Metadata of the Book that will be visualized online

Book Title	Molecular Methods for Evolutionary Genetics	
Book SubTitle		
Copyright Year	2012	
Copyright Holder	Springer Science+Business Media, LLC	

Corresponding Author	Family Name	**Wittkopp**
	Particle	
	Given Name	**Patricia J.**
	Suffix	
	Division	
	Organization	University of Michigan
	Address	1061 Natural Science Building, 830 North University Avenue, 48109-1048, Ann Arbor, MI, USA
	Email	wittkopp@umich.edu

Author's Proof

Chapter 18 1

Using Pyrosequencing to Measure Allele-Specific mRNA Abundance and Infer the Effects of *Cis*- and *Trans*-regulatory Differences 2 3 4

Patricia J. Wittkopp 5

Uncorrected Proof

Abstract 6

Changes in gene expression are an important source of phenotypic differences within and between species. 7 Differences in RNA abundance can be readily quantified between genotypes using a variety of tools, 8 including microarrays, quantitative real-time PCR, cDNA sequencing, and in situ hybridization, but 9 determining the genetic basis of heritable expression differences has historically been less straightforward. 10 Genetic changes that affect RNA abundance can be broadly classified into two groups depending on how 11 they affect gene expression: *cis*-acting changes affect expression of a single allele in a diploid cell and are 12 typically located close to the affected gene in the genome, whereas *trans*-acting changes affect expression 13 of both alleles of a gene in a diploid cell and can be located virtually anywhere within a genome. By com- 14 paring relative expression of two alleles in an F_1 hybrid with relative expression between the two parental 15 genotypes, the net effects of *cis*- and *trans*-acting changes can be discerned. Here, I describe how pyrose- 16 quencing can be used to obtain relative gene-specific and allele-specific expression. I also describe how 17 such data can be used to infer the relative contribution of *cis*- and *trans*-acting changes to expression 18 differences between genotypes. 19

Key words: Gene regulation, Allelic imbalance, Gene expression, Pyrosequencing, Quantitative, 20 mRNA transcripts, Divergence, Polymorphism, Evolution 21

1. Introduction 22

Quantifying allele-specific expression is a powerful way to examine 23 the regulation of gene expression and determine how this process 24 differs between individuals and between species. For example, mea- 25 surements of allele-specific expression can be used to test for 26 genomic imprinting (e.g., (1, 2)), for expression divergence between 27 paralogous genes (e.g., (3)), and for differences in *cis*- and *trans*- 28 regulation within and between species (e.g., (4–8)). Pyrosequencing 29

Virginie Orgogozo and Matthew V. Rockman (eds.), *Molecular Methods for Evolutionary Genetics*,
Methods in Molecular Biology, vol. 772, DOI 10.1007/978-1-61779-228-1_18, © Springer Science+Business Media, LLC 2011

Author's Proof

Uncorrected Proof

30 is a time- and cost-effective technique for quantifying allele-specific expression of single genes in a high-throughput manner. Here, I describe the protocols used in my laboratory for obtaining measurements of relative gene expression between two genotypes and relative allelic expression between two alleles within a single heterozygous genotype. These methods have been shown to produce estimates of relative gene expression that are consistent with microarrays (2), quantitative reverse transcriptase PCR (qRT-PCR) (2), and (most recently) cDNA sequencing (9).

Pyrosequencing is a method of DNA sequencing that couples primer extension with light production (10): each time a nucleotide is added to a 3′ end of DNA by DNA polymerase, a constant amount of light is produced. This light is captured by a camera monitoring each well of a 96-well plate and the amount of light produced in each well is recorded in a "pyrogram" (a sample of which is shown in Fig. 3a). The height of "peaks" observed in the pyrogram provides a quantitative readout of the number of nucleotides added to 3′ ends of DNA within a well at a particular time. This quantitative readout of primer extension makes pyrosequencing well suited not only for genotyping individuals but also for quantifying relative abundance in samples of DNA (or cDNA) containing more than one allele.

Pyrosequencing has many applications; however, this chapter focuses solely on using pyrosequencing to measure differences in transcript abundance between genotypes and between different alleles in heterozygous cells. Such data can be used to distinguish between *cis*- and *trans*-acting changes that contribute to expression differences between genotypes (6). Total expression differences are measured between two genotypes, and then relative allelic expression is measured in F_1 hybrids produced by crossing the two genotypes. These F_1 hybrids carry one *cis*-regulatory allele from each of the parental genotypes within the same cell, providing a direct readout of relative *cis*-regulatory activity because the two alleles are exposed to a common *trans*-regulatory environment. Any differences in allelic expression within these heterozygous genotypes must, therefore, be caused by differences in *cis*-regulatory activity (4). Any expression difference between parental genotypes for that gene that is in excess of the *cis*-regulatory difference observed in the F_1 hybrids is attributed to changes in *trans*-regulation (6). Differences in the activity or abundance transcription factors as well as differences in the environment and/or abundance of cell types will all be detected as *trans*-regulatory effects with this assay. The methods used to obtain such measures of transcript abundance are described below, divided into the following sections: (a) designing a pyrosequencing assay, (b) extraction of genomic DNA and RNA, (c) cDNA synthesis, (d) pyrosequencing, (e) experimental design, and (f) data analysis and interpretation.

Author's Proof

2. Materials

2.1. Designing a Pyrosequencing Assay

1. At least 50 bp of aligned, transcribed, exonic sequence from two alleles of each gene (or exon) of interest, containing at least one single-nucleotide polymorphism (SNP).
2. PyroMark Assay Design Software 2.0 (Qiagen). This software is recommended, but not essential (see step 3 in Subheading 3.1).

2.2. Extraction of Genomic DNA and Total RNA from Biological Tissue

1. Tissue samples to be analyzed. This may include samples from two (preferably inbred) genotypes and/or samples from a single heterozygous genotype. For the inference of *cis*- and *trans*-regulatory differences between genotypes, one sample should contain an approximately equal mix of tissue from the two genotypes of interest and another should contain tissue from F_1 heterozygotes produced by crossing the two genotypes of interest.
2. SV Total RNA Isolation Kit (Promega) containing: Spin Columns, Waste Collection tubes, Elution tubes, RNA Lysis Buffer, RNA Dilution Buffer, β-mercaptoethanol (BME, 48.7%), DNase I (lyophilized), $MnCl_2$, 0.09 M, Yellow Core Buffer, DNase Stop Solution, RNA Wash Solution, and Nuclease-Free Water.
3. 95% ethanol, prepared with RNAse-free water.
4. 70% ethanol, chilled to −20°C, does NOT need to be prepared with RNAse-free water.

