Am. J. Hum. Genet. 67:1348–1351, 2000

Report

Pitfalls in Homozygosity Mapping

Maria G. Miano,[1,2] Samuel G. Jacobson,[3] Andrew Carothers,[1] Isabel Hanson,[1] Peter Teague,[1] Jill Lovell,[1] Artur V. Cideciyan,[3] Neena Haider,[4,6] Edwin M. Stone,[5] Val C. Sheffield,[4,6] and Alan F. Wright[1]

[1]MRC Human Genetics Unit, Western General Hospital, Edinburgh; [2]International Institute of Genetics and Biophysics, Naples, Italy; [3]Department of Ophthalmology, Scheie Eye Institute, University of Pennsylvania, Philadelphia; and Departments of [4]Pediatrics and [5]Ophthalmology, and [6]Howard Hughes Medical Institute, University of Iowa, Iowa City

There is much interest in use of identity-by-descent (IBD) methods to map genes, both in Mendelian and in complex disorders. Homozygosity mapping provides a rapid means of mapping autosomal recessive genes in consanguineous families by identifying chromosomal regions that show homozygous IBD segments in pooled samples. In this report, we point out some potential pitfalls that arose during the course of homozygosity mapping of the enhanced S-cone syndrome gene, resulting from (1) unexpected allelic heterogeneity, so that the region containing the disease locus was missed as a result of pooling; (2) identification of a homozygous IBD region unrelated to the disease locus; and (3) the potential for inflation of LOD scores as a result of underestimation of the extent of inbreeding, which Broman and Weber suggest may be quite common.

We would like to draw attention to some potential problems associated with homozygosity mapping of autosomal recessive disorders in consanguineous pedigrees. Homozygosity mapping was used to study a rare autosomal recessive disorder, enhanced S-cone syndrome (ESCS [MIM 268100]), in three consanguineous pedigrees from North America. ESCS pedigree 1 (ESCS-1; fig. 1) was screened using a set of 386 microsatellite markers spaced at 10-cM intervals (Research Genetics Set 8) and pooled DNA samples—one from the four affected individuals, one from the unaffected sibs, and one from the parents. Markers that were homozygous in the affected pool but not in the other pools were examined further in families ESCS-2 and ESCS-3 (fig. 1), using both pooled and individual samples from each family. Only a single marker, D1S552, remained homozygous in all affected members but not in unaffected members of each family (the three families were unrelated and were homozygous for different alleles). The results using marker D1S552 were repeated and confirmed, and formal LOD-score analysis showed significant linkage with ESCS, with a LOD score of 3.69 (recombination fraction 0), consistent with a locus in chromosomal region 1p36.1 (a disease gene frequency [*q*] of .0045 was assumed).

We then analyzed the families for regions of homozygosity extending outside the linked marker. This was clearly present in ESCS-3 and tentatively present in the other two families. In ESCS-3, five of six chromosomes from affected individuals shared the same inferred haplotype for 18 consecutive markers spanning ⩾18 cM, consistent with autozygosity (probability of IBS 8.2×10^{-14}) (fig. 2). The sixth affected chromosome in individual IV:4 appeared to have recombined and was homozygous for the central five markers, including D1S552, spanning ~2 cM. However, since neither parent was available for study, it is also possible that the apparently recombined chromosome was in fact a different parental chromosome, which was identical-by-state, or that the haplotypes were incorrectly inferred. In addition, the marker order could be incorrect, although this was determined from the best available information, either using the Genetic Location Database or, for markers close to D1S552, from a YAC/BAC contig extending for ~4 megabases around D1S552. The GENEHUNTER program was used to calculate LOD scores in family ESCS-3 between the ESCS locus and D1S552 and flank-

Received June 1, 2000; accepted for publication September 11, 2000; electronically published September 27, 2000.

