beginning and at the ending of the melting curves would not be well resolved by the machine, only the values at the temperatures in a range defined by dsDNA fractions from 5% to 90% of the fastest melting reference (129 cDNA) were used for subsequent analysis. Within this range, all temperature pairs with a  $\Delta T$  of  $\geq 1^\circ\text{C}$  were subjected to calculation to yield a fractional component for that sample. The fractional values for all the temperature pairs were then averaged to yield  $\gamma_{DNA(Ob)}$  (observed fractional

value) for that sample. These  $\gamma_{DNA(Ob)}$  values were then corrected using the fraction value of the 1:1 mixture to yield the bias-corrected fraction value,  $\gamma_{DNA_C}$ . When the 1:1 mixture is a true 1:1 mass mixture of the two alleles, this  $\gamma_{DNA_C}$  is an accurate representation of the true allelic fraction. We have developed software to directly interface with data acquired by the Roche Light Cycler and will readily share this or advise on interfaces with other equipment.

For analysis of H19 gene expression, FRET probes 5'-CCTCTTGGCAATGCTGCCCA-Fluorescein and 5'-LC Red640-ACCCACCTGTCGTCATCTCCG-C3 Blocker were used as described above. The underscored G is an A residue in *domesticus* animals. For analysis of Kcnq1 gene expression, we used FRET probes 5'-GTGATCATCACAGACATGCTCCACCAGC-Fluorescein and 5'-LC Red640-GCTGTCCGTCGAACAAGGTG-C3 Blocker. The underscored C is an A residue in *domesticus* animals. All FRET probes were obtained from Idaho Technology, Inc.

### Acknowledgments

This work was supported by the NICHD Intramural Research Program. The NICHD ACUC approved all animal experimentation. We thank Claudia Gebert for many helpful discussions, especially in the design of FRET probes; John Van Druff for construction of the Kcnq1 reference plasmids; and Lamar Roos for advice on FRET probe design.

### References

- Farrall, M. 2004. Quantitative genetic variation: A post-modern view. *Hum. Mol. Genet.* **13**: R1–R7.
- Gimelbrant, A.A., Ensminger, A.W., Qi, P., Zucker, J., and Chess, A. 2005. Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. *J. Biol. Chem.* **280**: 1354–1359.
- Goldmit, M. and Bergman, Y. 2004. Monoallelic gene expression: A repertoire of recurrent themes. *Immunol. Rev.* **200**: 197–214.
- Gould, T.D. and Pfeifer, K. 1998. Imprinting of mouse *Kvlt1* is developmentally regulated. *Hum. Mol. Genet.* **7**: 483–487.
- Held, W., Roland, J., and Raulat, D.H. 1995. Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. *Nature* **376**: 355–358.
- Kaffer, C.R., Srivastava, M., Park, K., Ives, E., Hsieh, S., Batlle, J., Grinberg, A., Huang, S.P., and Pfeifer, K. 2000. A transcriptional insulator at the imprinted *H19/Igf2* locus. *Genes & Dev.* **14**: 1908–1919.
- Knight, J.C., Keating, B.J., Rockett, K.A., and Kwiatkowski, D.P. 2003. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat. Genet.* **33**: 469–475.
- Kuppuswamy, M.N., Hoffmann, J.W., Kasper, C.K., Spitzer, S.G., Groce, S.L., and Bajaj, S.P. 1991. Single nucleotide primer extension to detect genetic diseases: Experimental application to hemophilia B (factor IX) and cystic fibrosis genes. *Proc. Natl. Acad. Sci.* **88**: 1143–1147.
- Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H., and Lee, M.P. 2003. Allelic variation in gene expression is common in the human genome. *Genome Res.* **13**: 1855–1862.
- Park, K., Sellars, E., Grinberg, A., Huang, S., and Pfeifer, K. 2004. The H19 differentially methylated region marks the parental origin of a heterologous locus without gametic DNA methylation. *Mol. Cell. Biol.* **24**: 3588–3595.
- Pastinen, T. and Hudson, T.J. 2004. Cis-acting regulatory variation in the human genome. *Science* **306**: 647–650.
- Shykind, B.M. 2005. Regulation of odorant receptors: One allele at a time. *Hum. Mol. Genet.* **14**: R33–R39.
- Srivastava, M., Hsieh, S., Grinberg, A., Williams-Simon, L., Huang, S.P., and Pfeifer, K. 2000. H19 and *Igf2* monoallelic expression is regulated in two distinct ways by a shared cis acting element. *Genes & Dev.* **14**: 1186–1195.
- Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. 1998. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and *Igf2*. *Genes & Dev.* **12**: 3693–3702.
- Ueda, H., Howson, J.M., Esposito, L., Heward, J., Snook, H.,

- Chamberlain, G., Rainbow, D.B., Hunter, K.M., Smith, A.N., Digenova, et al. 2003. Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. *Nature* **423**: 506–511.
- Winter, E., Yamamoto, F., Almoguera, C., and Perucho, M. 1985. A method to detect and characterize point mutations in transcribed genes: Amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells. *Proc. Natl. Acad. Sci.* **82**: 7575–7579.
- Yan, H., Dobbie, Z., Gruber, S.B., Markowitz, S., Romans, K., Giardiello, F.M., Kinzler, K.W., and Vogelstein, B. 2002a. Small changes in expression affect predisposition to tumorigenesis. *Nat. Genet.* **30**: 25–26.
- Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. 2002b. Allelic variation in human gene expression. *Science* **297**: 1143.

*Received October 11, 2006; accepted in revised form April 10, 2007.*





## Accurate quantitation of allele-specific expression patterns by analysis of DNA melting

Sangkyun Jeong, Yoonsoo Hahn, Qi Rong, et al.

*Genome Res.* 2007 17: 1093-1100 originally published online June 1, 2007

Access the most recent version at doi:[10.1101/gr.6028507](https://doi.org/10.1101/gr.6028507)

---

### Supplemental Material

<http://genome.cshlp.org/content/suppl/2007/06/04/gr.6028507.DC1>

### References

This article cites 18 articles, 10 of which can be accessed free at:  
<http://genome.cshlp.org/content/17/7/1093.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

Affordable, Accurate  
Sequencing.



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>