The Effect of Inbreeding on the Distribution of Compound Heterozygotes: A Lesson from Lipase H Mutations in Autosomal Recessive Woolly Hair/Hypotrichosis

Lynn Petukhova^a Yutaka Shimomura^a Muhammad Wajid^a
Prakash Gorroochurn^d Susan E. Hodge^{b, d, e} Angela M. Christiano^{a, c}

Departments of ^aDermatology, ^bPsychiatry, and ^cGenetics & Development, and ^dDivision of Statistical Genetics, Department of Biostatistics, School of Public Health, Columbia University, and ^eDivision of Epidemiology, NY State Psychiatric Institute, New York, N.Y., USA

Key Words

LIPH • Autosomal recessive woolly hair • Consanguineous families • Compound heterozygous mutations

Abstract

Autozygosity mapping in consanguineous families has proven to be a powerful method for identifying recessive disease genes. Using this technique with whole genome SNP data generated from low density mapping arrays, we previously identified two genes that underlie autosomal recessive woolly hair (ARWH/hypotrichosis; OMIM278150), specifically *P2RY5* and Lipase H (*LIPH*). In the current study, we sought to identify a novel disease locus for ARWH/hypotrichosis by analyzing two large consanguineous families from Pakistan who had initially been excluded for mutations at either of these disease loci by haplotype analysis with microsatellite markers. A genome-wide analysis of 10 members from each of the two families failed to identify significant regions of autozygosity or linkage. Upon genotyping an additional 10 family members in one of the families, parametric linkage analysis identified a region on chromosome 3q27 with evidence for linkage ($Z = 2.5$). Surprisingly, this region contains the *LIPH* gene. Microsatellite markers located within the *LIPH* gene were used for haplotype analysis and demonstrated that not one, but two haplotypes were segregating with the

phenotype in each of these families. DNA sequencing identified two distinct *LIPH* mutations (280_369dup90 and 659_660delTA). Each affected individual ($n = 38$) was either homozygous for one mutation ($n = 7$ and 16 respectively), or compound heterozygous ($n = 15$). A review of the literature identified several reports of compound heterozygotes in consanguineous families. Prompted by this finding, we derived the probability that a patient affected with a recessive disease is carrying two mutations at the disease locus. We suggest that the validity of the IBD assumption may be challenged in large consanguineous families.

Copyright © 2009 S. Karger AG, Basel

Introduction

Gene mapping in consanguineous families provides a powerful method for identifying autosomal recessive genes. Inbreeding reduces the genetic variation in a population, so that when a rare recessive disease arises in a population with high consanguinity, it is more likely due to a mutation that is identical-by-descent (IBD), when compared to a population with no consanguinity. Auto-

L. Petukhova and Y. Shimomura contributed equally to this work.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2009 S. Karger AG, Basel
0001-5652/09/0682-0117\$26.00/0

Accessible online at:
www.karger.com/hhe

Angela M. Christiano, PhD
Department of Dermatology, Columbia University
College of Physicians & Surgeons
630 West 168th Street VC-15-204A, New York, NY 10032 (USA)
Tel. +1 212 305 9565, Fax +1 212 305 7391, E-Mail amc65@columbia.edu

zygosity mapping is a statistical technique that has been developed to identify regions of the genome that are shared among affected individuals and likely to be IBD. The success of utilizing highly inbred families to map-disease genes has been attributed to the fact that this strategy overcomes one of the largest obstacles in mapping studies, that is, genetic heterogeneity. Despite this commonly held notion, and the demonstrated power of the method, autozygosity may fail to determine the location of a disease locus when allelic heterogeneity is present. We encountered this scenario when performing mapping in two consanguineous families and were nevertheless successful in identifying the disease locus by extending our methodology to include parametric linkage analysis.

The use of SNP arrays has greatly increased the efficiency of gene mapping, particularly for Mendelian disorders. Even low density arrays (10K) have greatly increased marker density, while the ease and speed of sample preparation allows for a dramatic increase in genotyping throughput. For gene mapping in consanguineous families, we commonly use SNP array data to identify regions that provide evidence for IBD and recessive linkage. These intervals are then further refined with the use of more informative markers; we perform haplotype analysis with microsatellite data. Finally, direct sequencing of candidate genes leads to identification of the causative gene. By deploying this strategy, we have previously analyzed many consanguineous Pakistani families with autosomal recessive hair and nail disorders and have successfully identified homozygous mutations in causative genes, such as *HR* [1–3], *DSG4* [4–6], *RSPO4* [7, 8], *P2RY5* [9, 10], and *LIPH* [11].

Woolly hair (WH) refers to a group of hair shaft disorders that are characterized by fine and tightly curled hair. As compared with normal curly hair that is observed in some populations, WH grows slowly and stops growing after a few inches. Under light microscopy, WH shows some structural anomalies, such as trichorrexis nodosa and tapered ends. WH can appear as part of several syndromes, such as Naxos disease (OMIM 601214) and cardiofaciocutaneous syndrome (OMIM 115150). In addition to these syndromic forms of WH, isolated WH without associated findings (non-syndromic WH) has also been described. Non-syndromic WH can show either an autosomal dominant (ADWH; OMIM 194300) or recessive (ARWH; OMIM 278150) inheritance pattern.

