

Research Report

Use of Denaturing HPLC to Map Human and Murine Genes and to Validate Single-Nucleotide Polymorphisms

BioTechniques 28:740-745 (April 2000)

Lynn M. Schriml, Raymond J. Peterson, Bernard Gerrard and Michael Dean

National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD, USA

ABSTRACT

Linkage mapping has been extensively applied in the murine and human genomes. It remains a powerful approach to mapping genes and identifying genetic variants. As genome efforts identify large numbers of single-nucleotide polymorphisms, it will be critical to validate these polymorphisms and confirm their gene assignment and chromosomal location. The presence of pseudogenes can confuse such efforts. We have used denaturing HPLC to identify polymorphisms in human genes and to genotype individuals in selected CEPH pedigrees. The same approach has been applied to the mapping of murine genes in interspecies backcross animals. This strategy is rapid, accurate and superior in several respects to other technologies.

INTRODUCTION

Current efforts to identify large collections of single-nucleotide polymorphisms (SNPs) in the human genome (4) present a number of challenges to researchers. While high-throughput sequencing strategies allow the identification of many variants, validating

these SNPs requires considerable effort. Such validation must include a determination of the frequency of each SNP in several populations and confirm the location of the marker. The presence of pseudogenes can lead to the misidentification of polymorphisms or lead to the incorrect localization of the SNP. While a number of technologies exist for large-scale genotyping of SNPs, most of these methods involve considerable expense or effort to establish a working assay. Therefore, an intermediate validation step would be valuable to select the most useful SNPs for large-scale genotyping.

The use of radiation hybrids for gene mapping has largely replaced linkage mapping in human genome efforts (16). However, interspecies backcross mapping remains the principal means of gene mapping in the mouse (14). The advantage of backcross mapping is that errors are readily detectable as double recombinants and can be removed as the density of the map increases. A limitation of radiation hybrid mapping is that the errors are difficult to recognize. However, with higher-resolution radiation hybrid panels and more accurate typing, the resolution of radiation hybrids can exceed that of backcross analysis.

Denaturing HPLC is a rapid and efficient method for detecting variation in DNA as heteroduplexes (19). DNA is allowed to bind to a hydrophobic column in a buffer of triethyl ammonium acetate and is eluted with an increasing gradient of acetonitrile (6). When the column is run at an elevated temperature, partial denaturation of the dsDNA occurs. Heteroduplex molecules formed in heterozygotes denature at lower concentra-

tions of acetonitrile and are detected as a peak or peaks with shorter retention times than the homozygotes. The column can regenerate within a minute, resulting in run lengths of 6–10 min and a high sample throughput. The DNA is injected into the column directly from the PCR and the cost per reaction is low. The sensitivity of the method is quite high, making it competitive, if not superior, to most existing scanning methods (7), and its usefulness in genotyping has been demonstrated (18).

MATERIALS AND METHODS

PCR

For backcross mapping, amplification was performed with AmpliTaq® Gold DNA polymerase in a 25 µL reaction volume containing 40 ng DNA with 1× PCR buffer (PE Biosystems, Foster City, CA, USA) in a step-down program. For *Abca7*, a 267 bp product was generated with primers *Abca7F* (5'-CCTTAGAGTCAGTGTGTGGC-C-3') and *Abca7R* (5'-TGGATGGTG-TCCAGGTAGG-3'). The *JAK3* SNP was amplified with primers *JAK3F1* (5'-CTGAGGGGCTTTTTCTCTCT-3') and *JAK3R1* (5'-GCAACTATCC-AGCTTCCAA-3'), the *WT1* SNP with primers *WT1F2* (5'-GCTGTCTCGGT-TCTCTCTG-3') and *WT1R2* (5'-AAGACCCAGACGCAGAG-3'). Samples were denatured by heating at 94°C for 10 min, followed by 5 cycles of 94°C for 20 s; 60°C for 20 s; 72°C for 30 s and then by 35 cycles at 94°C for 20 s; 60°C for 20 s; 72°C for 30 s and a final extension at 72°C for 5 min. Following the PCR, samples were heated

Research Report

were carried out at 50°C for 2 min, 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and at 62°C for 1 min. Signal processing was performed on the TaqMan 7700 (PE Biosystems) using endpoint analysis.