2.3. cDNA Synthesis

1. RNA from tissue(s) of interest.
2. RNAse-free DNAse (Promega).
3. Primer for reverse transcription (500 μg/ml). A "poly-T" primer such as 5′-TTTTTTTTTTTTTTTTT-3′ or 5′-TTTT TTTTTTTTTTTTVN-3′ is most often used; however, a gene-specific primer could be used if the researcher is only interested in assaying expression of a single gene.
4. dNTP mix containing dATP, dCTP, dTTP, and dGTP at 10 mM each.
5. RNAsin (Promega).
6. Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV-RT, Promega), which is shipped with 5× First-Strand Buffer [250 mM Tris–HCl, pH 8.3 at room temperature, 375 mM KCl, 15 mM $MgCl_2$] and 0.1 M DTT. SuperScript™ II Reverse Transcriptase with its associated buffer (Invitrogen) has also been used with success.

Uncorrected Proof

Author's Proof

117 ### 2.4. Pyrosequencing

1. Nucleic acid template (genomic DNA or cDNA).
2. 10× PCR buffer: 100 mM Tris–HCl, 500 mM KCl, 15 mM $MgCl_2$, pH 8.3 at 25°C.
3. dNTP mix containing dATP, dCTP, dTTP, and dGTP at 1.5 mM each.
4. 10 μM solution of biotinylated PCR primer, see (1) in Fig. 1.
5. 10 μM solution of nonbiotinylated PCR primer, see (2) in Fig. 1.
6. *Taq* or equivalent thermostable DNA polymerase.
7. Sterile water.
8. Glassware for solutions, rinsed with deionized water to eliminate residual phosphate from detergent.
9. Pyrosequencing primer (100 μM), see (3) in Fig. 1, for each assay.
10. Streptavidin Sepharose, High Performance (GE Healthcare).
11. 70% ethanol.
12. 4 M hydrochloric acid and 1 M acetic acid for pH adjustments.
13. Binding Buffer: 10 mM Tris–HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20, pH 7.6.
14. Denaturing Buffer: 0.2 M NaOH.
15. Wash Buffer: 10 mM Tris–acetate, pH 7.6.
16. Annealing Buffer: 20 mM Tris–acetate, 2 mM Mg–acetate, pH 7.6.
17. PyroMark Q96 Vacuum Prep Troughs (Qiagen).
18. PyroMark Q96 Plate Low (Qiagen).
19. PyroMark Q96 Cartridge (Qiagen).
20. Pyro Gold Reagents or Pyro Gold SQA Reagents (Qiagen), containing: Enzyme mixture (DNA polymerase, ATP-sulfurylase, luciferase, and apyrase), Substrate mixture (luciferin, adenosine 5′ phosphosulfate), Nucleotides (dATP αS, dCTP, dGTP, and dTTP, separately).

Uncorrected Proof

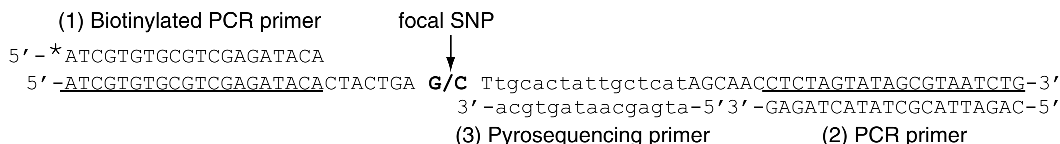


Fig. 1. Overview of primers and sequence components of a pyrosequencing assay. Three primers are required for each pyrosequencing assay: two primers (1 and 2) that are used for PCR amplification of a region including the focal SNP (*arrow*) and an internal primer that is used for pyrosequencing (3). The PCR primer (2) that is on the opposite strand from the pyrosequencing primer (3) is labelled with a biotin molecule (*asterisk*) at its 5′ end.

Author's Proof

21. PyroMark Vacuum Prep Tool (Qiagen).
22. PyroMark Vacuum Prep Tool Filter Probes (Qiagen).
23. PyroMark Q96 ID machine (Qiagen), or a predecessor such as the PSQ 96 machine (Biotage), with accompanying operating software and computer.

3. Methods

3.1. Designing a Pyrosequencing Assay

Pyrosequencing determines the relative abundance of alternative mRNA transcripts by using a SNP to discriminate between two alleles. Selection of the SNP to be assayed is, therefore, critical to the success of the experiment. Pyrosequencing assay design begins with at least 50 bp of transcribed, exonic sequence from each allele for a gene of interest (see Note 1). If not already available, allele-specific sequences can typically be obtained with traditional Sanger sequencing of PCR products produced from genomic DNA or cDNA. Using these sequences, three primers must be designed for each pyrosequencing assay, all centered around the chosen focal SNP (Fig. 1): two primers will be used to amplify a 50–200-bp region of sequence containing the focal SNP via PCR and the third primer will be annealed within 1–3 bp of the focal SNP and used for pyrosequencing.

1. Align sequences from the alleles to be compared and identify all SNPs and other sequence differences. This can easily be done in most sequence analysis programs.
2. Examine each SNP to identify the one(s) that are best suited to pyrosequencing, considering the following criteria:
   (a) At least 15 bp of sequence immediately adjacent to the SNP on at least one side must be identical between the two alleles. There must also be two regions of identical exonic sequence flanking the focal SNP that are located 50–200 bp apart, each at least 20 bp long, to allow for annealing of PCR primers (Fig. 1). To allow the direct comparison of amplicons from genomic DNA and cDNA, the amplified fragment (containing the focal SNP) should come from a single exon.
   (b) Try to avoid SNPs that are part of homopolymers, which are strings of two or more consecutive, identical nucleotides (e.g., GG or TTT), in either allele. All consecutive bases of the same type are incorporated at the same time in a pyrosequencing reaction, and the accuracy of such aggregate peaks is lower than peaks reflecting the incorporation of a single base in each sequence. If homopolymers cannot be avoided completely, choose a SNP in which one allele

Uncorrected Proof

Author's Proof

forms a homopolymer not more than two nucleotides long (e.g., G/TG, where G/T are the alternate alleles of the SNP and the following G is present in both alleles) (see Note 2).

(c) Give highest priority to G/C SNPs, next highest priority to G/A, G/T, C/A, and C/T SNPs, and lowest priority to A/T SNPs. This is because the incorporation of dATP during pyrosequencing (which cannot be avoided for A/T SNPs) produces more light than the incorporation of dTTP, dGTP, or dCTP, requiring the use of a correction factor (see Note 3) that can increase the variance of replicate measurements.

(d) If a "poly T" primer will be used for the production of cDNA, give preference to SNPs located near the 3′ end of the transcribed region to maximize the likelihood that the focal SNP is contained within each cDNA molecule synthesized from the mRNA of interest.