We recently established the role of two genes in which mutations can underlie ARWH/hypotrichosis. Initially, we identified homozygous pathogenic mutations in the *P2RY5* gene in several families with ARWH/hypotricho-

sis [9]. The *P2RY5* gene encodes a G-protein coupled receptor (GPCR) known as P2Y5 and is a nested gene, residing within intron 17 of the retinoblastoma 1 (*RB1*) gene. *P2RY5* is expressed abundantly in both Henle's and Huxley's layers of the inner root sheath of the hair follicle [9]. More recently, we demonstrated that homozygous pathogenic mutations in *LIPH* underlie ARWH/hypotrichosis. Patients who carry mutations in *LIPH* are clinically indistinguishable from patients who carry mutations in *P2RY5* [11].

In this study, screening for patterns of IBD with microsatellite markers revealed two consanguineous ARWH/hypotrichosis families (Family 1 and Family 2) that had been excluded from linkage to either *P2RY5* or *LIPH*. These consanguineous families are unrelated and originate from two different regions in Pakistan that are separated by a geographic barrier. We selected 10 members from each of these families for a mapping study and performed whole-genome SNP genotyping. Our initial analysis showed that these data were uninformative for linkage and autozygosity. We therefore performed whole-genome genotyping on an additional 10 family members from Family 1 and repeated the analysis. While the evidence for linkage using a simple recessive model increased modestly on chromosome 3q ($Z = 2.5$), autozygosity mapping under the assumption of IBD remained uninformative. Placement of additional microsatellite markers located within *LIPH* allowed us to identify not one IBD disease allele, but rather, two different mutations that were segregating in each family. The presence of compound heterozygotes in these consanguineous families negates the assumption of IBD that is critical for successful autozygosity mapping. Our findings suggest that in addition to locus heterogeneity or reduced marker informativeness, an alternative hypothesis of compound heterozygosity should be considered.

Materials and Methods

DNA Samples

After obtaining informed consent, we collected peripheral blood samples in EDTA-containing tubes from members of Pakistani families and 100 population-matched unrelated, unaffected control individuals (under institutional approval and in adherence to the Declaration of Helsinki Principles). Genomic DNA was isolated from these samples according to standard techniques.

Genotyping

The Affymetrix GeneChip Human Mapping 10K 2.0 array was used to perform whole genome scans on individuals from two

consanguineous families. Sample preparation followed the Affymetrix 10K protocol. Hybridization was performed by the Columbia University Gene Chip Facility.

In order to confirm linkage to this region on chromosome 3q27, genomic DNA from family members was amplified by PCR using primers for microsatellite markers close to the *LIPH* gene. We analyzed three markers for the first analysis (D3S3592, D3S1602 and D3S1262), and four additional markers (LIPH-MS1–4) for the second analysis [11]. The amplification conditions for each PCR were 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, with a final extension at 72°C for 7 min. PCR products were run on 8% polyacrylamide gels and genotypes were assigned by visual inspection.

Linkage Analysis

Genespring GT (Agilent Software) was used for quality control measures and to perform a number of analyses. After removing SNPs that showed Mendelian inconsistencies, Genespring GT was used to infer haplotypes from the data. By using haplotypes rather than SNPs, we minimized the effect of linkage disequilibrium on multipoint linkage analysis, thus reducing Type I error. Initial analysis included whole-genome autozygosity mapping to identify regions of IBD that are shared among affected individuals. Details about the methodology employed by this test can be found at <https://www.chem.agilent.com/cag/bsp/products/gsgt/Downloads/pdf/autozygosity.pdf>. Multipoint parametric linkage analysis was performed on inferred haplotypes, assuming a recessive mode of inheritance with 100% penetrance and a disease allele frequency of 0.001. Others have previously demonstrated that misspecification of penetrance does not greatly affect power to detect linkage [12]. We therefore expect that reduced penetrance would not have significantly altered the conclusions.

Mutation Analysis of the LIPH Gene

Using genomic DNA from family members, all exons and exon-intron boundaries of the *LIPH* gene were amplified by PCR using gene-specific primers [11]. The amplified PCR products were run on 1.5% agarose gels, and purified with QIAquick Gel Extraction Kit (Qiagen). Subsequently, the products were directly sequenced in an ABI Prism 310 Automated Sequencer, using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems). Screening assays for the mutation 659_660delTA were performed as described previously [11].

Results

Clinical Features

We ascertained two large consanguineous families of Pakistani origin with 28 (Family 1) and 10 (Family 2) affected individuals, respectively. Pedigrees were consistent with autosomal recessive inheritance and show several inbreeding loops (fig. 1). All affected individuals in these families had tightly curled hair on their scalp from birth (fig. 1). The hair grew slowly and stopped growing after a few inches. The hair density is variable from normal to less dense, and the disease was nonsyndromic. For

all affected individuals, facial and body hair, teeth, nails, and sweating were normal, and palmoplantar hyperkeratosis was not evident. There was no family history of heart disease, cancers, sudden death or neurologic abnormalities.