RESULTS

Human SNP Detection and Validation

As part of identifying SNPs in cancer-related genes, we have scanned a large number of gene segments in parents from CEPH families. To validate the SNPs, we tested the feasibility of using denaturing HPLC to genotype the remaining members of the family. The primary limitation of this approach is that denaturing HPLC cannot distinguish the two homozygous classes of an SNP without mixing with a homozygous control sample. In pedigrees in which only one parent is heterozygous, this is not a limitation because there is only one homozygous class. Figure 1A shows CEPH pedigree 1331, in which a polymorphism in the first intron of the *Janus kinase 3* (*JAK3*) gene (406T/C) is heterozygous in the father and homozygous in the mother. The *JAK3* SNP segregates with the two closest flanking microsatellite markers (*D19S899* and *D19S407*), which confirms that this is a true polymorphism and the gene assignment is correct.

Figure 1B displays the same pedigree segregating a variant in the *WT1* gene in which the SNP is heterozygous in both parents. To resolve the two classes of homozygotes, each homozygous sample was mixed with a sample homozygous for allele 1. Samples 4, 5, 6, 10, 13 and 14 remained homozygous and samples 11, 16 and 17 appeared heterozygous. These genotypes were all consistent with those of the flanking microsatellites at the *D11S1324* and *D11S914* loci.

5' Nuclease Assay Genotyping of *JAK3* SNP

The *JAK3* SNP was genotyped in 480 members of 40 CEPH pedigrees (5). Segregation in all families was consistent with Mendelian transmission as de-

termined by the algorithms implemented in Pedcheck (13). In 102 CEPH grandparents, allele frequency of the C allele was 0.671. This frequency was 0.483 in 30 African Americans and 0.583 in 24 Hispanics. None of the genotype distributions departed from the Hardy-Weinberg expectation as determined by a contingency table χ^2 test (20).

Denaturing HPLC for Mouse Backcross Mapping

To characterize all murine ATP-binding cassette genes, we have been mapping new genes as they are identified in the expressed sequence tag (EST) database. One of these genes, *Abca7*, was mapped in the T31 mouse/hamster radiation hybrids (WGRH, Research Genetics) to chromosome 10 at 44-48 cM (manuscript in press). To more precisely position *Abca7*, we amplified a segment of the gene from the two parental strains of the Jackson Laboratory (Bar Harbor, ME, USA; <http://www.jax.org/resources/documents/cmdata>) interspecific backcross panel, C57BL/6J and *Mus spretus*, and a third sample comprised of a mixture of the two parental DNAs. Denaturing HPLC analysis revealed that both parental species produced a trace with a single peak (Figure 2, A and B). The hybrid DNA yielded a clear heterozygous pattern with two peaks, indicating the presence of one or more sequence variants between the two species. The 94 F1 backcross animal DNA samples from

each of the BSB and BSS backcross panels were typed by denaturing HPLC and the results submitted to the Jackson Laboratory Backcross Mapping service. The results clearly placed *Abca7* at 44 cM from the centromere of chromosome 10 (Figure 3). Figure 3A represents the markers and mapping data surrounding *Abca7* and Figure 3B represents the haplotype data for *Abca7*. Because there are a larger number of markers typed on the BSS panel (4160) compared with BSB (920), a more precise localization was obtained. Single recombinants with the *D10Mit22* and *D10Jcs30* loci allow *Abca7* to be placed between these two markers.

DISCUSSION

The National Human Genome Research Institute (NHGRI, a part of the National Institutes of Health) has funded efforts to identify 100 000 SNPs (3) and a consortium of pharmaceutical companies is supporting the discovery and public release of 300 000 SNPs. In addition, several companies are assembling private SNP collections (12). The maximal use of all of these SNPs will require the proper validation of the allele frequency of the SNPs in the major racial groups, the analysis of haplotype frequencies and linkage disequilibrium with surrounding polymorphisms.