3. Design PCR and pyrosequencing primers to interrogate focal SNP. The easiest way to do this is to use the PyroMark Assay Design Software 2.0 (http://www.pyrosequencing.com/DynPage.aspx?id=7257) with the "allele quantification" option selected. This software allows multiple assays to be designed in a high-throughput, "batch processing" manner. If this proprietary software is not available, however, the three primers required for each pyrosequencing assay can be designed manually using a freely available primer design program such as Primer3 (11). Care must be taken to check all primers and the PCR product for hairpins and complementary sequences within and between primers that could cause unintended nucleotide bases to be added to 3′ ends of DNA during pyrosequencing. Because the pyrosequencing reaction occurs at 28°C complementary sequence of as few as four 3′ bases of DNA can prime DNA synthesis. Such extension introduces background signal that can interfere with the accurate quantification of alleles at the focal SNP. Additional tips for designing primers well suited for pyrosequencing are available from the technical support division of Qiagen.

4. Synthesize the three primers designed for each assay, adding a biotin molecule to the 5′ end of the PCR primer that anneals to the opposite strand from the pyrosequencing primer (Fig. 1). Most commercial suppliers of oligonucleotides (e.g., Integrated DNA Technologies) will add a 5′ biotin molecule to an oligonucleotide for an additional fee.

Uncorrected Proof

Author's Proof

3.2. Extraction of Genomic DNA and Total RNA from Biological Tissue

This chapter specifically describes analysis of the two types of biological samples required to infer the relative effects of *cis*- and *trans*-regulatory variation: a mixed sample of tissue from two highly inbred “parental” genotypes and a tissue sample from the heterozygous F_1 hybrid genotype produced by crossing the two parental genotypes together (see Note 4). For the mixed parental tissue sample, and in any case where expression is compared between two genotypes, similar amounts of tissue from the two genotypes should be used. For the hybrid tissue sample, and in any case where only a single genotype is examined, all of the tissue is derived from the same (heterozygous) genotype. Genomic DNA and total RNA are extracted from each tissue sample. Genomic DNA and RNA should always be extracted from the same homogenate of a mixed tissue sample to allow the relative abundance of both cells and RNA from each genotype to be determined. The extraction of genomic DNA and RNA from the same tissue homogenate is not essential for samples containing only a single genotype, but is recommended for consistency if comparing to mixed tissue samples. Regardless of extraction method, pyrosequencing analysis of genomic DNA from heterozygous cells is critical to measure and control for any inequality in PCR amplification and/or pyrosequencing detection of the two alleles (5, 6).

A protocol for extracting genomic DNA and total RNA from a single tissue homogenate is described below. It was developed for use with the SV Total RNA isolation kit from Promega (Madison, WI), and was modified from (12). It has been tested most extensively with *Drosophila* (fruit fly) species, but should also be suitable for tissue from most multicellular eukaryotes. The protocol is divided into three sections: tissue homogenization (steps 1–8), RNA extraction (steps 9–22), and genomic DNA extraction (steps 23–29). The genomic DNA and RNA recovery can proceed in parallel; however, if they are processed in series, RNA recovery (steps 9–22) should be completed first to minimize the likelihood of degradation.

1. Collect tissue, freeze in liquid nitrogen, and store at −80°C until ready for extraction.
2. Homogenize ~15 mg of tissue (see Note 5) in 175 μl SV RNA Lysis Buffer with BME added. Add 350 μl of SV RNA Dilution Buffer and vortex at least for 1 min.
3. Centrifuge for 10 min on maximum speed at room temperature.
4. Transfer supernatant to RNAse-free microcentrifuge tube and add 75 μl of 95% ethanol (RNAse-free).
5. Mix by inverting ten times.
6. Transfer to SV Total RNA Spin Column Assembly and let stand for 5 min at room temperature. Genomic DNA binds to the column resin during this step.

Uncorrected Proof

Author's Proof

280 7. Centrifuge for 1 min on maximum speed at room temperature.
281 8. Transfer flow-through, which contains RNA and proteins, to a
282 clean RNAse-free microcentrifuge tube.
283 9. Add 300 μl of RNAse-free 95% ethanol to the flow-through
284 from step 8 above.
285 10. Mix gently by inversion.
286 11. Transfer solution to a *new* SV Total RNA Spin Column
287 Assembly and let stand for 1 min at room temperature. RNA
288 binds to the column resin during this step.
289 12. Centrifuge on high speed for 1 min. Discard the flow-through.
290 13. Apply 600 μl of SV RNA Wash solution to the column.
291 Centrifuge on high for 1 min. Discard the flow-through.
292 14. Mix 40 μl of Yellow Core buffer, 5 μl of 0.09 M $MnCl_2$, 5 μl
293 of DNAse (in this order), and pipette mixture onto the col-
294 umn. Be sure to cover column surface completely.
295 15. Let stand for 15 min at room temperature.
296 16. Add 200 μl of SV DNAse Stop Solution, let stand at room
297 temperature for 1 min, and then centrifuge for 1 min on high
298 speed. Discard the flow-through.
299 17. Apply 600 μl of SV RNA Wash solution to the column.
300 Centrifuge on high for 1 min. Discard the flow-through.
301 18. Apply 250 μl of SV RNA Wash solution to column. Centrifuge
302 on high for 1 min. Discard the flow-through.
303 19. Centrifuge again for 1 min to remove any residual wash solu-
304 tion. Discard the flow-through.
305 20. Place Spin Column into sterile Elution tube. Add 100 μl of
306 Nuclease-free water to column, covering surface completely.
307 Let stand at room temperature for 5 min. Centrifuge on high
308 for 1 min. This eluate contains total RNA.
309 21. Quality and quantity can be crudely examined using agarose
310 gel electrophoresis. Strong bands corresponding to 18S and
311 28S rRNA should be clearly visible along with faint a "smear"
312 of RNA of other lengths.
313 22. RNA samples should be stored at –80°C until used for cDNA
314 synthesis to avoid degradation.
315 23. Add 700 μl of cold 70% ethanol to the Spin Column used for
316 steps 1–8. Let stand for 1 min at room temperature. Centrifuge
317 on high for 1 min. Discard the flow-through.
318 24. Repeat previous step.
319 25. Centrifuge for 1 min on high to remove any residual ethanol.
320 Discard the flow-through.