Analysis of Genome-Wide SNP Data Shows Linkage to Chromosome 3 without Evidence for Autozygosity

Using genomic DNA of affected individuals from each of these two families, we first tested for linkage to the *P2RY5* gene using microsatellite markers [9] and excluded linkage to this region (data not shown). We next tested for linkage to the *LIPH* gene with 3 microsatellite markers that span a 1.8 Mb region, the two most proximal of which were located 0.8 Mb upstream and downstream from *LIPH* (D3S3592, D3S1602 and D3S1262). We saw no evidence for an IBD pattern consistent with an autozygous recessive inheritance among affected individuals. Instead, we found that some affected individuals were homozygous for the same haplotype, while others were heterozygous for distinct haplotypes. Because we assumed there should be IBD in such large consanguineous pedigrees, we thereby excluded both families from linkage to *LIPH*. Embarking on a whole genome scan to search for a third locus for ARWH/hypotrichosis, we genotyped 10 members of each family, using the Affymetrix GeneChip Human Mapping 10K array. Statistical analysis revealed no evidence for linkage or autozygosity. We then genotyped an additional 10 members for Family 1 and analyzing under a recessive model, we identified a region of suggestive linkage ($Z = 2.5$) on chromosome 3q26.3. Within the interval of linkage, the LOD score for autozygosity was -14.07 , indicating strong evidence against IBD alleles (fig. 2).

Haplotype Analysis with Additional Microsatellite Markers Identifies Two Distinct Haplotypes Segregating with ARWH/Hypotrichosis in Both Families

Since re-analysis of Family 1 with SNP array data on 20 individuals suggested linkage to the *LIPH* locus, we increased the density of microsatellite markers in this region. Four new markers, LIPH-MS1–4, were placed between D3S3592 and D3S1602 (fig. 3). Two of these markers are intronic: LIPH-MS2 and LIPH-MS3 are located respectively within intron 5 and intron 2 of the *LIPH* gene. We performed haplotype analysis and found two distinct haplotypes that were shared by affected members of both families. One is between D3S3592 and LIPH-MS4, and the other is between LIPH-MS1 and

Fig. 1. Clinical presentation. All affected individuals have features of ARWH/hypotrichosis, which is characterized by fine, tightly curled hair that grows slowly and stops growing after a few inches (A, B, C, and D). Hair density and hair shaft pigmentation phenotypes display variable expressivity. The patient in B displays sparse hair and the patient in C displays depigmented hair shafts.

The pedigrees shown in E (Family 1) and F (Family 2) show that disease segregation is consistent with autosomal recessive inheritance and there is a high degree of inbreeding in both families. Individuals who participated in this study are numbered in the pedigree.

LIPH-MS4, which we termed haplotypes A and B, respectively (fig. 3). Every affected individual in both families has one of three different combinations of these two haplotypes: homozygous for haplotype A; homozygous for haplotype B; or heterozygous for haplotypes A and B (fig. 3 and 5). This result suggested linkage to the *LIPH* gene in both families. Each consanguineous fam-

ily was segregating not one, but two distinct disease alleles.

Identification of Mutations in the *LIPH* Gene

On the basis of the results of our mapping studies, we sequenced the *LIPH* gene in both families. First, we amplified all exons and exon-intron boundaries of *LIPH* in

Fig. 2. Results of whole-genome genotyping in Family 1. The graph shows lod scores as a function of location within the genome. Chromosomes are indicated by alternating shades of gray. A total of 20 members of Family 1 were genotyped with a low density Affymetrix mapping array. Parametric linkage analysis under a recessive mode of inheritance identified a maximum LOD score of 2.5 on chromosome 3q26.33 (**A**). The maximum LOD score identified by autozygosity mapping did not coincide with a region of linkage (**B**).

several affected individuals and resolved the PCR products on 1.5% agarose gels. As shown in figure 4A, two distinct fragments were detected by PCR amplification of exon 2, and the fragment patterns were variable among affected individuals (fig. 5A). Direct sequencing of the shorter fragment, 540 bp in size, showed the wild type sequence (fig. 4B), while that of the longer one, 630 bp in size, contained a 90 bp insertion at position 369 of the *LIPH* gene, which is predicted to result in an in-frame 30-amino acid insertion (fig. 4C). The sequence of this insertion is a tandem duplication of the sequence between positions 280 and 369, thus designated 280_369dup90 (fig. 4C). We refer to this as mutation A.

Affected individuals who are homozygous for mutation A do not have any sequence variants in other exons of the *LIPH* gene. However, affected individuals who are heterozygous for mutation A have a second mutation in exon 5 of the *LIPH* gene. This second mutation consists of a 2-nucleotide deletion at positions 659 and 660, designated 659_660delTA (fig. 4D). We refer to this as mutation B. Furthermore, affected individuals who do not carry mutation A on either allele are homozygous for mutation B (fig. 5). Neither mutation A nor mutation B was detected in 100 unrelated healthy control individuals of Pakistani origin (data not shown). As shown in figure 5C and D, affected individuals are either homozygous for mutation A, homozygous for mutation B, or compound

heterozygous for mutations A and B. Unaffected individuals either have no mutation, or carry one of the mutations in the heterozygous state (fig. 5).