Address for correspondence and reprints: Dr. Alan F. Wright, MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

0002-9297/2000/6705-0034$02.00

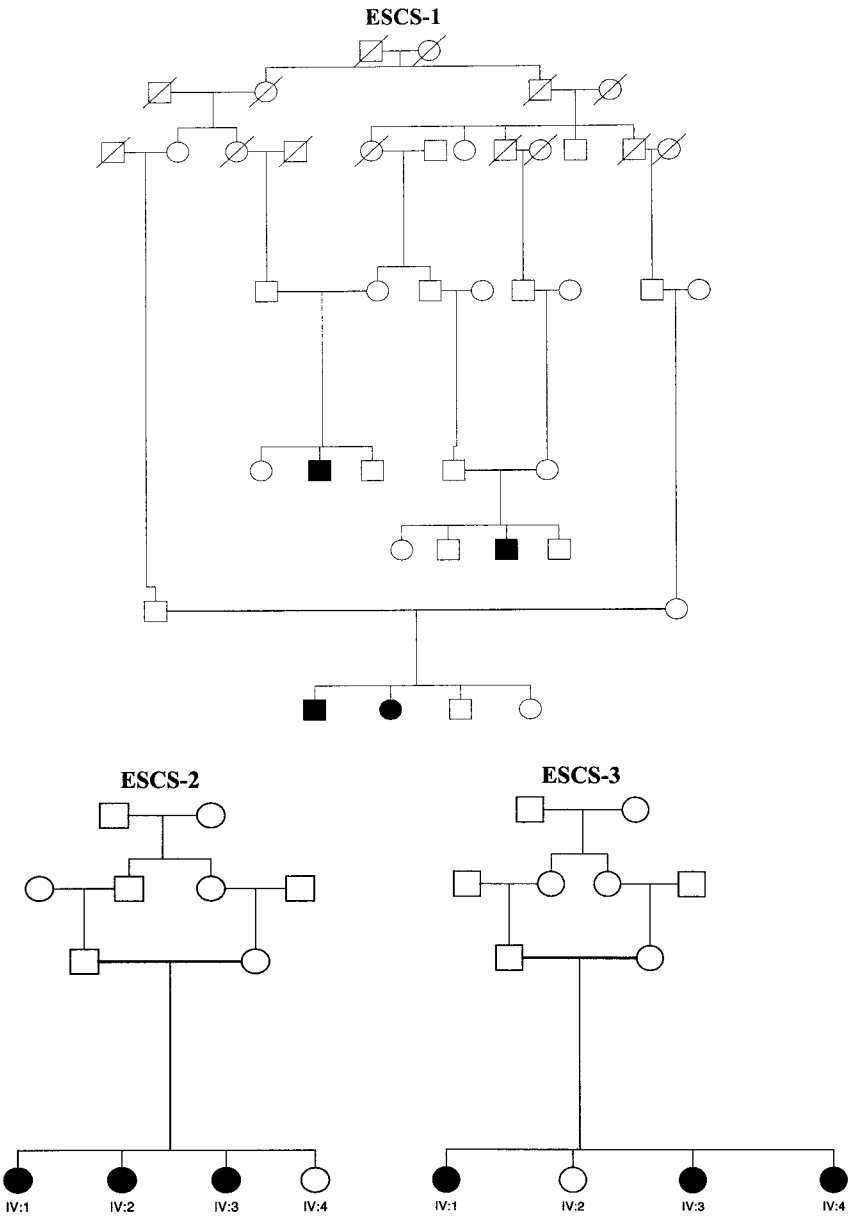
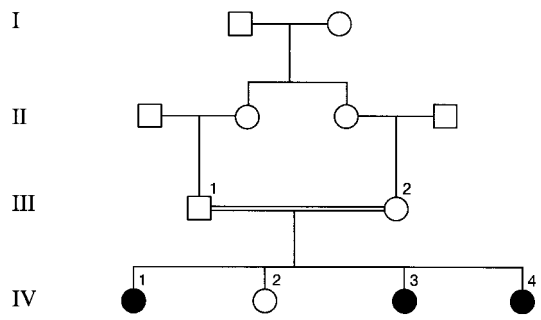


Figure 1 Pedigrees of three consanguineous ESCS kindreds. Pedigree ESCS-1 is the corrected version of the one shown in figure 3 in the study by Haider et al. (2000).

ing markers. This showed a LOD score of 2.5 at a location 54.8 cM (sex-averaged) from distal 1p. The regions of homozygosity were smaller in families ESCS-1 and ESCS-2, extending between two and seven markers around D1S552 but spanning only 1–2 cM. This was attributed to the distal chromosomal location of D1S552, which is often associated with high recombination rates.