Uncorrected Proof

Author's Proof

26. Place Spin Column into a sterile Elution tube. Add 100 μl of nuclease-free water to the Spin Column. Let stand at room temperature for 5 min. (Heating the column to 55–60°C while standing increases recovery, but can also increase degradation.) Centrifuge on high for 1 min. This eluate contains the genomic DNA. 321 322 323 324 325 326
27. Repeat previous step and combine eluates. 327
28. Quality and quantity of genomic DNA can be crudely examined using agarose gel electrophoresis. A large, single band should be most prominent. 328 329 330
29. Genomic DNA should be stored at either −20°C or −80°C until needed. 331 332

3.3. cDNA Synthesis

Prior to pyrosequencing, the RNA must be converted into (single- or double-stranded) cDNA. Any cDNA synthesis protocol can be used for this purpose. "Poly T" primer, which will allow cDNA to be synthesized from all polyadenylated mRNA, or one or more gene-specific primers can be used depending on the number of genes that will be examined. A DNAse treatment of the RNA template immediately prior to cDNA synthesis is strongly recommended to remove any genomic DNA that survived the DNAse treatment administered during RNA extraction. 333 334 335 336 337 338 339 340 341

1. Combine 8.0 μl (~1.5 μg) total RNA extracted as described in Subheading 3.2, 8.4 μl of 5× First-Strand Buffer, and 2.0 μl of DNAse. Incubate at 37°C for 1 h. 342 343 344
2. Heat to 65°C for 15 min to inactivate the DNAse. 345
3. Add 5.4 μl (500 μg/ml) of primer for reverse transcription and slowly cool to 37°C for over 10 min. 346 347
4. Add 4.2 μl of dNTP mix (10 mM per nucleotide), 1 μl of RNAsin, and 1.2 μl of M-MLV-RT and incubate at 37°C for 1 h. (Inclusion of 3 μl of DTT is optional.) 348 349 350
5. Dilute as appropriate. We typically add 69.8 μl of nuclease-free water to dilute the 30.2 μl of cDNA synthesis reaction to 100 μl total. For most genes, 1–2 μl of this diluted single-stranded cDNA is sufficient to produce a strong PCR product suitable for pyrosequencing (see below). For lowly expressed genes, increasing the amount of cDNA used may help to achieve a strong PCR product. 351 352 353 354 355 356 357

3.4. Pyrosequencing

Pyrosequencing involves three major steps: PCR amplification of sequence containing the focal SNP, recovery of single-stranded PCR product with the pyrosequencing primer annealed, and pyrosequencing, which is the real-time monitoring of a controlled primer extension reaction by a machine such as the PyroMark Q96 ID (Qiagen) or PSQ 96 (Biotage). 358 359 360 361 362 363

Uncorrected Proof

Author's Proof

364 365 366 367 368 369 370 371

3.4.1. PCR Amplification of Sequence to Be Analyzed

Virtually any PCR protocol that produces a single amplified product with no evidence of primer dimers or residual primers can be used for pyrosequencing. Each pair of PCR primers may require some optimization (e.g., by changing the annealing temperature, extension time, and/or amount of DNA or cDNA template) to obtain a single amplified product that produces enough material for pyrosequencing. Starting conditions used for testing each new pair of PCR primers for pyrosequencing in my laboratory follow:

372 373 374 375 376 377 378 379 380 381

- 2 μl of nucleic acid template (i.e., genomic DNA or cDNA, ~0.1–10 ng).
- 5 μl of 10× PCR buffer (including 15 mM $MgCl_2$).
- 5 μl of dNTP mix (i.e., dATP, dCTP, dTTP, and dGTP, each at 1.5 mM).
- 2 μl of 10 μM biotinylated PCR primer, (1) in Fig. 1.
- 2 μl of 10 μM nonbiotinylated PCR primer, (2) in Fig. 1.
- 0.2 μl of *Taq* polymerase (5 units/μl) (or equivalent thermostable DNA polymerase).
- 33.8 μl of sterile water.

382 383 384 385 386 387 388 389 390 391

This 50-μl reaction is subject to 35–50 cycles of 94°C for 30 s, 55–68°C (depending on primer sequences) for 30 s, and 72°C for 15–60 s (depending on length of PCR product). These cycles are preceded by heating the reaction to 94°C for 3 min and followed by cooling the reaction to 4°C. The amount of PCR product is assessed by running 5 μl of the completed PCR reaction on an agarose gel. An easily visible band containing at least 30 ng of DNA should be observed. PCR products must be arrayed in a 96-well plate format for pyrosequencing, and it is easiest to simply perform the PCR reactions in a 96-well plate.

392 393 394 395 396 397 398 399 400 401 402 403 404 405 406

3.4.2. Recovery of Single-Stranded PCR Product with Pyrosequencing Primer Annealed

Prior to pyrosequencing, the double-stranded PCR product must be denatured and the strand that is not complementary to the pyrosequencing primer removed. The pyrosequencing primer is then annealed to the remaining, complementary strand. Recovery of the appropriate strand is mediated by the interaction between biotin (which is located at the 5′ end of the appropriate PCR primer (Fig. 1)) and streptavidin-coated sepharose. The most efficient way to process a 96-well plate of samples is to use the PyroMark Vacuum Prep Tool (Qiagen), as described below. This hand-held vacuum device is fitted with 96 PyroMark Vacuum Prep Tool Filter Probes, which allow liquids, but not the streptavidin sepharose, to pass through. A protocol for using this tool to prepare a 96-well plate of PCR products for pyrosequencing follows. An alternative protocol using magnetic beads instead of the PyroMark Vacuum Prep Tool is available from Qiagen.

Uncorrected Proof

Author's Proof

1. Preheat a heating block that holds a 96-well plate to 90°C. 407
2. Add 3 μl of streptavidin sepharose and 40 μl of Binding Buffer to each PCR product (which should be 40–45 μl after analysis of 5 μl by gel electrophoresis), and cover tightly (see Note 6). 408 409 410
3. Vortex the plate containing PCR products and streptavidin sepharose at maximum speed for 10 min at room temperature to maximize binding between the biotinylated strands of PCR products and the streptavidin sepharose. 411 412 413 414
4. While the plate is shaking, do the following: 415
   (a) Fill five PyroMark Q96 Vacuum Prep Troughs with the following solutions, arranging the trays from left to right in the following order: deionized water, 70% ethanol, Denaturing Buffer, Wash Buffer, and deionized water. 416 417 418 419
   (b) Prepare a PyroMark Q96 Plate Low (which is a specially sized 96-well plate) for pyrosequencing by adding 40 μl of 0.5 μM pyrosequencing primer diluted in Annealing Buffer (20 mM Tris–Acetate, 2 mM Mg–Acetate, pH 7.6) to each well. The pyrosequencing primer in each well should match the PCR product in the corresponding well of the original PCR plate. 420 421 422 423 424 425 426
   (c) Add 620 μl of water to the Pyro Gold (or Pyro Gold SQA) Reagent vial labelled "Enzyme". Mix thoroughly by shaking. 427 428 429
   (d) Add 620 μl of water to the Pyro Gold (or Pyro Gold SQA) Reagent vial labelled "Substrate". Mix thoroughly by shaking. 430 431 432
   (e) Pipet all 620 μl of the resuspended "Enzyme" and "Substrate" solutions as well as 200 μl of each nucleotide into the PyroMark Q96 Cartridge as shown in Fig. 2 (see Note 7). 433 434 435 436
   (f) Allow solutions to warm completely to room temperature. 437

Uncorrected Proof

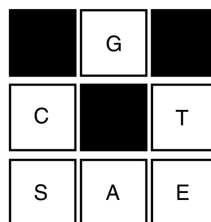


Fig. 2. Placement of enzyme (E), substrate (S), and nucleotides (A, C, G, T) in the PyroMark Q96 Cartridge. Arrangement shown is looking down at the top of the cartridge with the label closest to the researcher, as indicated.