Discussion

Lipase H Mutations in Autosomal Recessive Woolly Hair

We and others recently demonstrated that two genes, *P2RY5* and *LIPH*, underlie ARWH/hypotrichosis [9–11, 13]. *LIPH* transcripts are abundantly and widely expressed in the human hair follicles. *LIPH* is a member of phosphatidic acid-selective phospholipase A1 and is a key enzyme in the synthesis of 2-acyl-lysophosphatidic acid (LPA) [14, 15], an extracellular mediator of many biological functions [16]. *P2Y5* is a LPA receptor [13], and thus provides a link between these discoveries. Collectively, these data suggest a crucial role of the *LIPH*/LPA/*P2Y5* signaling pathway in the pathogenesis of ARWH/hypotrichosis.

During these studies, we identified two large consanguineous families with ARWH/hypotrichosis. We inadvertently excluded *P2RY5* and *LIPH* because we incorrectly assumed that a single disease allele would be segregating in each family. We expected to map a third, novel disease locus in these ARWH/hypotrichosis fami-

Fig. 3. Haplotype analysis with densely spaced microsatellite markers was performed on all available family members. Results are presented for a subset of individuals from Family 1 and Family 2. Two distinct haplotypes are segregating with the disease in both families.

lies. Instead, we discovered that two *LIPH* disease alleles were segregating in each consanguineous family, and more than half of the genotyped affected individuals were compound heterozygous.

For linkage mapping in consanguineous families, we typically utilize genotyping data generated on low density SNP arrays and combine results from autozygosity mapping with parametric linkage analysis to identify candidate regions. By combining the results from these two statistical methods, we aim to reduce Type I error associated with each method individually. We have used this technique to successfully identify disease genes in several large consanguineous families [1–11].

At the onset of this study, we had excluded *LIPH* by haplotype analysis with genotype data from three microsatellite markers, the two most proximal of which were located 0.8 Mb upstream and downstream, respectively, from *LIPH*. Given the extent of inbreeding in these families, we expected these markers to be sufficient to detect true linkage. In fact, for ARWH/hypotrichosis families that segregate a single disease mutation in *LIPH*, these three markers were sufficient for establishing linkage and identifying IBD [11].

Having excluded autozygous linkage, we proceeded to perform whole-genome analysis with SNP data. For Family 1, we genotyped a total of 20 affected individuals and achieved a maximum LOD score for linkage to a recessive

Fig. 4. Identification of mutations. PCR amplification of exon 2 revealed the presence of two distinct fragments of 630 and 540 bp (A) in some affected individuals. Molecular weight marker is loaded in lane 1, lanes 2 and 3 contain PCR products from 2 different patients and lane 4 is from an unaffected family member. Direct

sequencing of the two PCR products showed that the shorter fragment is the wild type sequence (B), while the larger product contained a 90 bp insertion (C). ARWH/hypotrichosis patients who were heterozygous or null for the exon 2 mutation were found to carry a 2 bp deletion in exon 5 (D).

locus of 2.49 on chromosome 3q27 (fig. 2A). Autozygosity mapping was consistent with exclusion and produced a LOD score for IBD in the region of linkage of -14.07 . Across the genome, the maximum IBD LOD score was 1.34 on chromosome X, and the median score was -18.35 , indicating no evidence for IBD anywhere in the genome (fig. 2B). Despite the lack of evidence for autozygosity, because the region of linkage under a recessive model coincided with the *LIPH* locus, we increased the density of microsatellite markers and repeated haplotype analysis. We designed four new microsatellite markers, two of which are located within *LIPH*, and observed two haplotypes that were segregating with the disease, named

Haplotype A and Haplotype B (fig. 3). By direct sequencing of the *LIPH* gene in affected individuals, we identified two mutations: 280_369dup90 segregated with Haplotype A and 659_660delTA segregated with Haplotype B (fig. 4). All affected individuals were either homozygous for one of these mutations or compound heterozygous (fig. 5).

The mutation 280_369dup90 in the *LIPH* gene probably occurred through an in-frame duplication event of nucleotide sequences between positions 280 and 369, which results in the tandem repeat of 30 amino acid residues at the protein level. Pathogenic mutations due to in-frame duplication events have previously been reported in sev-

Fig. 5. Mutation analysis in all samples showed that ARWH/hypotrichosis patients are either homozygous for one of the mutations or compound heterozygotes. The unaffected family members were found to be heterozygous for one of the mutations or null for both of the mutations. For Family 2, visualization of PCR products from the amplification of exon 2 is presented in **A** and from exon 5 in **B**. Note that the PCR products from the 659_660delTA allele were digested with *Hpy188I* restriction enzyme. Red font indicates the presence of ARWH/hypotrichosis. Molecular weight marker is loaded in lane M and amplification of an unrelated unaffected control is loaded in lane C. In **C**, Family 2 results are indicated in the pedigree. In **D**, results of PCR analysis for Family 1 are indicated in the pedigree.

