

ARTICLE

# Challenges and solutions for gene identification in the presence of familial locus heterogeneity

Atteeq U Rehman<sup>1,12</sup>, Regie Lyn P Santos-Cortez<sup>2,12</sup>, Meghan C Drummond<sup>1</sup>, Mohsin Shahzad<sup>3,4</sup>, Kwanghyuk Lee<sup>2</sup>, Robert J Morell<sup>1</sup>, Muhammad Ansar<sup>2,5</sup>, Abid Jan<sup>5</sup>, Xin Wang<sup>2</sup>, Abdul Aziz<sup>5</sup>, Saima Riazuddin<sup>3,4</sup>, Joshua D Smith<sup>6</sup>, Gao T Wang<sup>2</sup>, Zubair M Ahmed<sup>4</sup>, Khitab Gul<sup>3</sup>, A Eliot Shearer<sup>7</sup>, Richard J H Smith<sup>7</sup>, Jay Shendure<sup>6</sup>, Michael J Bamshad<sup>6</sup>, Deborah A Nickerson<sup>6</sup>, University of Washington Center for Mendelian Genomics<sup>13</sup>, John Hinnant<sup>8,14</sup>, Shaheen N Khan<sup>3</sup>, Rachel A Fisher<sup>9</sup>, Wasim Ahmad<sup>5</sup>, Karen H Friderici<sup>9,10</sup>, Sheikh Riazuddin<sup>3,11</sup>, Thomas B Friedman<sup>1</sup>, Ellen S Wilch<sup>10</sup> and Suzanne M Leal<sup>\*,2</sup>

Next-generation sequencing (NGS) of exomes and genomes has accelerated the identification of genes involved in Mendelian phenotypes. However, many NGS studies fall short of identifying causal variants, with estimates for success rates as low as 25% for uncovering the pathological variant underlying disease etiology. An important reason for such failures is familial locus heterogeneity, where within a single pedigree causal variants in two or more genes underlie Mendelian trait etiology. As examples of intra- and inter-sibship familial locus heterogeneity, we present 10 consanguineous Pakistani families segregating hearing impairment due to homozygous variants in two different hearing impairment genes and a European-American pedigree in which hearing impairment is caused by four variants in three different genes. We have identified 41 additional pedigrees with syndromic and nonsyndromic hearing impairment for which a single previously reported hearing impairment gene has been identified but only segregates with the phenotype in a subset of affected pedigree members. We estimate that locus heterogeneity occurs in 15.3% (95% confidence interval: 11.9%, 19.9%) of the families in our collection. We demonstrate novel approaches to apply linkage analysis and homozygosity mapping (for autosomal recessive consanguineous pedigrees), which can be used to detect locus heterogeneity using either NGS or SNP array data. Results from linkage analysis and homozygosity mapping can also be used to group sibships or individuals most likely to be segregating the same causal variants and thereby increase the success rate of gene identification.

*European Journal of Human Genetics* (2015) 23, 1207–1215; doi:10.1038/ejhg.2014.266; published online 10 December 2014

## INTRODUCTION

Identification of genetic variants that cause Mendelian disorders, which segregate in large families, has been facilitated through linkage analysis coupled with Sanger sequencing.<sup>1–3</sup> In the past few years, next-generation sequencing (NGS) has supplanted this experimental strategy for variant detection for monogenic traits,<sup>4</sup> and many articles report successful gene identification.<sup>5–7</sup> Failures are rarely published. Estimates of the success rate for gene identification of Mendelian traits using NGS in clinical settings is as low as 25%.<sup>8</sup> However, for studies using large pedigrees with Mendelian segregation, the gene identification success rate can be much higher, for example, in a study of 24 families with multiple affected individuals the success rate was 60% (95% confidence interval (CI): 36%, 78%).<sup>9</sup>

A frequent strategy used for Mendelian trait gene identification is to select DNA samples from one or more affected or both affected

and unaffected family members and perform NGS. If multiple family members have undergone NGS, filtering is then performed based upon variant sharing in affected family members and lack of sharing in unaffected family members. Additional filtering is performed using variant databases such as Exome Variant Server or 1000 Genomes retaining low-frequency variants (eg, <0.1%) that are predicted to be deleterious by bioinformatics tools. The selected variants are then tested for co-segregation with the phenotype in the entire family. However, this approach is based on the assumptions that clinical information is reliable, the disease is fully penetrant, no phenocopies exist and there is locus homogeneity. If these conditions do not hold, identification of the causal variant can be problematic, because segregation with disease will not be observed. In our cohort of pedigrees segregating hearing impairment (HI), initially unrecognized locus heterogeneity was

