

SHORT REPORT

Application of haplotype pair analysis for the identification of hemizygous loci

B C Hendrickson, D Pruss, E Lyon, T Scholl

J Med Genet 2003;40:346-347

An expectation maximisation based prediction algorithm was created to identify unusual haplotypes in patient samples that may be caused by small intragenic deletions. In this approach, unphased SNP genotypes are compared to pairs of canonical haplotypes to identify potentially hemizygous regions. This method was successfully applied to identify five deletions in the 3' region of *BRCA1*.

Most of the human SNP diversity at a given locus may be described as a set of "canonical" haplotypes, representing common haplotypes in a population. Of the few unusual haplotypes not in the canonical set, some are found in a genotype context that is similar to a genotype expected from a combination of a pair of canonical haplotypes, except that one or several polymorphic positions appear, unexpectedly, homozygous. These unusual haplotypes could represent hemizygous loci resulting from intragenic deletions. However, these could also be bona fide haplotypes that are too rare to be represented in the canonical set, resulting from recombinations between common haplotypes, or single base changes.

In the present study, we used unphased genotype data for 12 common biallelic *BRCA1* SNPs (located from exons 4 to 16) generated during sequence based clinical mutation testing to obtain *BRCA1* SNP haplotypes for 5911 anonymised samples, by applying an expectation maximisation (EM) algorithm similar to those described elsewhere.^{1,2} Ten canonical SNP haplotypes are known for *BRCA1*.³ Interestingly, two *BRCA1* haplotypes account for the bulk of the genotypes, the consensus at 59% frequency, and the most common non-consensus haplotype at 21% (fig 1).

We identified 78 (1.3%) samples with non-canonical haplotypes, which in 62 cases were paired with a canonical haplotype. Among these samples, 14 were identified with rare haplotypes created by a change at the SNP in exon 16 (fig 1), suggestive of a possible deletion of this exon. This group was selected for molecular analysis because of its abundance, representing 18% of the unusual haplotypes. Also, additional sequence information available for one of these samples showed a heterozygous SNP near exon 17, placing a 3' limit on the size of the putative deletion. This discovery increased the likelihood that the unusual haplotype in this sample, and perhaps others in this group, were the result of intragenic deletions, and not just recombination.

Within the remaining samples, 42 contained rare haplotypes that appeared to arise from changes in one out of five SNP loci in exon 11, potentially indicative of a partial deletion of the exon. These samples were excluded from deletion testing because, to date, all clinically significant large deletions that have been characterised in *BRCA1* are *Alu* mediated, whole exon deletions.⁴⁻⁸ An additional 17 samples were excluded from deletion testing because their haplotypes were defined by changes at two, non-adjacent haplotype defining

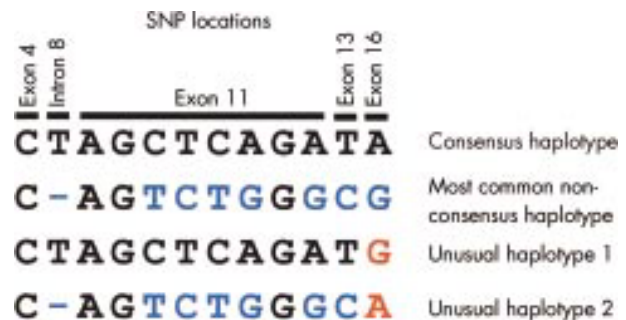


Figure 1 Haplotype definitions and unusual haplotypes in *BRCA1*. Twelve SNPs were used to define haplotypes in *BRCA1*. The most common non-consensus haplotype is defined by changes at eight of the 12 sites (indicated by blue text in the second haplotype; a dash indicates a single deleted base). Fourteen samples were identified with two haplotypes that could suggest a possible deletion of *BRCA1* exon 16. The first haplotype, which was found in nine samples, is similar to the consensus haplotype with the exception of the non-consensus base in exon 16. In these samples, this haplotype was paired with the most common non-consensus haplotype. The second rare haplotype was identified in five samples and was paired with the consensus haplotype. This haplotype was similar to the most common non-consensus haplotype with the exception of the wild type base in the last SNP location. Divergent bases are shown in red text. In all cases, the haplotype pair data appear as heterozygous at the eight SNPs except exon 16, which was homozygous for either the consensus or non-consensus base.

SNPs, which could not be explained by a single deletion event. Several other samples were identified with haplotypes that may have been the result of intragenic deletions. However, the DNA from these samples was not available.

Long range PCR of the region between exons 14 and 18 showed that seven of the selected 14 samples produced smaller than the expected products. Restriction digests of the mutant fragments localised breakpoint regions, which were then characterised by sequencing (fig 2) to show five distinct novel deletions, ranging in size from 5629 to 7183 bases. All five deletions appear to be *Alu* mediated. Additional work is continuing to determine if the remaining seven samples may contain deletions with breakpoints outside the amplified region.

Our example shows that haplotype analysis is effective for identifying samples that may contain hemizygous regions. The use of haplotype analysis permitted rapid processing of a large sample set and selected the best candidate samples for biochemical analysis. Of the selected specimens in our target group, 50% were shown to contain deletions. Also, this example illustrates the potential increased information available from phased data that can be overlooked in unphased genotype data. Finally, this method may have broader application in large SNP based assays by identifying potentially hemizygous data that might confound downstream analyses.