SNP genotyping can be accomplished by a number of methods that include restriction enzyme digestion,

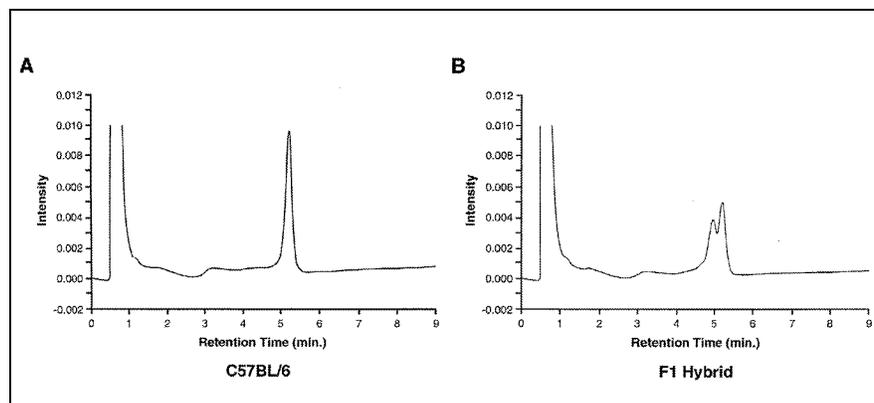


Figure 2. Denaturing HPLC traces of mouse parental and F1 DNA. DNA from the backcross parental C57BL6/JEi strain (A) and from a pool of C57BL6/JEi and SPRET/Ei DNA (F1 hybrid) (B) were used to amplify a segment of the *Abca7* gene. The DNA was analyzed by denaturing HPLC at 61°C resulting in unique traces for parental and F1 hybrid DNA. Traces for the two parental strains, SPRET/Ei and C57BL6/JEi (data not shown), were identical.

Research Report

allele-specific hybridization, primer extension, DNA arrays, mass spectrometry, 5' nuclease cleavage and molecular beacons (2,8–11,15,17). These techniques require the design of a specific assay and in many cases the synthesis of costly modified oligonucleotides. When the assay is used to genotype thousands of samples, on average this cost can be inexpensive. However, considerable care will be required to choose the SNPs in a given gene or region that will be most informative for that locus. A rapid method for prescreening and characterizing SNPs will be useful in selecting markers for large-scale genotyping. Denaturing HPLC has a number of advantages for this step. The method is highly sensitive to identifying polymorphisms and will be useful for most variants. A wide range of fragment sizes can be employed and modified primers are not required. Because of this, the same sequence-tagged site that was used to identify the SNP can often be used for denaturing HPLC validation.

Where denaturing HPLC was used

to discover the SNP, denaturing HPLC validation studies can be performed without modifying the discovery conditions. By using CEPH parents as part of the SNP discovery, we quickly moved to a three-generation pedigree to demonstrate segregation. The presence of a large number of microsatellite markers in the radiation hybrid database allowed us to easily select loci flanking the gene of interest. In the absence of family samples, denaturing HPLC can still be used to estimate allele frequencies. Although the homozygotes will not be distinguished without a mixing step, the frequency of heterozygotes provides an accurate measure of the allele frequency. While the throughput with current technology is about 150 samples per day, this num-

ber may improve with the use of fluorescent or mass spectrometry-tagged primers, decreases in run time or multi-column machines. The cost of denaturing HPLC is currently about \$0.80 per sample, which is less than TaqMan (\$1.00/sample for 1000 assays) and PCR-RFLP (\$1.00 or more, depending on the enzyme).