<sup>1</sup>Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, USA; <sup>2</sup>Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; <sup>3</sup>National Center of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan; <sup>4</sup>Department of Otorhinolaryngology Head and Neck Surgery, School of Medicine, University of Maryland, Baltimore, MD, USA; <sup>5</sup>Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan; <sup>6</sup>Department of Genome Sciences, University of Washington, Seattle, WA, USA; <sup>7</sup>Molecular Otolaryngology and Renal Research Labs, Department of Otolaryngology—Head and Neck Surgery, University of Iowa, Iowa City, IA, USA; <sup>8</sup>Department of Religious Studies, Michigan State University, East Lansing, MI, USA; <sup>9</sup>Department of Pediatrics and Human Development, Michigan State University, East Lansing, MI, USA; <sup>10</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA; <sup>11</sup>Allama Iqbal Medical College-Jinnah Hospital Complex, University of Health Sciences, Lahore, Pakistan

\*Correspondence: Professor SM Leal, Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, 1 Baylor Plaza 700D, Houston, TX 77030, USA. Tel: +1 713 798 4011; Fax: +1 713 798 4012; E-mail: sleal@bcm.edu

<sup>12</sup>These authors contributed equally to this work.

<sup>13</sup>See appendix.

<sup>14</sup>Present address: Jalan Mawar, Kalibukbuk, Lovina, Bali, Indonesia 81119.

Received 8 August 2014; revised 25 October 2014; accepted 31 October 2014; published online 10 December 2014

found to be a recurring hindrance to gene identification. In the presence of locus heterogeneity, the causal variant may be falsely rejected as non-causal or may lie outside the mapped region, because affected individuals in different branches or even in the same sibship do not segregate the same causal variant.

Here we present 10 consanguineous families from Pakistan and one European-American family segregating fully penetrant HI. For each of

the 10 consanguineous families, homozygous putatively causal variants were identified in two different HI genes, while four variants in three different genes underlie HI etiology in the European-American family. We propose strategies by which locus heterogeneity can be detected in pedigrees, and this information, in turn, can facilitate gene identification.

## MATERIALS AND METHODS

### Institutional Review Board approval

Before the onset of our study, Institutional Review Board (IRB) approval was acquired from: Baylor College of Medicine and Affiliated Hospitals, Cincinnati Children's Hospital Research Foundation, Combined Neuroscience IRB at the National Institutes of Health, Michigan State University, National Centre of Excellence in Molecular Biology at the University of the Punjab, Quaid-i-Azam University, and University of Iowa. Written informed consent was obtained from all study participants.

### Screening and identification of HI variants

Different methods were used to screen for and identify putatively causal HI variants, including PCR, restriction fragment length polymorphism genotyping, Sanger sequencing, comparative genome hybridization and NGS. Table 1 lists

**Table 1 Genes and variants screened in 11 hearing impairment families**

Gene <sup>a</sup>	Exon or variant	Screening method	Families screened
<i>GJB2</i>	Exon 2	Sanger sequencing	1–11
<i>GJB2</i>	c.35delG	RFLP	6
<i>CIB2</i>	c.272T>C (p.(Phe91Ser))	Sanger sequencing	1–5, 7–11
<i>HGF</i>	c.482 +1986_8delTGA	Sanger sequencing	1–5, 7–11
<i>SLC26A4</i>	c.1337A>G (p.(Gln446Arg))	Sanger sequencing	1–4

Abbreviation: RFLP, restriction fragment length polymorphism.

<sup>a</sup>MIM and accession numbers: *GJB2*, MIM 121011, NM\_004004.5; *CIB2*, MIM 605564, NM\_006383.3; *HGF*, MIM 142409, NM\_000601.4; *SLC26A4*, MIM 605646, NM\_000441.1.