On the basis of these data, a physical map of the region was constructed and candidate-gene analysis commenced. In the meantime, the photoreceptor nuclear receptor gene *NR2E3* was mapped to chromosomal region 15q23, and mutations were identified by candidate-gene screening in ESCS patients (Haider et al. 2000). Furthermore, we identified, in family ESCS-3, a disease-causing mutation—a homozygous Arg311Gln missense mutation—which was found by Haider et al. (2000) in 13 other North American families with ESCS. Homozygosity was confirmed in family ESCS-3 for markers flanking the *NR2E3* gene in 15q23, and no doubts remain that this mutation is causal. Mutations in *NR2E3* were also found in affected members of families ESCS-



D1S407	154	154	150	154	154	154	150	154
D1S2144	202	202	202	202	202	202	202	202
D1S436	206	206	206	206	206	206	200	206
D1S199	92	92	92	92	92	92	104	92
D1S2644	222	222	210	222	222	222	222	222
D1S478	170	170	152	170	170	170	170	170
D1S2647	181	181	175	181	181	181	181	181
D1S552	252	252	256	252	252	252	252	252
D1S483	228	228	228	228	228	228	228	228
ATA43C09	127	127	127	127	127	127	137	127
D1S2843	181	181	181	181	181	181	181	181
D1S2732	262	262	262	262	262	262	262	262
GGAA30B06	171	171	195	171	171	171	167	171
D1S2702	145	145	147	145	145	145	145	145
D1S234	232	232	226	232	232	232	232	232
D1S2734	110	110	124	110	110	110	122	110
D1S2885	237	237	251	237	237	237	237	237
D1S2749	216	216	216	216	216	216	212	216

Figure 2 Homozygosity for chromosome 1 markers, suggesting IBD, in affected members of family ESCS-3.

1 and ESCS-2, consisting of an A→C splice-acceptor mutation in intron 1 (ESCS-1) and an Arg97His mutation (ESCS-2), as reported by Haider et al. (2000).

Two questions arise from these findings. First, how was the genuine locus missed? The initial screening was performed in consanguineous kindred ESCS-1, in which the four affected members were the offspring of cousins. However, two different *NR2E3* mutations are segregating in this kindred. Three affected individuals were homozygous for the intron 1 splice-acceptor mutation; the fourth was heterozygous for this mutation and is presumed to carry another, as yet unidentified, mutation in *NR2E3*. This would be expected to result in loss of homozygosity in the pool of affected individuals. The results from markers flanking *NR2E3* were therefore reexamined to see whether this resulted in exclusion of the region. *NR2E3* is close to three markers—D15S643, GATA151F03, and D15S211—used in the 10-cM genome scan. The results showed that all three have either two or three alleles in the DNA pool from affected individuals in ESCS-1; hence, they were excluded. Despite the fact that ESCS is rare and that *NR2E3* mutations are rare in the general population ($\ll$1 in 500), allelic heterogeneity is present in ESCS-1, creating problems in screening by DNA pooling.

The second question raised is the likelihood of observing the large region of IBD segregating with the disease in at least five and possibly six of six disease chromosomes in family ESCS-3. The offspring of a first-cousin marriage have an inbreeding coefficient (F) of .0625, and the probability that the parents share an allele that is identical by descent at any locus is .25. The probability of autozygosity occurring at any particular locus in all six chromosomes of the three affected offspring of family ESCS-3 (fig. 2) is then .0039 ($= \frac{1}{4} \times (\frac{1}{4})^3 = 1/256$). If a sex-averaged autosomal length of 3,488 cM were assumed (Collins et al. 1996), then a total length of autozygosity shared between affected individuals of, on average, ~14 cM ($= 3{,}488 \times 1/256$), which is presumed to be distributed in multiple segments across the genome, would be implied. Also, the average length of an individual autozygous segment in ESCS-3 is estimated to be ~20 cM (Wright et al. 1997). These considerations, coupled with the assumption that the number of autozygous regions is Poisson distributed, imply a probability of .46 that at least one such region is present. Furthermore, any such region with an average length of ~20 cM has a high probability of detection by a 386-marker (~10-cM) scan. Similar calculations in the case when only five of the six chromosomes have an autozygous region lead to the conclusion that at least one such region is almost certain to be present and to be detected by a 386-marker scan. The results in this family alone are therefore readily explicable on a chance basis.