Mapping of mouse genes has largely been accomplished by interspecies backcross mapping (1). The advantage of a cross between species is the high probability of finding a sequence variant that can be used to map the gene of interest. Traditionally, backcross mapping required that a variant be identified by Southern blotting, single-strand conformation polymorphism analysis or sequencing. The advantage of dena-



Group Leader Analytical Biochemistry:
Midsize publicly held biotech company is looking for a Ph.D. with strong Analytical Protein Biochemistry. Must have industry experience and a broad knowledge of analytical techniques such as HPLC, gel electrophoresis, Mass Spec., etc. Interested candidates should send their information to Amy@techfind.com or call (508) 647-0111 for more information.

Group Leader or Director Molecular Biology
Privately held start-up biotech company seeks a Ph.D. in classical Molecular Biology with experience in assay development and phage display. Minimum of 5 years technical and managerial, industry experience is a must. Interested candidates should send their information to Amy@techfind.com or call (508) 647-0111 for more information.

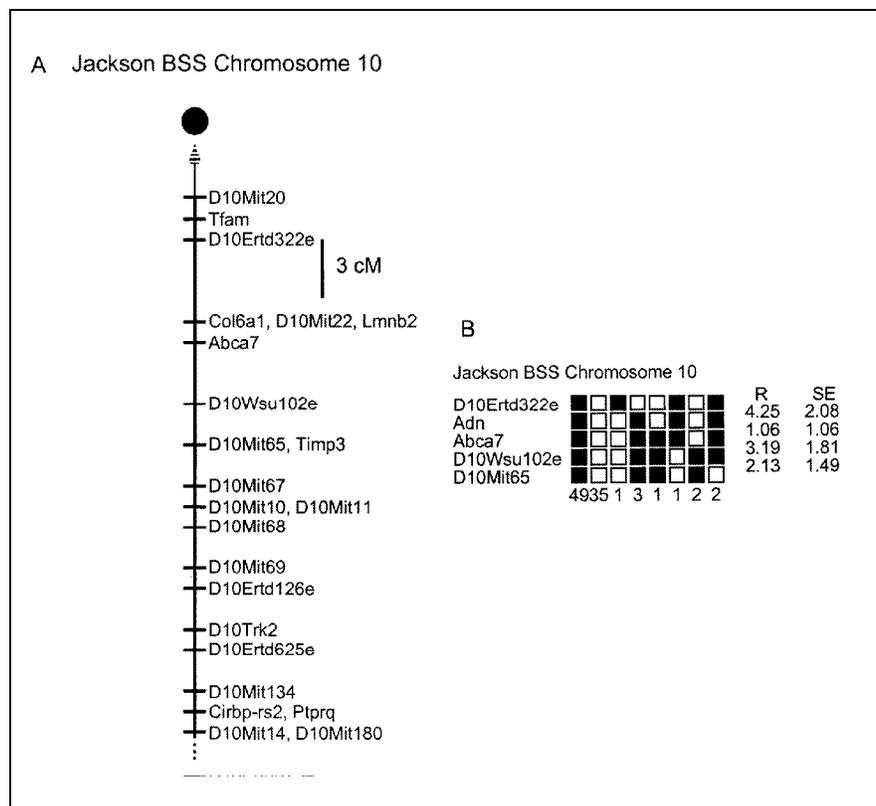


Figure 3. Localization of the *Abca7* gene on mouse chromosome 10. (A) Map figure from the Jackson Laboratory interspecific BSS backcross panel showing a portion of Chromosome 10. The map is depicted with the centromere toward the top with a 3 cm scale bar at the right of the figure. Loci mapping to the same position are listed in alphabetical order. Missing typings were inferred from surrounding data where assignment was unambiguous. Raw data were obtained from the Jackson Laboratory. (B) Haplotype figure from the Jackson Laboratory interspecific BSS backcross panel showing part of Chromosome 10 with loci linked to *Abca7*. Loci are listed in order with the most proximal at the top. The black boxes represent the C57BL6/JEi allele and the white boxes the SPRET/Ei allele. The number of animals with each haplotype is given at the bottom of each column of boxes. The percent recombination (R) between adjacent loci is given to the right of the figure, with the standard error (SE) for each R. Missing typings were inferred from surrounding data where assignment was unambiguous.

turing HPLC for backcross mapping is that a variant can be readily identified and the samples typed without gels or radioactivity, steps that can easily be carried out in one day. With denaturing HPLC, backcross mapping becomes as simple as radiation hybrid mapping and has the added benefit of confidence in localization provided by the analysis of recombination data.