Author's Proof

5. Remove the plate containing PCR products and streptavidin beads from the vortex, remove plate cover carefully to avoid contamination of neighboring wells. Turn on vacuum so that air is being drawn through the PyroMark Vacuum Prep Tool Filter Probes of the PyroMark Vacuum Prep Tool. Place in the left-most trough containing water for 30 s, drawing water through the pins.

   (a) Move the PyroMark Vacuum Prep Tool into the 96-well plate containing the PCR products and streptavidin beads for 2–3 s, moving it up and down. The goal is to capture all of the beads from each well on the bottom of a Vacuum Prep Tool Filter Probe.

   (b) Move the PyroMark Vacuum Prep Tool into the second trough (ii) containing 70% ethanol for 5–10 s.

   (c) Move the PyroMark Vacuum Prep Tool into the third trough (iii) containing Denaturing Buffer for 5–10 s.

   (d) Move the PyroMark Vacuum Prep Tool into the fourth trough (iv) containing Wash Buffer for 5–10 s.

   (e) Remove the PyroMark Vacuum Prep Tool from the fourth trough and (after all of the liquid has been pulled away from the sepharose), *turn off the vacuum*, holding the PyroMark Vacuum Prep Tool in the air (ideally, just above the PyroMark Q96 Plate Low). Make sure the suction has stopped completely prior to proceeding with the next step.

   (f) Carefully place the PyroMark Vacuum Prep Tool into the PyroMark Q96 Plate Low containing the pyrosequencing primers in annealing buffer. Be sure to line up the pins and wells correctly such that the beads originally collected from well A1 in the PCR plate match up with the pyrosequencing primer for well A1 in the PyroMark Q96 Plate Low.

   (g) Gently swirl the PyroMark Vacuum Prep Tool in the PyroMark Q96 Plate Low to shake off sepharose.

   (h) Place the PyroMark Q96 Vacuum Prep Tool into the fifth trough (v) containing water and shake to further remove any remaining sepharose.

   (i) Transfer the PyroMark Q96 Plate Low to a heat block at 90°C and let it stand for 2 min.

   (j) Remove the PyroMark Q96 Plate Low from the heating block and allow to cool slowly to room temperature (10–15 min). This plate contains the single-stranded PCR product with the pyrosequencing primer annealed and is ready to be used for pyrosequencing.

Uncorrected Proof

Author's Proof

6. Clean the PyroMark Q96 Vacuum Prep Tool by turning the vacuum on and allowing it to sit in the first trough (i) containing clean water for 30 s. With the vacuum still running, carefully lift the PyroMark Q96 Vacuum Prep Tool out of the trough and shake in the air for 30 s to remove any excess water. Store on a piece of dry paper towel (see Note 8).

3.4.3. Pyrosequencing

Before pyrosequencing can be performed, each assay must be programmed into the software and the arrangement of assays within a plate must be entered into the operating software. For each assay, the following information must be entered:

- "Sequence to analyze", which is the sequence immediately following the pyrosequencing primer, including the focal SNP. For the assay shown in Fig. 1, the sequence to analyze is 5′-AG/CTCAGTAGT-3′. Note that this is the reverse complement of the sequence shown in Fig. 1 because of the strand that anneals to the pyrosequencing primer.
- "Dispensation order", which is the order in which nucleotides will be dispensed. For the assay shown in Fig. 1, one possible dispensation order is tAGCaTCAGTAGT. The lower case "t" and "a" nucleotides are not expected to be added to the growing oligonucleotide chain, but rather are used to detect background signal in the assay. The PyroMark operating software will determine the dispensation order based on the sequence to analyze; however, the user can also manually edit the dispensation order and may want to do so in some cases (see Note 9).

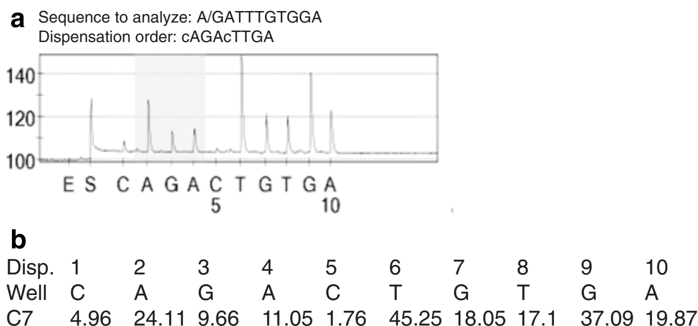
Place PyroMark Q96 Plate Low containing streptavidin beads bound to biotinylated single-stranded PCR product with pyrosequencing primer annealed and the PyroMark Q96 Cartridge filled as described in Subheading 3.4.2 step 4(a) into the pyrosequencing machine and run as instructed in the operating manual. At the end of a pyrosequencing run, which takes approximately 15 min for a single 96-well plate, the results from each well are provided graphically in a pyrogram (Fig. 3a). The height of each nucleotide specific peak can also be exported in a tabular format (Fig. 3b). For large projects involving many different assays and/or many different biological samples, custom programs can be written to most efficiently calculate relative allele-specific expression levels using the table of peak height information (see Note 10). [AU1]

3.5. Experimental Design

This section deals with the most effective experimental design for inferring *cis*- and *trans*-regulatory differences between genotypes, including critical controls, an optional titration series analysis, and considerations for levels of replication required to obtain reliable data.