**Table 2 Variant information for 11 families with hearing impairment and locus heterogeneity**

Family	Method of identification	Sequenced gene	Variant	Gene region
1	Sanger sequencing	<i>GJB2</i>	c.71G>A (p.(Trp24*))	13q12.11
	Sanger sequencing	<i>HGF</i>	c.482 +1986_8delTGA	7q21.11
2	Sanger sequencing	<i>GJB2</i>	c.231G>A (p.(Trp77*))	13q12.11
	Sanger sequencing	<i>HGF</i>	c.482 +1986_8delTGA	7q21.11
3	Sanger sequencing	<i>GJB2</i>	c.71G>A (p.(Trp24*))	13q12.11
	Sanger sequencing	<i>SLC26A4</i>	c.1337A>G (p.(Gln446Arg))	7q22.3
4	Sanger sequencing	<i>GJB2</i>	c.231G>A (p.(Trp77*))	13q12.11
	Sanger sequencing	<i>SLC26A4</i>	c.1337A>G (p.(Gln446Arg))	7q22.3
5	Sanger sequencing	<i>GJB2</i>	c.71G>A (p.(Trp24*))	13q12.11
	Sanger sequencing	<i>MYO7A</i> <sup>a</sup>	c.397dupC (p.(His133Profs*7))	11q13.5
6 <sup>b</sup>	Sanger or RFLP	<i>GJB2</i>	c.35delG (p.(Gly12Valfs*2))	13q12.11
	Array CGH or PCR	<i>GJB2</i>	131-kb deletion	13q12.11
	RFLP	<i>SLC26A4</i> <sup>c</sup>	c.1334T>G (p.(Leu445Trp))	7q22.3
	PCR	<i>POU3F4</i> <sup>d</sup>	3.1-Mb deletion	Xq21.1
7	Sanger sequencing	<i>HGF</i>	c.482 +1986_8delTGA	7q21.11
	Sanger sequencing	<i>TMCI</i> <sup>e</sup>	c.1114G>A (p.(Val372Met))	9q21.13
8	Sanger sequencing	<i>GJB2</i>	c.35delG (p.(Gly12Valfs*2))	13q12.11
	Sanger sequencing	<i>GJB2</i>	c.71G>A (p.(Trp24*))	13q12.11
9	Exome NGS <sup>f</sup>	<i>SLC26A4</i>	c.1337A>G (p.(Gln446Arg))	7q22.3
	Sanger sequencing	<i>CIB2</i>	c.272T>C (p.(Phe91Ser))	15q25.1
10	Exome NGS <sup>f</sup>	<i>SLC26A4</i>	c.716T>A (p.(Val239Asp))	7q22.3
	Exome NGS	<i>LHFPL5</i>	c.246delC (p.(Leu84fs*1)) <sup>g</sup>	6p21.31
11	Sanger sequencing	<i>GJB2</i>	c.71G>A (p.(Trp24*))	13q12.11
	Targeted NGS <sup>h</sup>	<i>SLC26A4</i>	c.691G>A (p.(Val231Met))	7q22.3

Abbreviation: CGH, comparative genomic hybridization.

<sup>a</sup>Carriers of this *MYO7A* (MIM 276903, NM\_000260.3) variant<sup>18</sup> were re-evaluated for retinitis pigmentosa and reclassified to have Usher syndrome.

<sup>b</sup>In the subpedigree of family 6 shown in Figure 2, there are 17 genotyped individuals with NSHI, of which 9 are homozygous for *GJB2* c.35delG, while 4 are compound heterozygous for *GJB2* c.35delG and a 131-kb deletion that is telomeric to *GJB2* and *GJB6* and that reduces mRNA expression of both genes.<sup>31</sup>

<sup>c</sup>Individual 99 of family 6 was previously reported to have enlarged vestibular aqueducts and is homozygous for *SLC26A4* c.1334T>G (p.(Leu445Trp)).<sup>19</sup> Hearing-impaired individual 82 is heterozygous for *SLC26A4* c.2171A>G (p.(Asp724Gly)), but a second variant was not identified.

<sup>d</sup>In individual 93 of family 6, temporal bone imaging revealed bilateral communication between the cochlear basal turns and internal auditory canals (IAC), with distal patulous IAC dilatation, absent modioli, dysplastic vestibules and dysmorphic lateral semicircular canals. These findings are consistent with loss-of-function of *POU3F4* (MIM 300039, NM\_000307.4).<sup>20</sup> The approximate boundaries of the novel 3.1-Mb deletion at the *DFNX2* locus (chrX:80.5–86.6 Mb), which includes *POU3F4*, was initially localized by performing PCR across the region, and then the deletion breakpoints were mapped using restriction digest analysis of a 15-kb PCR product and by Sanger sequencing.

<sup>e</sup>When branch 2 of family 7 was mapped to the 9q region, all coding exons of *TMCI* (MIM 606706, NM\_138691.2) were sequenced.

<sup>f</sup>For family 10, sequence capture was performed with the SeqCap EZ Human Exome Library v2.0 (~36.6-Mb target) while for families 8 and 9, EZ Exome v3.0 (~64-Mb target) was used.

Sequencing was performed using an Illumina HiSeq to average read depths of ~65× for individual 33 of family 8 and individual 19 of family 9 and 92× for individual 12 of family 10.