ACKNOWLEDGMENTS

We thank Lucy Rowe and Mary Barter for assistance with backcross mapping. This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does any mention of trade names, commercial products or organizations imply endorsement by the US government.

REFERENCES

1. **Bedell, M.A., N.A. Jenkins and N.G. Copeland.** 1997. Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. *Genes Dev.* 11:1-10.
2. **Case-Green, S.C., K.U. Mir, C.E. Pritchard and E.M. Southern.** 1998. Analysing genetic information with DNA arrays. *Curr. Opin. Chem. Biol.* 2:404-410.
3. **Collins, F.S., L.D. Brooks and A. Chakravarti.** 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8:1229-1231.
4. **Collins, F.S., M.S. Guyer and A. Chakravarti.** 1997. Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580-1581.
5. **Dausset, J., H. Cann, D. Cohen, M. Lathrop, J.M. Lalouel and R. White.** 1990. Centre d'étude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 6:575-577.
6. **Huber, C.G., P.J. Oefner, E. Preuss and G.K. Bonn.** 1993. High-resolution liquid chromatography of DNA fragments on non-porous poly(styrene-divinylbenzene) particles. *Nucleic Acids Res.* 21:1061-1066.
7. **Jones, A.C., J. Austin, N. Hansen, B. Hoogendoorn, P.J. Oefner, J.P. Cheadle and M.C. O'Donovan.** 1999. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clin. Chem.* 45:1133-1140.
8. **Koster, H., K. Tang, D.J. Fu, A. Braun, D. van den Boom, C.L. Smith, R.J. Cotter and C.R. Cantor.** 1996. A strategy for rapid and efficient DNA sequencing by mass spectrometry. *Nat. Biotechnol.* 14:1123-1128.
9. **Landegren, U. and M. Nilsson.** 1997. Locked on target: strategies for future gene diagnostics. *Ann. Med.* 29:585-590.
10. **Livak, K. J.** 1999. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet. Anal.* 14:143-149.
11. **Marras, S.A., F.R. Kramer and S. Tyagi.** 1999. Multiplex detection of single-nucleotide variations using molecular beacons. *Genet. Anal.* 14:151-156.
12. **Masood, E.** 1999. As consortium plans free SNP map of human genome. *Nature* 398:545-546.
13. **O'Connell, J.R. and D.E. Weeks.** 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* 63:259-266.
14. **Rowe, L.B., J.H. Nadeau, R. Turner, W.N. Frankel, V.A. Letts, J.T. Eppig, M.S. Ko, S.J. Thurston and E.H. Birkenmeier.** 1994. Maps from two interspecific backcross DNA panels available as a community genetic mapping. *Mamm. Genome* 5:253-274.
15. **Schena, M., R.A. Heller, T.P. Theriault, K. Konrad, E. Lachenmeier and R.W. Davis.** 1998. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* 16:301-306.
16. **Schuler, G.D., M.S. Boguski, E.A. Stewart, L. D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tome et al.** 1996. A gene map of the human genome. *Science* 274:540-546.
17. **Syvanen, A.C.** 1999. From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum. Mutat.* 13:1-10.
18. **Underhill, P. A., L. Jin, A.A. Lin, S.Q. Mehdi, T. Jenkins, D. Vollrath, R.W. Davis, L.L. Cavalli-Sforza, et al.** 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography. *Genome Res.* 7:996-1005.
19. **Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. and Cavalli-Sforza.** 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA* 93:196-200.
20. **Weir, B.** 1996. Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, Sunderland, MA, USA.

Received 2 August 1999; accepted 1 December 1999.

Address correspondence to:

Dr. Michael Dean
Bldg. 560, Rm. 21-18
NCI-FCRDC
Frederick, MD 21702, USA
Internet: dean@ncifcrf.gov