Fig. 6. The probability of observing a compound heterozygous mutation is plotted as a function of the number of disease alleles (n). The extent of inbreeding is indexed by an inbreeding coefficient, f , over a range of values within the theoretical boundaries (0, 1). The range of f values actually observed within inbred populations is highlighted in gray (0.05, 0.15). For a large, randomly mating population, $f=0$ and most recessive disease will be caused by compound heterozygous mutations. As the amount of inbreeding increases, this probability decreases, but will become 0 only when $f=1$. Part **A** shows the full range of probabilities, from 0 to 1. Part **B** focuses on the range of probabilities from 0 to 0.1, to allow interpolation for the *LIPH* locus ($n=11$).

Fig. 7. The effect of compound heterozygous mutations on linkage analysis (black) and autozygosity mapping (gray). The graphs show lod scores as a function of location for chromosome 3, in two consanguineous pedigrees affected with ARWH/hypotrichosis. *LIPH* is located at 187Mb. In **A**, results are presented from Family 1, shown to be segregating two disease mutations. Strong evidence against IBD alleles is found in a region of moderate linkage ($Z = 2.5$). In **B**, results are presented from analysis of an ARWH family found to be segregating a single mutation. Evidence for autozygosity coincides with weak evidence for linkage ($Z = 0.6$).

eral other genes, such as *HOXA13* [17], *HOXD13* [18], and *K6a* [19]. The additional amino acid sequences in the N-terminal catalytic domain would severely affect the structure and/or function of the lipase H protein. The mutation 659_660delTA in the *LIPH* gene leads to a frameshift and a premature termination codon 25 amino acid residues downstream from the deletion (Ile220ArgfsX25). A *LIPH* transcript with this mutation is predicted to be largely degraded through nonsense-mediated mRNA decay. This recurrent mutation was recently identified in several Pakistani families with ARWH/hypotrichosis [11, 20].

The haplotype associated with the mutation 280_369dup90, is at least 800kb larger than the haplotype associated with the mutation 659_660delTA (fig. 3). In general, haplotypes associated with novel mutations are quite large. Over time, recombination events reduce the number of loci in linkage disequilibrium with the mutation and the haplotype is reduced in size. We also expect that older mutations will be more frequent in the population than ones that arose more recently. To date, we and others have identified *LIPH* mutations in several ARWH/hypotrichosis families, some of which are recurrent at

different frequencies. The mutation 659_660delTA is segregating in 6 families [11]. However, the mutation 280_369dup90 has only been identified in the two families we report here. Based on these two observations: (1) the length of the homozygous tracts surrounding the mutations, and (2) the frequency of the mutations in the population, it appears that the 280_369dup90 mutation has emerged more recently than the mutation 659_660delTA. The fact that this mutation exists in a much higher frequency in Family 2 than in Family 1 could possibly indicate that it originated in founders of Family 2 and was more recently introduced into Family 1. Alternatively, genetic drift, which is known to have stronger effects in highly inbred populations, may explain this observation.

The two families under investigation in our study originate from two different provinces in Pakistan and are separated by approximately 1,100 km and a river that serves as a geographical barrier. In addition, these two families belong to different tribes and speak distinct languages (Family 1 speaks Sindi; Family 2 speaks Saraiki). It is believed that these two families have had no common

Table 1. Published reports of compound heterozygous mutations segregating in consanguineous families

First author	Year	Journal	Population	Disease	Gene (MIM)	Total published counts	
						variants in reported gene	genes implicated in disease
Carrasquillo	1997	Hum Mol Gen	Israeli-Arab	Nonsyndromic autosomal recessive deafness	GJB2 (121011)	+50	28
Coucke	1999	J Med Genet	Turkey	Pendred syndrome (hearing, thyroid involvement: goiter)	SLC26A4 (605646)	29	1
Pannain	1999	J Clin Endocrinol and Metab	Amish of Swiss descent	Congenital hypothyroidism	TPO (606765)	12	1
Bolk	1999	Am J Hum Genet	Amish	Congenital nephrotic syndrome	nephrin (602716)	+80	1
Miano	2000	Am J Hum Genet	North American	Enhanced S-cone syndrome	NR2E3 (604485)	6	1
Gershoni-Baruch	2002	Am J Med Genet	Israel, of Muslim Arab descent	Familial Mediterranean Fever	MEFV (608107)	20	1
deBrouwer	2003	Hum Genet	Dutch	Nonsyndromic autosomal recessive deafness	GJB2 (121011) CDH23 (114021)	+50 +36	28
Ducroq	2006	Eur J Hum Genet	Israel, of Christian Arab descent	Cone-rod dystrophy (may include retinal pigmentation)	ABCA4 (601691)	29	3
Farah	2006	Thromb Haemost	Lebanon	F5F8D (coagulation)	LMAN1 (601567)	24	2
Laurier	2006	Eur J Hum Genet	Lebanon	Bardet-Biedl Syndrome (renal abnormalities, plus retinal dystrophy, polydactyly, mental retardation, and mild obesity)	BBS10 (610148)	118	12
Frishberg	2007	Genet Med	Israel, of Muslim Arab descent	Congenital nephrotic syndrome	nephrin (602716)	+80	1
Seidel	2007	Eur J Pediatr	Turkey	Familial Mediterranean Fever	MEFV (608107)	20	1
Lezirovitz	2008	Eur J Hum Genet	Brazil	Nonsyndromic autosomal recessive deafness	MYO15A (602666)	+14	28
Petukhova	2008	Hum Hered	Pakistani	Autosomal recessive woolly hair	LIPH (607365)	11	2

ancestors for at least 150 years, suggesting that these mutations arose more than 6 generations ago.