<sup>g</sup>This *LHFPL5*<sup>32</sup> (MIM 609427, NM\_182548.1) variant has not been reported previously, was not found in 1000 Genomes or the Exome Variant Server, was not identified in 140 ethnically matched control chromosomes and is predicted by MutationTaster<sup>33</sup> to initiate nonsense-mediated mRNA decay.

<sup>h</sup>For family 11, the *SLC26A4* variant was identified by using a gDNA library that contained probes for coding exons of all known nonsyndromic and selected syndromic HI genes, and then using DNA from individual 7, NGS was performed on a SOLiD5500.

the families screened and the methods used for initial evaluation of specific genes and variants. These screening methods led to the discovery of some but not all the variants identified in each family. All causal variants and the method by which each variant was identified in the families are listed in Table 2.

### Evaluating pedigree informativeness

Simulation studies were performed to evaluate the maximum LOD (MLOD) scores for the Pakistani pedigrees, which have an autosomal recessive mode of inheritance. Using SLINK,<sup>10</sup> replicates were generated for each given pedigree structure conditional on affection status by generating a marker with a minor allele frequency of 0.0001 completely linked (recombination fraction ( $\theta$ ) = 0.0) to a fully penetrant autosomal recessive disease locus with an allele frequency of 0.0001. Parametric linkage analysis was then performed using MSIM<sup>10</sup> on the generated replicates using the same disease model that was used to generate the data, and MLOD scores were obtained for each pedigree structure.

### Linkage analysis

Genotypes for 366 microsatellite markers for family 7 and ~6000 SNP markers for families 8–10 were generated. Homozygosity mapping was performed using HomozygosityMapper.<sup>11</sup> Genome-wide parametric linkage analysis<sup>12</sup> was performed using a disease frequency of 0.0001 and assuming a fully penetrant autosomal recessive model with no phenocopies, that is, penetrance is 1 for the homozygous minor allele genotypes and 0 for heterozygous and homozygous wild-type genotypes. Marker allele frequencies were estimated using the founders and reconstructed founders from these pedigrees and additional Pakistani pedigrees, which were genotyped using the same marker panel. Two-point linkage analysis was performed using Superlink.<sup>13</sup> Multipoint linkage analysis and haplotype reconstruction were performed using MERLIN.<sup>14</sup>