Uncorrected Proof

Author's Proof



Disp.	1	2	3	4	5	6	7	8	9	10
Well	C	A	G	A	C	T	G	T	G	A
C7	4.96	24.11	9.66	11.05	1.76	45.25	18.05	17.1	37.09	19.87

Fig. 3. Sample output from pyrosequencing. (**a**) A "pyrogram" showing the peaks reflecting the amount of light produced following the addition of the enzyme (E), substrate (S), and individual nucleotides (C, A, G, A, C, T, G, T, G, A) to the pyrosequencing reaction. The sequence to analyze and dispensation order are also shown. *Lower case bases* are not expected to be incorporated. Note the presence of some background signal when the first cytosine was added to the reaction. Also note that this assay involves a homopolymer (AA) in one of the two alleles. The dispensation order was also altered from the default suggested by the pyrosequencing software to completely separate peaks from the two alleles. In this case, the software recommended a dispensation order of "cGAc", whereas the dispensation order used was "cAGAc". This change allows the first A peak to represent only one allele (AATTTGTGGA), whereas the G and s A represent only the other allele (GATTTGTGGA). The rationale for this modification is further described in Note 2. *Peaks* reflecting incorporation of nucleotides corresponding to the alternate alleles are *shaded gray*. The *first A peak* represents incorporation of two successive A nucleotides. (**b**) *Peak heights* shown in the tabular output format. In panel (**a**), the percentage of the molecules attributed to the A allele of the SNP reported by the pyrosequencing software (using its proprietary formulas) was 51.5%. Using the data from the peak heights table and the default correction factor for each adenine of 0.86, the percentage of molecules attributed to the A allele of the SNP is $(0.86 \times 24.11)/((0.86 \times 24.11) + 9.66 + (0.86 \times 11.05)) = 52\%$.

Uncorrected Proof

3.5.1. Controls

Prior to the use of any new pyrosequencing assay to collect biologically relevant data, there are three important control pyrosequencing reactions that must be run:

- *Confirm allele-specific genotypes*: Pyrosequencing reactions should be run to analyze PCR products derived from each individual genotype examined to confirm the focal SNP alleles that the assay was designed to detect.
- *Template only*: A reaction should be performed by replacing the annealing buffer containing the pyrosequencing primer in the PyroMark Q96 Plate Low with annealing buffer alone. Any signal pyrosequencing signal observed in this reaction is attributed to a hairpin forming at the 3′ end of the single-stranded PCR product that primes DNA synthesis (see Note 11).
- *Pyrosequencing primer only*: A reaction should be performed in which the PCR product is replaced with water. Sepharose from this well should be transferred to a well in the PyroMark Q96 Plate Low containing the pyrosequencing primer for each assay

Author's Proof

in annealing buffer. Any pyrosequencing signal observed in this reaction is attributed to the pyrosequencing primer hybridizing to itself (in either a hairpin or primer dimer) at the 3′ end that primes DNA synthesis (see Note 11).

3.5.2. Titration Series (Optional)

Analysis of a titration series is recommended for first-time users and for trouble-shooting; however, it is not strictly required for all assays once the protocol has been well established in a laboratory. It can be performed as follows:

1. Use PCR primers for a given assay to amplify each of the alleles to be compared in separate PCR reactions, as described for the "*confirm allele-specific genotypes*" control reaction in the preceding section.
2. Dilute both PCR products to approximately 1 ng/μl. Estimates based on the intensity of bands on an agarose gel are sufficient for this purpose.
3. Perform serial dilutions with each PCR product to get a range of concentrations. For example, to examine a titration series from 70:30 to 30:70, with steps of ~10%:
   (a) Combine 16 μl of original PCR product and 4 μl water to get dilution #1, and then combine 16 μl of dilution #1 and 4 μl of water to get dilution #2. Repeat this process five more times to produce dilutions #1–7 for each PCR product.
   (b) Combine 3 μl of dilution #3 from PCR product A with 3 μl of dilution #7 from PCR product B to generate a solution that is 70.94% PCR product A by volume. Add 0.6 μl water to dilute DNA concentration to be equal to the 50% sample described in step (d) below.
   (c) Combine 3 μl of dilution #4 from PCR product A with 3 μl of dilution #6 from PCR product B to generate a solution that is 60.98% PCR product A by volume. Add 0.15 μl water to dilute DNA concentration to be equal to the 50% sample.
   (d) Combine 3 μl of dilution #5 from PCR product A with 3 μl of dilution #5 from PCR product B to generate a solution that is 50% PCR product A by volume.
   (e) Combine 3 μl of dilution #6 from PCR product A with 3 μl of dilution #4 from PCR product B to generate a solution that is 39.02% PCR product A by volume. Add 0.15 μl of water to dilute DNA concentration to be equal to the 50% sample.
   (f) Combine 3 μl of dilution #7 from PCR product A with 3 μl of dilution #3 from PCR product B to generate a

Uncorrected Proof

Author's Proof

solution that is 29.06% PCR product A by volume. Add 0.6 μl of water to dilute DNA concentration to be equal to the 50% sample.

(g) For each of these five combinations of PCR products, use 2 μl as the template for new PCR reactions and analyze the PCR products with pyrosequencing.

4. A PCR/pyrosequencing reaction should also be performed in parallel using genomic DNA from cells known to be heterozygous for the two alleles under study. This provides a pyrosequencing measurement for a template that is known to contain exactly 50% of each allele and measures the relative PCR amplification efficiency of the two alleles.
5. Using the peak heights from the focal SNP from each pyrosequencing reaction, divide the peak height from the nucleotide present in allele 1 by the peak height from the nucleotide present in allele 2.
6. Compare these ratios from the heterozygous DNA pyrosequencing reaction and the 50:50 PCR product pyrosequencing reaction by dividing the allele 1–allele 2 ratio from the 50:50 DNA sample by that from the heterozygous DNA sample. This value is called "*C*", and represents the concentration of PCR product 1 at the beginning of the serial dilutions relative to the concentration of PCR product 2 at the beginning of the serial dilutions. The 50:50 DNA sample is affected by the same (if any) PCR amplification bias between alleles observed in the heterozygous DNA sample, which is cancelled out by taking the ratio, leaving only the relative concentration of the starting PCR products used to construct the titration series (6).
7. To compare the pyrosequencing estimates of relative allelic abundance in the titration series to the "true" relative allele abundance in each sample, the volume-based proportions of the two alleles must be adjusted for concentration differences between the two PCR products used to construct the titration series. These adjusted values are calculated as: (proportion by volume of allele $1 \times C$)/(proportion by volume of allele $1 \times C +$ proportion by volume of allele 2). For example, in the [AU2] sample containing 30% by volume of allele 1, the "true" proportion of allele 1 in the mixed sample (by number of molecules) would be $(0.3 \times C)/((0.3 \times C) + 0.7))$.