The Occurrence of Compound Heterozygotes in Consanguineous Families

There are two main advantages to studying recessive diseases in highly inbred populations. First, inbreeding increases the prevalence of recessive diseases. Within an inbred population, the number of affected individuals is proportional to the disease allele frequency, whereas in a large, randomly mating population, the number of affecteds is proportional to the square of the disease allele frequency [21]. Thus, inbreeding increases the efficiency of identifying patients to study.

A second reason for gene mapping in consanguineous families is the expectation of a reduction in genetic heterogeneity. Our a priori assumption is that when a recessive disease arises in a consanguineous family, it must be

due to a mutation that is IBD. In fact, if we consider a consanguineous family for which there are only two founders, and in which a recessive disease has arisen, the probability that the disease is being caused by a single mutation is determined by the frequency of that allele in the population, while the probability that the disease is due to two different mutations at the same locus is determined by the product of the respective allele frequencies. For example, if the disease allele frequencies are each on the order of 0.001, then the probability that a consanguineous family affected with a recessive disease is segregating both mutations is only one in a million or 1×10^{-6} . Thus, it would appear that our surprise at discovering compound heterozygous mutations in two consanguineous kindreds was justified.

Our initial gene mapping efforts were confounded because our initial assumption of IBD was, in fact, incorrect. An analysis of the literature uncovered at least 13

other such reports, summarized in table 1. A number of these publications also report surprise at their findings and no single report provides a comprehensive list of prior studies with similar findings. The earliest report that we identified indicated that allelic diversity may be a consequence of extensive inbreeding in populations that descend from a small number of founders, and furthermore predicted that this could be 'a persistent feature of many inbred communities' [22].

We next sought to precisely define how often one would expect to find compound heterozygous individuals in consanguineous families. In a large randomly mating population, the proportion of people with a recessive disease caused by a compound heterozygous mutation, in contrast to one caused by a homozygous mutation, is a function of the number of disease alleles, as well as the frequencies of these alleles. We can assume that because the population is large and mating randomly, the force of genetic drift is relatively weak and selection acts to remove disease-causing mutations. Therefore, in the absence of heterozygote advantage, the frequencies of disease mutations are kept small and we will assume equal. If n disease-causing mutations exist in a gene, and each allele exists at frequency q , then the proportion of affected individuals that are compound heterozygous is given by the following formula:

$$\frac{\binom{n}{2} 2q^2}{\binom{n}{2} 2q^2 + nq^2} = \frac{n-1}{n}$$

The allele frequency cancels out and the probability remains a function of only the number of disease alleles. It converges to 1 as $n \rightarrow \infty$, which means that as the number of disease alleles increases, the probability that the disease is caused by a compound heterozygous mutation increases. For example, if a locus has 10 disease alleles, there is a 90% chance that a patient is carrying a compound heterozygous mutation.

In the presence of inbreeding, with an inbreeding coefficient of f , the above formula becomes:

$$\frac{\binom{n}{2} 2q^2 (1-f)}{\binom{n}{2} 2q^2 (1-f) + nq^2 (1-f) + nqf} = \frac{(n-1)q(1-f)}{nq(1-f) + f}$$

Now the probability that a patient carries a compound heterozygous mutation depends not only on the number of disease alleles (n) as in the first equation, but also on the allele frequencies (q) and the extent of inbreeding (f).

This formula captures our expectations about gene mapping in consanguineous families. As f approaches 1, the probability that an affected person is a compound heterozygote approaches zero, and thus the probability of homozygosity for a single disease allele approaches 1. Furthermore, a small increase in the amount of inbreeding results in a dramatic decrease in the probability of observing a compound heterozygote. For example, to observe the effect of introducing inbreeding, we can consider a locus with 10 disease alleles. The probability that a patient with a recessive disease is carrying two different mutations at the disease locus drops from 90% when $f = 0$ to 7.5% when $f = 0.1$. More often than not, inbreeding will produce affected individuals who are homozygous for a disease allele.

Figure 6 plots the probability that a recessive disease is caused by a compound heterozygous genotype as a function of the number of disease alleles at a given locus for a range of inbreeding coefficients (f) within the theoretical boundaries of this statistic (0, 1). The frequencies of all disease alleles are fixed at 0.001. When f is set equal to 0 (plotted in black), there is an absence of inbreeding, such as would be observed in a large randomly mating population. The probability of identifying a compound heterozygous patient rapidly approaches 1 as the number of disease alleles increases.