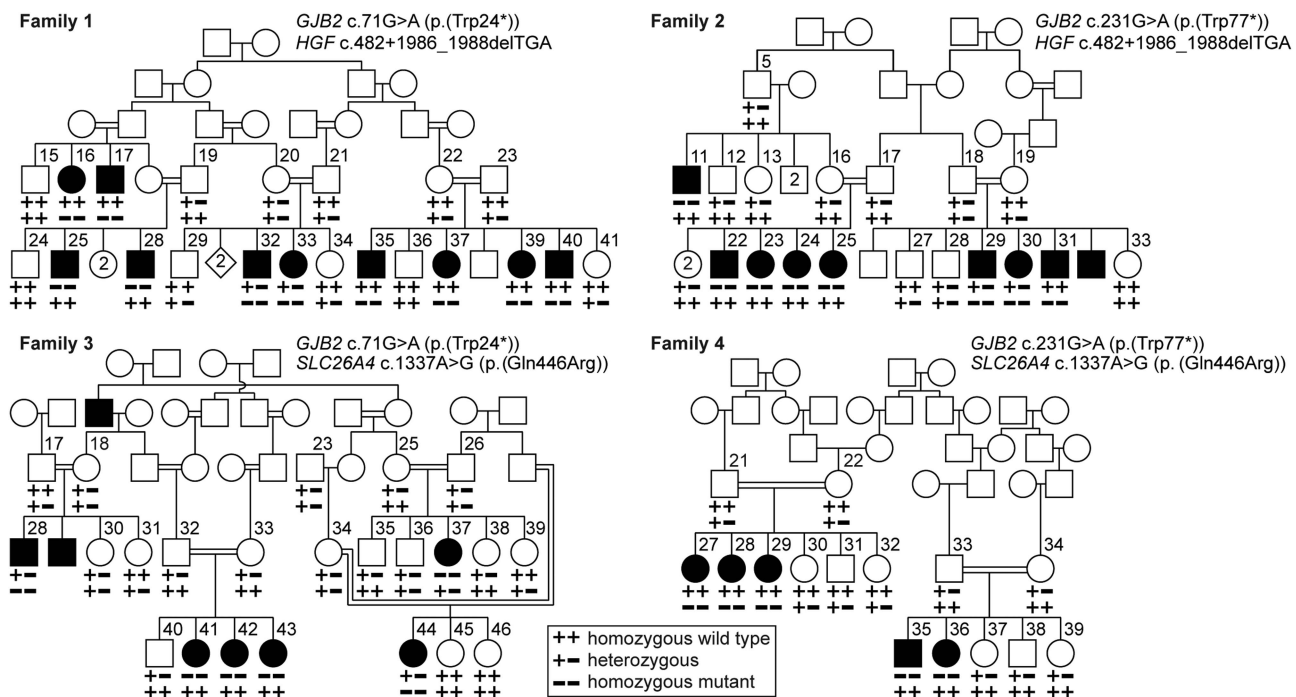
## RESULTS

Depending on the phenotype and how many different variants and genes are involved in disease etiology, screening known genes or

variants can be a useful tool prior to NGS. For nonsyndromic (NS) HI, causal variants in ~80 genes have been identified and some of these variants are frequent in individuals with NSHI in particular populations, for example, variants in *GJB2*.<sup>2–3,15–17</sup> DNA samples from hearing-impaired family members were initially screened for putatively causal *GJB2* variants and, for Pakistani families, common NSHI variants in *CIB2*, *HGF* and *SLC26A4*. Using this approach, we were able to resolve the locus heterogeneity in four families: variants in *GJB2* and *HGF* are responsible for HI in families 1 and 2, while variants in *GJB2* and *SLC26A4* are responsible for HI in families 3 and 4 (Figure 1). For six additional pedigrees, homozygous or compound heterozygous variants in *GJB2* (families 5, 6, 8 and 11), *HGF* (family 7) or *CIB2* (family 9) were initially identified for some but not all hearing-impaired individuals (Figures 2 and 3, Table 2).

The phenotype was similar across hearing-impaired family members, except for families 5 and 6 which demonstrate both locus and phenotypic heterogeneity. Family 5 segregates the *GJB2* c.71G>A (p.(Trp24\*)) variant in branch 2 (Figure 2). In branch 1, progressive vision loss due to retinitis pigmentosa in three hearing-impaired siblings provided evidence for Usher syndrome (MIM 276900) and led to the identification of the *MYO7A* c.397dupC variant within this sibship (Figure 2).<sup>18</sup>

The majority of hearing-impaired members of European-American family 6 (Figure 2) had no other clinical features, except for two individuals. Individual 99 presented with enlargement of the vestibular aqueducts (EVA), which gave a clue that HI may be due to *SLC26A4* variants (Table 2 and Figure 2). Recessively inherited *SLC26A4* variants can cause NSHI with or without EVA (MIM 600791) as well as Pendred syndrome (MIM 274600).<sup>15,19</sup> Additionally, individual 93 had temporal bone abnormalities, which is consistent with the



**Figure 1** Four pedigrees in which intra-familial locus heterogeneity was discovered by Sanger-sequencing genes that are commonly involved in the etiology of hearing impairment. In each family, two homozygous variants in different genes segregate with hearing impairment. The box on the lower side shows the legend for genotypes. Genotypes beneath each symbol follow the order of variants listed above each pedigree. The *HGF* c.482 +1986\_8delTGA variant is denoted as *del3*. Families 1–3 include individuals without congenital hearing loss who are double heterozygotes for variants in *GJB2* and *SLC26A4* and for variants in *GJB2* and *HGF*, providing evidence against digenic inheritance for these specific combinations of genes and variants.

distinctive phenotype due to variants in *POU3F4* (Figure 2).<sup>20</sup> Screening of *SLC26A4* identified the c.1334T>G (p.(Leu445Trp)) variant in individual 99,<sup>19</sup> while a 3.1-Mb deletion that includes *POU3F4* was detected in individual 93. Thus, based on clinical information and screening of *GJB2*, *SLC26A4* and *POU3F4*, the genetic etiology of HI for the majority of hearing-impaired members of European-American family 6 has been resolved, with the exception of individuals 82 and 96 (Figure 2).