However, the combined LOD score for all three families was 3.69 for D1S552 and 4.67 by multipoint analysis, which normally provides a conservative evaluation of the linkage likelihood. The question then arises as to whether the LOD score is influenced by underestimation of the extent of inbreeding in such consanguineous families, bearing in mind that, if apparently outbred CEPH kindreds show evidence of recent IBD (Broman and Weber 2000), this may well be underestimated in consanguineous kindreds. We therefore examined the effect of changing the inbreeding coefficient in the three ESCS families on the observed LOD score with marker D1S552 (data available on request). This shows that, depending on the disease allele q, the maximum likelihood LOD score declines from 3.69 to 2.97 by increasing the value of F by a factor of 2 (equivalent to making all founder members first cousins). In a genomewide linkage scan, the proposed threshold for demonstration of linkage is 3.3 ($P < .05$), so the LOD score falls below this threshold for several values of q and F.

In summary, a number of pitfalls in homozygosity mapping have come to light in the course of these studies. The first is the occurrence of allelic heterogeneity within a single consanguineous kindred, which can result in loss of homozygosity in flanking markers and failure to detect linkage. The second is the detection, in affected members of at least one family with ESCS, of autozygosity that is unrelated to the disease. The presence of

"hidden" consanguinity in pedigree founders may add to the likelihood of detection of such regions. Third, the usual safeguards against detecting false-positive linkage failed to prevent such an occurrence in this case. The simplest explanation is that a LOD score of 3.69 after a genomewide linkage scan is expected by chance in ~1/50 studies. The reassurance provided by the detection of large regions (>18 cM) of autozygosity flanking the linked marker, indicating IBD, in at least one of the families and smaller regions in all three families, also proved to be misleading. The probability of detecting such a region, cosegregating with disease in a small family, is high in the context of a whole-genome scan; therefore, caution is required in equating autozygosity with linkage. Finally, underestimation of the extent of inbreeding, which is suggested by the results of Broman and Weber's study (2000), can potentially inflate the LOD scores, increasing the chance of a false-positive linkage.

Acknowledgments

We acknowledge the National Institutes of Health (for grant EY05627), the Foundation Fighting Blindness, the Medical Research Council (UK), the Horvitz Family Foundation for its support, and Mr. David Hanna and Ms. Elaine deCastro for help with DNA sample collection. V.C.S. is an associate investigator at the Howard Hughes Medical Institute.

Electronic-Database Information

The accession number and URLs for data in this article are as follows:

Genetic Location Database, http://cedar.genetics.soton.ac.uk/public_html/ldb.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for ESCS [MIM 268100])

References

Broman KW, Weber JL (2000) Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. Am J Hum Genet 65:1493–1500

Collins A, Frezal J, Teague J, Morton NE (1996) A metric map of humans: 23,500 loci in 850 bands. Proc Natl Acad Sci USA 93:14771–14775

Haider NB, Jacobson SG, Cideciyan AV, Swiderski R, Streb LM, Searby C, Beck G, Hockey R, Hanna DB, Gorman S, Duhl D, Carmi R, Bennett J, Weleber RG, Fishman GA, Wright AF, Stone EM, Sheffield VC (2000) Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. Nat Genet 24:127–131

Wright AF, Teague PW, Bruford E, Carothers A (1997) Problems in dealing with linkage heterogeneity in autosomal recessive forms of retinitis pigmentosa. In: Edwards JH, Pawlowitzki IH, Thompson E (eds) Genetic mapping of disease genes. Academic Press, London, pp 255–272