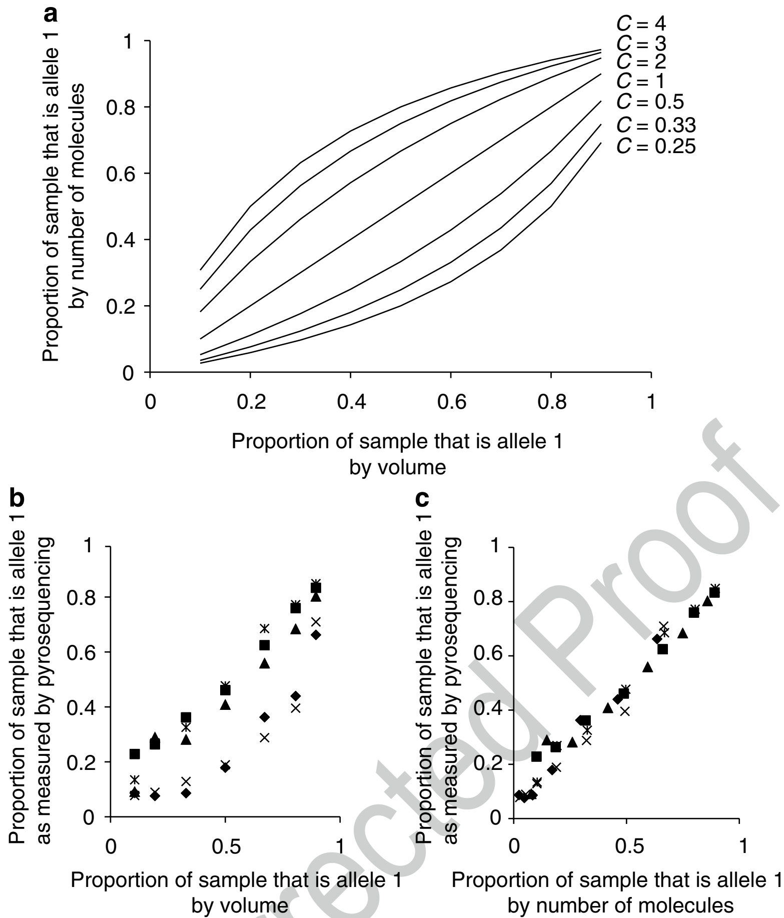
Figure 4 shows a theoretical and an empirical example of results from a titration series.

3.5.3. Replication

To obtain the most reliable estimates of relative allelic expression, an experiment should include replicate biological samples as well as replicate pyrosequencing measurements of each biological sample. The most appropriate amount and type of replication will vary

Uncorrected Proof

Author's Proof



Uncorrected Proof

Fig. 4. Theoretical and empirical results for titration series analysis of allele abundance using pyrosequencing. (**a**) Theoretical relationship between the relative abundance of alleles in mixed samples constructed by volume and the relative abundance of the same allele in the same mixed samples after adjusting for differences in the starting concentrations of samples used to construct the titration series. Relative concentrations (C = allele 1/allele 2) ranging from 0.25 to 4 (i.e., 1/4 to 4/1) are shown. For real samples, C can be calculated empirically as described in step 6 of Subheading 3.5 of the main text. (**b**) Measurements of relative allele abundance determined by pyrosequencing are compared to the relative volume of each allele combined for the titration series. (**c**) Measurements of relative allele abundance determined by pyrosequencing are compared to the relative number of molecules from each type of allele as determined by the volume of each allele combined for the titration series as well as the empirically determined relative concentration C, as described in step 7 of Subheading 3.5 of the main text. Data shown in (**b**) and (**c**) are from a titration series performed for pyrosequencing assays used to analyze five *D. melanogaster* genes (for more details, *see* the Supplementary Information section in ref. 6).

according to the researcher's experimental design. In my experience working with genomic DNA and cDNA samples from *Drosophila* species, the overwhelmingly largest source of variance is replicate samples of biological tissue, with replicate RNA samples having greater variance than replicate genomic DNA samples; replicate samples containing tissue from two different genotypes nearly always show greater variance than replicate samples containing a

Author's Proof

single genotype. Of the technical sources of variance, cDNA synthesis, PCR, and pyrosequencing reactions, in this order, usually have progressively decreasing contributions to variance. A typical experiment performed in my laboratory will include analysis of both genomic DNA and cDNA from at least three replicate biological tissue samples. cDNA synthesis is performed in duplicate at least for each RNA sample, and PCR reactions for each genomic DNA or cDNA sample are performed in triplicate at least. Each PCR product is analyzed in a single pyrosequencing reaction.

3.6. Data Analysis and Interpretation

Before calculating any summary statistics, the pyrogram from each pyrosequencing reaction should be examined manually to identify and exclude any obviously failed reactions (e.g., pyrograms with no peaks), reactions with background signal contaminating the peaks for the focal SNP nucleotides, and cases where the peak heights are outliers (i.e., either very high or very low compared to most other reactions). From the remaining wells, the relative expression of the two alleles is calculated as the ratio of peak heights resulting from the incorporation of alternative nucleotides at the SNP. The percentages of each allele reported by the pyrosequencing software can also be used in some cases without further calculation.

To test for (a) a significant expression difference between genotypes, (b) a significant expression difference between alleles in F_1 hybrids, and (c) a significant difference in relative expression between the parental genotypes and between alleles in the F_1 hybrids, we fit the data to a mixed general linear model in SAS using proc MIXED that incorporates the replication structure of the experiment and performs the pairwise *t*-tests used to infer *cis*- and *trans*-regulatory differences (for examples of appropriate models, see refs. 5, 7). A significant difference in expression between two genotypes is interpreted as divergent expression. A significant difference in expression between alleles in F_1 hybrids is interpreted as evidence for *cis*-regulatory differences between the two parental alleles. And a significant difference in relative expression between parental genotypes and between alleles in F_1 hybrids is interpreted as evidence for *trans*-regulatory differences between the two parental genotypes.

4. Notes

1. Because pyrosequencing uses a single SNP to compare expression of two alleles, measurements reflect expression only of the exon containing this focal SNP. If this exon is not present in all isoforms of mRNA transcribed from the gene (e.g., because of alternative splicing), the estimate of relative transcript

Uncorrected Proof

Author's Proof

abundance in a cDNA pool provided by a single pyrosequencing assay may not accurately reflect the relative abundance of all exons of the gene.

2. Longer homopolymers or more complex SNP sequence contexts can be used (usually with an increased variance of replicate measurements) if the user can find a way to estimate the proportion of one or more peaks attributable to each allele. For example, if allele 1 contains the sequence GGG and allele 2 contains the sequence TGG, a dispensation order of GTG could be used in which the first peak would measure the incorporation of the three consecutive Gs from allele 1, the second peak would measure the incorporation of the T from allele 2 and the third peak would measure the incorporation of the two Gs following the T in allele 2. In this case, the relative abundance of allele 1 to allele 2 could be calculated as the height of peak 1 divided by the combined heights of peaks 2 and 3.