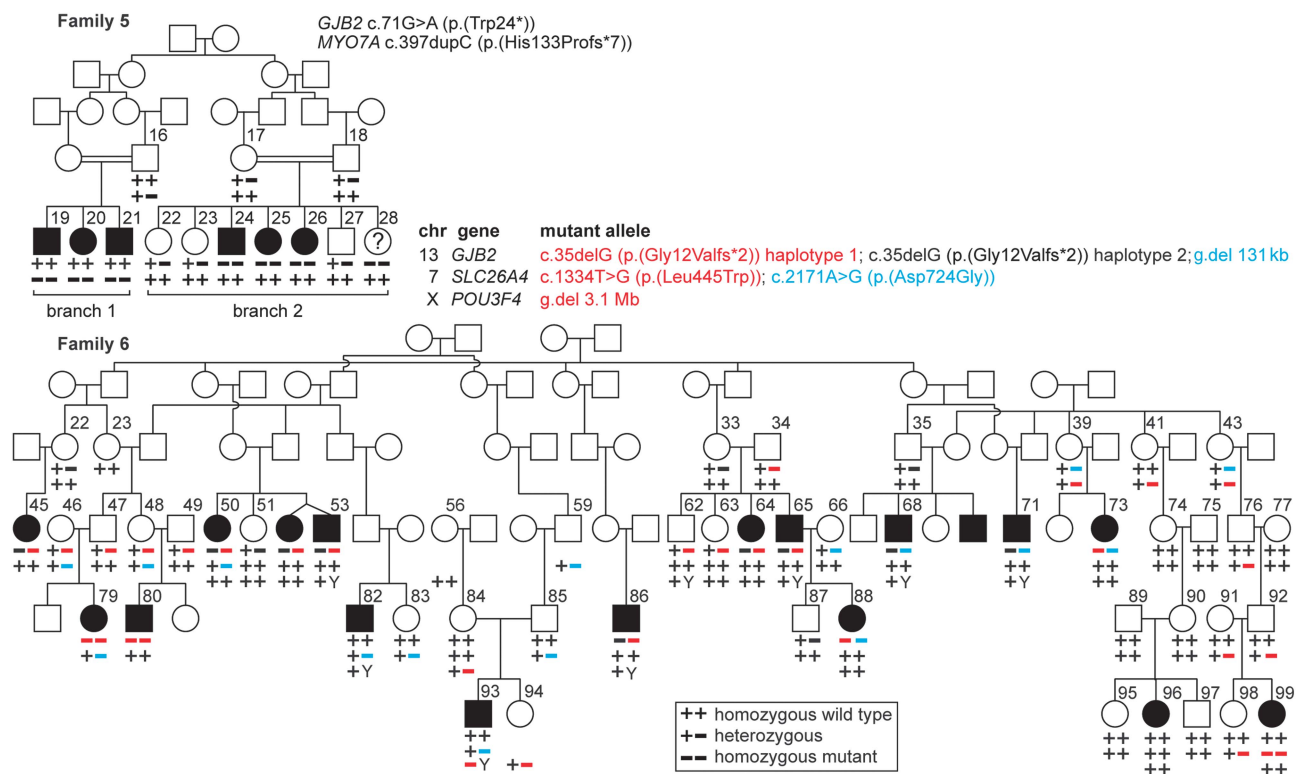
Although for families 1–6 phenotypic information and/or genetic screening aided identification of HI-causal variants even in the presence of locus heterogeneity, for many pedigrees and traits, screening of selected genes or variants is not possible due to: (a) the large number of genes known to be involved in disease; (b) gene(s) responsible for the greatest disease burden may be difficult to screen (eg, has a large number of exons and/or harbors many causal variants); and (c) many genes that harbor causal variants have not been identified. For families 7–11, we elucidated the presence of locus heterogeneity by the identification of NSHI variants using linkage analysis, NGS or both.

Linkage analysis and homozygosity mapping were used to identify individuals segregating the same NSHI variant in families 7–10 (Figure 3), which were genotyped with either microsatellite (family 7) or SNP (families 8–10) markers (Table 3). For families 7–9, the observed maximum LOD scores were much lower than the MLOD scores obtained from simulation (Table 3). In family 7, initial screening revealed the *HGF* c.482 +1986\_1988delTGA variant,

however this variant only segregated in branch 1 but not in branch 2 (Figure 3). Additionally, the MLOD score for this family was 5.73, but the observed maximum LOD score was 1.48. When linkage analysis was performed separately for each branch, for branch 1 a LOD score of 3.59 was obtained at the *HGF* region, while for branch 2 a LOD score of 2.53 was obtained at the 9q21.12–q21.13 region, which contains *TMCI*. Sanger sequencing of all coding exons of *TMCI* identified the c.1114G>A (p.(Val372Met)) variant within *TMCI*<sup>21</sup> and confirmed co-segregation of this variant with HI in branch 2.