Interestingly, although the theoretical bounds of f are (0, 1), the range that is observed in actual populations is much smaller (0, 0.15). In figure 6, we have indicated a range that is found in inbred populations by gray shading. One of the most highly inbred animal populations is reported to have had an inbreeding coefficient of 0.149 [23]. Human populations in general have lower levels of inbreeding than animal populations and a recent publication that reported estimates of human inbreeding coefficients found a range of 0.000 to 0.125 within their relatively small sample [24]. From our studies of ARWH/hypotrichosis, we have identified a total of 11 unique mutations at the LIPH locus. From figure 6B, the probability of identifying a compound heterozygote is therefore between 5 and 16%. While inbreeding reduces the probability that a single mutation will be found at a recessive disease locus in an affected individual, the chance of ascertaining a compound heterozygote is not unlikely, but is, in fact, a highly viable alternative hypothesis.

Carrasquillo et al. [22] attributed the discovery of compound heterozygotes in a consanguineous family to the increased force of genetic drift in inbred populations. Two consequences of increased genetic drift are (1) an increase in the number of disease alleles, and (2) an in-

crease in the frequencies of disease alleles. We have demonstrated here that these two increases are positively correlated with the probability of observing compound heterozygote patients. Furthermore, small changes in the amount of inbreeding within a population will greatly impact this probability. In light of these considerations, the violation of the IBD assumption observed in our study should not have been unexpected.

Conclusion

We report here two mutations in the *LIPH* gene that underlie ARWH/hypotrichosis and are segregating together in two unrelated consanguineous families from Pakistan. This finding should caution against making the a priori assumption of IBD in consanguineous families. Our data illustrate that evidence for linkage, in the absence of evidence for homozygosity, points to the possibility of allelic heterogeneity (fig. 7). Parametric linkage analysis is more robust to a violation of the IBD assumption than autozygosity mapping, and it is important to consider results from both methods when gene mapping in consanguineous families. Finally, we have demonstrated that the assumption of IBD is sensitive to the true amount of inbreeding, which is often difficult to estimate accurately, as well as the number of disease alleles and the frequencies of those alleles.

Our initial efforts at gene mapping in two consanguineous families were inconclusive because we had incorrectly assumed that the disease alleles were IBD. In table 1, we have summarized publications that report finding compound heterozygote patients in consanguineous families, and draw attention to the fact that half of these papers report on a disease known to be caused by a

single gene. In such a situation, investigators are naturally prompted to question microsatellite data that excludes linkage to an IBD disease allele and are consequently more likely to identify compound heterozygosity. It is not possible to determine how often failure to observe a pattern of IBD among patients from a consanguineous family is incorrectly attributed to locus heterogeneity in the disease.

Our report demonstrates that compound heterozygosity is a viable and reasonable explanation for such null findings. We are not recommending that gene mapping in consanguineous families should be initiated with an expectation of compound heterozygosity. However, we wish to point out that investigators should be aware that locus heterogeneity or reduced marker polymorphisms are not the only alternative hypotheses to be explored. Furthermore, exploring an alternative of compound heterozygosity may be more reasonable than assuming these other explanations.

Gene mapping in consanguineous families is clearly one of the most powerful methods for identifying recessive disease genes. While it is easy to accept without question assumptions that underlie methods with demonstrated success, our report illustrates the importance of remaining vigilant for the validity of commonly held assumptions.

Acknowledgements

We are grateful to the family members for their participation in this study, to Ha Mut Lam for technical assistance, and to Krzysztof Koryluk for helpful discussions. This work was supported by NIDDK grant DK-31813 (to S.E.H.) and USPHS NIH/NIAMS grant R01AR44924 (to A.M.C.).

References

- Ahmad W, Faiyaz ul Haque M, Brancolini V, Tsou HC, ul Haque S, Lam H, Aita VM, Owen J, deBlaquiere M, Frank J, Cserhalmi-Friedman PB, Leask A, McGrath JA, Peacocke M, Ahmad M, Ott J, Christiano AM: Alopecia universalis associated with a mutation in the human hairless gene. *Science* 1998;279:720–724.
- Ahmad W, Zlotogorski A, Panteleyev AA, Lam H, Ahmad M, ul Haque MF, Abdallah HM, Dragan L, Christiano AM: Genomic organization of the human hairless gene (*hr*) and identification of a mutation underlying congenital atrichia in an Arab Palestinian family. *Genomics* 1999;56:141–148.
- Kim H, Wajid M, Kraemer L, Shimomura Y, Christiano AM: Nonsense mutations in the hairless gene underlie apl in five families of Pakistani origin. *J Dermatol Sci* 2007;48:207–211.
- Kljuic A, Bazzi H, Sundberg JP, Martinez-Mir A, O'Shaughnessy R, Mahoney MG, Levy M, Montagutelli X, Ahmad W, Aita VM, Gordon D, Uitto J, Whiting D, Ott J, Fischer S, Gilliam TC, Jahoda CA, Morris RJ, Panteleyev AA, Nguyen VT, Christiano AM: Desmoglein 4 in hair follicle differentiation and epidermal adhesion: Evidence from inherited hypotrichosis and acquired pemphigus vulgaris. *Cell* 2003;113:249–260.
- Schaffer JV, Bazzi H, Vitebsky A, Witkiewicz A, Kovich OI, Kamino H, Shapiro LS, Amin SP, Orlow SJ, Christiano AM: Mutations in the desmoglein 4 gene underlie localized autosomal recessive hypotrichosis with monilethrix hairs and congenital scalp erosions. *J Invest Dermatol* 2006;126:1286–1291.
- Wajid M, Bazzi H, Rockey J, Lubetkin J, Zlotogorski A, Christiano AM: Localized autosomal recessive hypotrichosis due to a frameshift mutation in the desmoglein 4 gene exhibits extensive phenotypic variability within a Pakistani family. *J Invest Dermatol* 2007;127:1779–1782.