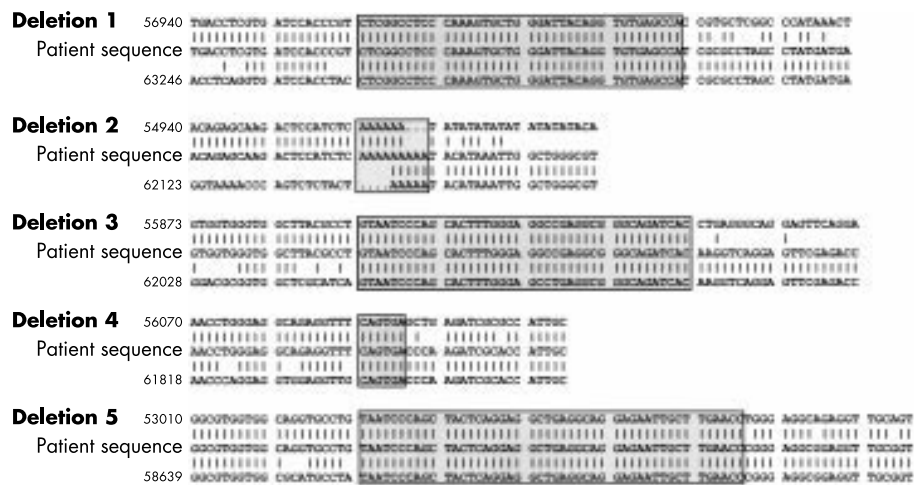


Figure 2 Results of molecular analysis. Deletions were discovered in seven of the 14 samples. Three samples that claimed Latin American ancestry had the same deletion of exons 16 and 17 (deletion 1). Also, three additional deletions involving exons 16 and 17 (deletions 2-4), and one deletion of exons 15 and 16 (deletion 5, also reported elsewhere during the preparation of this manuscript⁹) were identified. All numerical designations correspond to GenBank L78833. Top strands are genomic sequence at the upstream breakpoint region and lower strands are sequence from the downstream breakpoint region. Vertical lines indicate identity between patient data (middle sequence) and expected sequence. Grey boxes highlight the recombination regions.

Authors' affiliations

B C Hendrickson, D Pruss, T Scholl, Myriad Genetics Inc, Salt Lake City, UT 84108, USA

E Lyon, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT 84132, USA

Correspondence to: Dr T Scholl, Clinical Research and Development, Myriad Genetic Laboratories Inc, 320 Wakara Way, Salt Lake City, Utah 84108, USA; tscholl@myriad.com

Revised version received 24 January 2003

Accepted for publication 27 January 2003

REFERENCES

- 1 **Ceppellini R**, Siniscalco M, Smith CAB. The estimation of gene frequencies in a random mating population. *Ann Hum Genet* 1955;**20**:97-115.
- 2 **Excoffier L**, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;**12**:921-27.
- 3 **Shattuck-Eidens D**, Oliphant A, McClure M, McBride C, Gupte J, Rubano T, Pruss D, Tavtigian SV, Teng DH, Adey N, Staebell M, Gumpfer K, Lundstrom R, Hulick M, Kelly M, Holmen J, Lingenfelter B, Manley S, Fujimura F, Luce M, Ward B, Cannon-Albright L, Steele L, Offit K, Gilewski T, Norton L, Brown K, Schulz C, Hampel H, Schluger A, Giulotto E, Zoli W, Ravaiani A, Nevanlinna H, Pyrhonen S, Rowley P, Loader S, Osborne MP, Daly M, Tepler I, Weinstein PL, Scalia JL,

Michaelson R, Scott RJ, Radice P, Pierotti MA, Garber JE, Isaacs C, Peshkin B, Lippman ME, Dosik MH, Caligo MA, Greenstein RM, Pilarski R, Weber B, Burgemeister R, Frank TS, Skolnick MH, Thomas A. BRCA1 sequence analysis in women at high risk for susceptibility mutations: risk factor analysis and implications for genetic testing. *JAMA* 1977;**278**:1242-50.

- 4 **Montagna M**, Santacatterina M, Torri A, Menin C, Zullato D, Chieco-Bianchi L, D'Andrea E. Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. *Oncogene* 1999;**18**:4160-5.
- 5 **Petrij-Bosch A**, Peelen T, van Vliet M, van Eijk R, Olmer R, Drusedau M, Hogervorst FB, Hageman S, Arts PJ, Ligtenberg MJ, Meijers-Heijboer H, Klijn JG, Vasen HF, Cornelisse CJ, van't Veer LJ, Bakker E, van Ommen GJ, Devilee P. BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat Genet* 1997;**17**:341-5.
- 6 **Puget N**, Torchard D, Sinilnikova OM, Lynch HT, Feunteun J, Lenoir GM, Mazoyer S. A 1-kb Alu-mediated germ-line deletion removing BRCA1 exon 17. *Cancer Res* 1997;**57**:828-31.
- 7 **Puget N**, Stoppa-Lyonet D, Sinilnikova OM, Pages S, Lynch HT, Lenoir GM, Mazoyer S. Screening for germ line rearrangements and regulatory mutations in BRCA1 led to the identification of four new deletions. *Cancer Res* 1999;**59**:455-61.
- 8 **Swensen J**, Hoffman M, Skolnick MH, Neuhausen SL. Identification of a 14 kb deletion involving the promoter region of BRCA1 in a breast cancer family. *Hum Mol Genet* 1997;**6**:1513-17.
- 9 **Casilli F**, Rocco ZC, Gad S, Tournier I, Stoppa-Lyonet D, Frebourg T, Tosi M. Rapid detection of novel BRCA1 rearrangements in high-risk breast-ovarian cancer families using multiplex PCR of short fluorescent fragments. *Hum Mutat* 2002;**20**:218-26.