Using the entire pedigree structure of family 9, a LOD score of 1.65 was observed although the MLOD score was 6.27. Additionally, the mapped region 4q21.21 did not harbor any known NSHI genes. Sanger sequencing showed co-segregation of the *CIB2* c.272T>C (p.(Phe91Ser)) variant with NSHI within branch 2 only. When each branch of pedigree 9 was analyzed separately, there were several linked regions per branch (Table 3). However, the 15q24.1–q26.1 region that contains *CIB2* provided the highest LOD score of 2.41 for branch 2, while for branch 1, one of the linked regions was at chromosome 7q, which encompasses *SLC26A4*. A DNA sample from hearing-impaired individual 19 who is homozygous for the 7q haplotype underwent exome NGS, which revealed a known variant *SLC26A4* c.716T>A (p.(Val239Asp))<sup>15</sup> that co-segregates with HI within branch 1 (Figure 3).

For families 8, 10 and 11 (Figure 3), subdividing the pedigrees into branches consisting of sibships could not completely resolve the problem of locus heterogeneity due to intra-sibship heterogeneity. Two of these families (10 and 11) consist of only one branch with



**Figure 2** Two families segregating autosomal recessive hearing impairment where variable phenotype of affected individuals helped to identify intra-familial locus heterogeneity. The box on the lower side shows the legend for genotypes. Genotypes beneath each symbol follow the order and colors of variant names listed above each pedigree. In family 5, a *MYO7A* frameshift variant segregates with Usher syndrome in branch 1, while the *GJB2* c.71G>A (p.(Trp24\*)) variant segregates with nonsyndromic hearing impairment in branch 2. Family 6 is from a community isolate, which descended from four ancestors that immigrated to North America from Germany in the seventeenth century. Four NSHI-causal variants segregate in this family. The *GJB2* c.35delG haplotype 1 (red) was inherited by 27 of >200 genotyped community members, while *GJB2* c.35delG haplotype 2 (black) was identified in 21 genotyped family members who descended from two immigrants originating from nineteenth century Austro-Hungary. Individuals 82 and 96 of family 6 were screened using OtoSCOPE,<sup>30</sup> a NGS capture array targeting NSHI and selected hearing impairment syndrome genes, but no causal variants were identified.

intra-sibship locus heterogeneity, while family 8 displays intra-sibship heterogeneity in one of its branches (branch 3). Therefore, for the analysis of these branches, the affection status of the siblings needs to be varied.

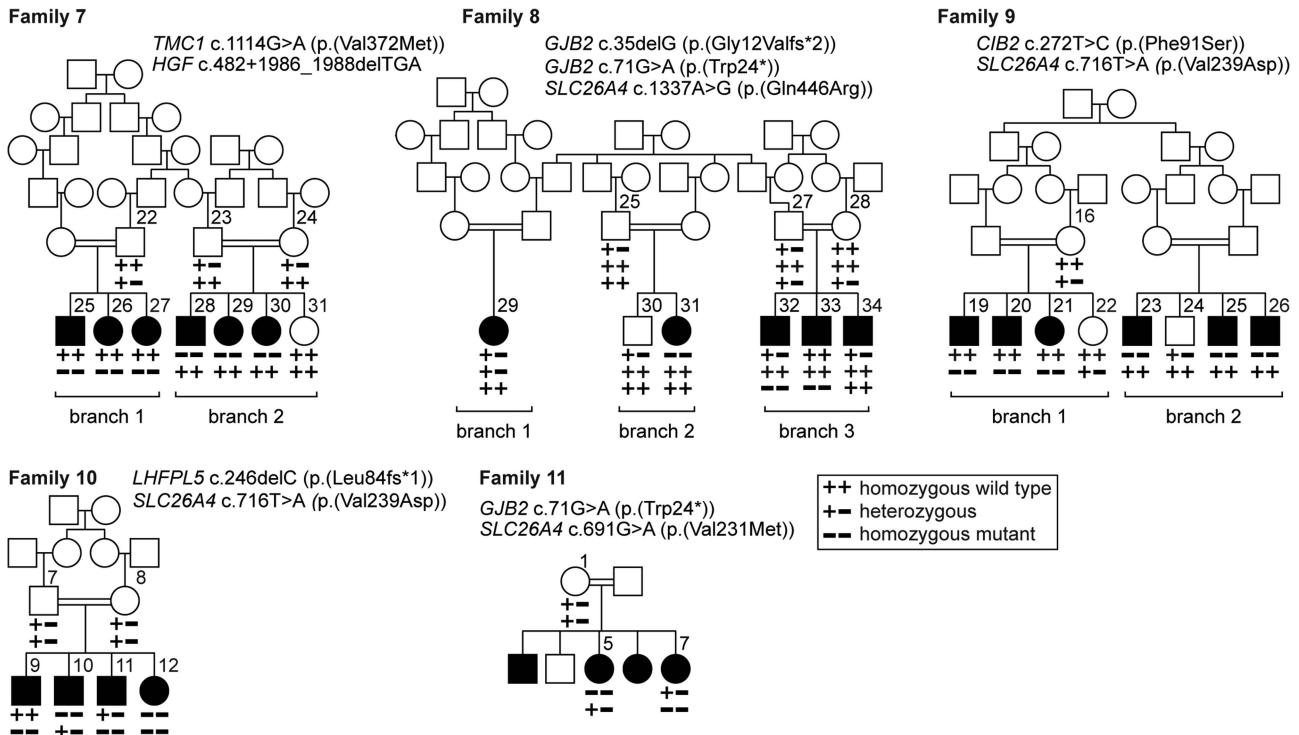
For family 8 (Figure 3), when all three branches are analyzed together the region of homozygosity at 10q26.3-qter did not encompass *GJB2* and *SLC26A4*, which are involved in NSHI (Table 3). Sanger sequencing of *GJB2* revealed that individual 31 of branch 2 is homozygous for the c.35delG variant while individual 29 of branch 1 is compound heterozygous for c.35delG and c.71G>A (p.(Trp24\*)). For branch 3 (Table 3), the chromosome 4p region was homozygous in all three affected sibs and produced a LOD score of 2.4, but this region does not contain any known NSHI gene. When individual 34 was considered to have unknown affection status, a number of regions were homozygous and produced the same LOD score of 1.8; two of these regions contained NSHI genes *RDX* and *SLC26A4* (Table 3). DNA from individual 33 of branch 3 underwent exome NGS and was found to be homozygous for the known variant *SLC26A4* c.1337A>G (p.(Gln446Arg)). Sanger sequencing confirmed that his hearing-impaired brother (individual 32) is also homozygous for this variant while his other hearing-impaired brother (individual 34) is wild type. The cause of HI for individual 34 remains unknown.