- 7 Blaydon DC, Ishii Y, O'Toole EA, Unsworth HC, Teh MT, Ruschendorf F, Sinclair C, Hopsu-Havu VK, Tidman N, Moss C, Watson R, de Berker D, Wajid M, Christiano AM, Kelsell DP: The gene encoding r-spondin 4 (rspo4), a secreted protein implicated in wnt signaling, is mutated in inherited anonychia. *Nat Genet* 2006;38:1245–1247.
- 8 Ishii Y, Wajid M, Bazzi H, Fantauzzo KA, Barber AG, Blaydon DC, Nam JS, Yoon JK, Kelsell DP, Christiano AM: Mutations in r-spondin 4 (rspo4) underlie inherited anonychia. *J Invest Dermatol* 2008;128:867–870.
- 9 Shimomura Y, Wajid M, Ishii Y, Shapiro L, Petukhova L, Gordon D, Christiano AM: Disruption of p2ry5, an orphan G protein-coupled receptor, underlies autosomal recessive woolly hair. *Nat Genet* 2008;40:335–339.
- 10 Petukhova L, Sousa, EC, Martinez-Mir, A, Vitebsky, A, dos Santos, LG, Shapiro, L, Haynes, C, Gordon, D, Shimomura, Y, and Christiano, AM: Genome-wide linkage analysis of an autosomal recessive hypotrichosis identifies a novel p2ry5 mutation. *Genomics* 2008;92:273–278.
- 11 Shimomura Y, Wajid M, Petukhova L, Shapiro L, Christiano AM: Mutations in the lipase H (liph) gene underlie autosomal recessive woolly hair/hypotrichosis. *J Invest Dermatol* 2008;129:622–628.
- 12 Greenberg DA: Inferring mode of inheritance by comparison of lod scores. *Am J Med Genet* 1989;34:480–486.
- 13 Jin W, Broedl UC, Monajemi H, Glick JM, Rader DJ: Lipase h, a new member of the triglyceride lipase family synthesized by the intestine. *Genomics* 2002;80:268–273.
- 14 Sonoda H, Aoki J, Hiramatsu T, Ishida M, Bandoh K, Nagai Y, Taguchi R, Inoue K, Arai H: A novel phosphatidic acid-selective phospholipase a1 that produces lysophosphatidic acid. *J Biol Chem* 2002;277:34254–34263.
- 15 Moolenaar WH, van Meeteren LA, Giepmans BN: The ins and outs of lysophosphatidic acid signaling. *Bioessays* 2004;26:870–881.
- 16 Pasternack SM, von Kugelgen I, Aboud KA, Lee YA, Ruschendorf F, Voss K, Hillmer AM, Molderings GJ, Franz T, Ramirez A, Nurnberg P, Nothen MM, Betz RC: G protein-coupled receptor p2y5 and its ligand LPA are involved in maintenance of human hair growth. *Nat Genet* 2008;40:329–334.
- 17 Mortlock DP, Innis JW: Mutation of hoxa13 in hand-foot-genital syndrome. *Nat Genet* 1997;15:179–180.
- 18 Muragaki Y, Mundlos S, Upton J, Olsen BR: Altered growth and branching patterns in synpolydactyly caused by mutations in hoxd13. *Science* 1996;272:548–551.
- 19 Smith FJ, Liao H, Cassidy AJ, Stewart A, Hamill KJ, Wood P, Joval I, van Steensel MA, Bjorck E, Callif-Daley F, Pals G, Collins P, Leachman SA, Munro CS, McLean WH: The genetic basis of pachyonychia congenita. *J Invest Dermatol Symp Proc* 2005;10:21–30.
- 20 Jelani M, Wasif N, Ali G, Chishti M, Ahmad W: A novel deletion mutation in liph gene causes autosomal recessive hypotrichosis (lah2). *Clin Genet* 2008;74:184–188.
- 21 Lander ES, Botstein D: Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children. *Science* 1987;236:1567–1570.
- 22 Carrasquillo MM, Zlotogora J, Barges S, Chakravarti A: Two different connexin 26 mutations in an inbred kindred segregating non-syndromic recessive deafness: Implications for genetic studies in isolated populations. *Hum Mol Genet* 1997;6:2163–2172.
- 23 Templeton AR: *Population Genetics and Microevolutionary Theory*. Hoboken, N.J., Wiley-Liss, 2006.
- 24 Carothers AD, Rudan I, Kolcic I, Polasek O, Hayward C, Wright AF, Campbell H, Teague P, Hastie ND, Weber JL: Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. *Ann Hum Genet* 2006;70:666–676.