3. The manufacturers of pyrosequencing machines recommend the use of 0.86 as a correction factor for the incorporation of dATP αS. That is, the height of a peak produced by the addition of adenine to the pyrosequencing primer should be multiplied by 0.86 to make its height equal to a peak reflecting incorporation of the same number of molecules of dCTP, dGTP, or dTTP. This correction factor can also be determined empirically for a given reaction by comparing the peak height resulting from an A that is present in both allele sequences to the peak height for a C, T, or G that is present in both allele sequences, assuming that none of the nucleotide peaks examined is part of a homopolymer.

4. If a researcher is interested only in testing for *cis*-regulatory differences between alleles in a heterozygous genotype, only tissue from this genotype is required. This is the most common experimental design used for analysis of human genes, since tissue from only a single heterozygous individual is typically available (e.g., (13)). To compare expression between recently duplicated ("paralogous") genes within a genome (e.g., (3)), tissue is also needed from only one genotype. To test for genomic imprinting, two heterozygous genotypes should be examined separately that are produced by performing reciprocal crosses between two parental genotypes (e.g., (1, 2)).

5. For *Drosophila melanogaster*, 15 whole adult flies can be easily homogenized by hand in an microfuge tube using a pestle homogenizer. For other tissue types, different homogenization tools may be more suitable. Regardless of the method used, samples should be kept on ice (ideally dry ice) until homogenization to prevent the degradation of RNA. The recommended 15 mg of tissue is based on *Drosophila*. Other amounts of tissue

Uncorrected Proof

Author's Proof

726 may be appropriate for other species, but care should be taken
727 to avoid saturating the SV Total RNA Isolation spin columns;
728 additional guidelines are provided in (12).

729 6. For extremely abundant PCR products, a subset of the PCR
730 reaction should be used. The ideal amount can be determined
731 empirically by examining a pyrogram peak resulting from a
732 nucleotide found in both alleles that is not part of a homopo-
733 lymer. If this peak height is <10 units, greater variance among
734 replicate reactions is often observed; if it is >50 units, the
735 amount of light produced may be detected in neighboring
736 wells, artificially inflating their peak heights, again increasing
737 variance among replicate reactions.

738 7. Volumes given are for analysis of a full 96-well plate of PCR
739 products. If a partial plate is analyzed, smaller volumes of
740 Enzyme, Substrate, and nucleotides can be loaded into the car-
741 tridge. Formulas for determining the required amounts are
742 provided in the reagent use guide included with the Pyro Gold
743 (or Pyro Gold SQA) Reagents kit. Any unused Enzyme,
744 Substrate, or nucleotides can be stored for at least 6 months at
745 –20°C. All components should be warmed to room tempera-
746 ture before use, and the resuspended Enzyme and Substrate
747 should not be refrozen once thawed.

748 8. If preparing more than one 96-well plate for pyrosequencing
749 in the same session, you do not need to perform this full clean-
750 ing step between plates. Simply, leave the PyroMark Vacuum
751 Prep Tool in the fifth trough containing water between plates,
752 completing the full cleaning and drying protocol after the last
753 plate.

754 9. The default dispensation order determined by the pyrose-
755 quencing machine is generally adequate, but I often prefer to
756 alter the dispensation order to maximize use of all nucleotides
757 (i.e., choose nucleotides that are most rare in the sequence to
758 analyze as the unincorporated bases at the start of the sequence
759 and following the SNP) and to separate base incorporation at
760 and near the SNP as completely as possible. For example, the
761 software typically suggests a dispensation order of "CG" to
762 assay the sequence G/CG. This dispensation order results in a
763 peak from the incorporation of C that reflects only one allele,
764 but a peak from the incorporation of G that includes the G fol-
765 lowing the C in one allele as well as the two consecutive Gs in
766 the other allele. By comparison, a dispensation order of "GCG"
767 will result in the first G peak reflecting only one allele and the
768 following two peaks (C and G) reflecting the other allele.

769 10. Custom programs can be written in languages such as Perl or
770 Python that can easily parse the exported file containing a table
771 of peak heights for each well. The "notes" section of the

Uncorrected Proof

Author's Proof

pyrosequencing software can also be used to provide descriptive information for the sample in each well, which can then be extracted from exportable summary files that allow automated data processing and analysis. Any such program will need to be customized to fit the researcher's needs.

11. If the source and sequence of any background signal can be determined, it may be possible to adjust the dispensation order of nucleotides for the assay to prevent the extension of hairpins and primer dimers. If this can be done, it is still possible to obtain reliable data from the focal SNP in reactions containing both the PCR template and pyrosequencing primer.

References

1. Wang X, Sun Q, McGrath SD et al (2008) Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS One **3**:e3839
2. Wittkopp PJ, Haerum BK, Clark AG (2006) Parent-of-origin effects on mRNA expression in *Drosophila melanogaster* not caused by genomic imprinting. Genetics **173**: 1817–1821
3. Anderson DW, Evans BJ (2009) Regulatory evolution of a duplicated heterodimer across species and tissues of allopolyploid clawed frogs (*Xenopus*). J Mol Evol **68**:236–247
4. Cowles CR, Hirschhorn JN, Altshuler D et al (2002) Detection of regulatory variation in mouse genes. Nat Genet **32**:432–437
5. Landry CR, Wittkopp PJ, Taubes CH et al (2005) Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics **171**:1813–1822
6. Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. Nature **430**:85–88
7. Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between *Drosophila* species. Nat Genet **40**:346–350
8. Tung J, Primus A, Bouley AJ et al (2009) Evolution of a malaria resistance gene in wild primates. Nature **460**:388–391
9. McManus CJ, Coolon JD, Duff M et al (2010) Regulatory divergence in *Drosophila* revealed by mRNA-Seq. Genome Research **20**:826–25
10. Ahmadian A, Gharizadeh B, Gustafsson AC et al (2000) Single-nucleotide polymorphism analysis by pyrosequencing. Anal Biochem **280**:103–110
11. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers, in Bioinformatics Methods and Protocols: Methods in Molecular Biology. (Krawetz S, Misener S, Eds.), pp 365-386, Humana Press, Totowa, NJ
12. Otto P, Kephart D, Bitner R et al (1998) Separate isolation of genomic DNA and total RNA from single samples using the SV Total RNA Isolation System. Promega Notes **69**:19
13. Wang H, Elbein SC (2007) Detection of allelic imbalance in gene expression using pyrosequencing. Methods Mol Biol **373**:157–176

Uncorrected Proof

Author's Proof

Author Queries

Chapter No.: 18 0001278785

Queries	Details Required	Author's Response
AU1	In the sentence 'Place PyroMark Q96 Plate Low containing…' 'in 3.4.4.a' has been changed to 'Subheading 3.4.2 step 4(a)'. Please check if appropriate.	
AU2	Parenthesis count mismatch in the following sentence 'For example, in the sample containing…' Please check.	

Uncorrected Proof