Family 10 has four affected children whose parents are first cousins (Figure 3). For family 10, when all sibs are considered affected the observed maximum LOD score and the MLOD score were equal (3.01), but the mapped region on chromosome 7 does not include the causal variants (Table 3). For these four hearing-impaired siblings, we varied the affection status by alternately assigning unknown affection status for each sib and also different pairs of sibs. Although this approach led to the identification of many regions (data not shown), it revealed which among these affected siblings were likely to segregate

the same causal variants. When affection status of individual 10 was made unknown, a large region on chromosome 7, which includes *SLC26A4*, was homozygous in individuals 9, 11 and 12, and linkage analysis produced a LOD score of 2.4, which is the same as the MLOD score for the selected siblings. Linkage analysis with individuals 9 and 11 having unknown affection status led to mapping of several regions, including chromosome 6p24.3-p12.3, which contains *LHFPL5*. DNA from individual 12 who is homozygous for chromosomal regions 6p and 7q underwent exome NGS. This individual is homozygous for two variants *SLC26A4* c.716T>A (p.(Val239Asp)) and *LHFPL5* c.246delC. Sanger sequencing revealed that her brothers (9 and 11) were homozygous for the *SLC26A4* variant but not the *LHFPL5* variant, whereas individual 10 was homozygous for the *LHFPL5* variant but not the *SLC26A4* variant (Figure 3).

Family 11 also presents intra-sibship heterogeneity (Figure 3). For this pedigree, DNA is available for two hearing-impaired family members. Hearing-impaired individual 5 is homozygous for *GJB2* c.71G>A (p.(Trp24\*)), whereas her affected sister is heterozygous for this variant. DNA from individual 7 underwent NGS after targeted capture for coding exons of all previously reported HI genes. NGS and confirmatory Sanger sequencing revealed that individual 7 is homozygous for *SLC26A4* c.691G>A (p.(Val231Met)).<sup>15</sup>

In Pakistani families with HI for which we have performed linkage analysis and sequencing, we detected intra-familial locus heterogeneity in 15.3% (95% CI: 11.4%, 19.9%) (Table 4). The families with locus heterogeneity (*N*=45) have at least one variant in a previously published HI gene that segregates with HI in some but not all hearing-impaired family members. Conversely, in those families with locus homogeneity (*N*=249), a variant within a known HI gene completely segregates with HI.



**Figure 3** Pedigrees in which linkage analysis and/or NGS were used to identify intra-familial or intra-sibship locus heterogeneity. In families 7–10, linkage analysis and homozygosity mapping prior to NGS helped to identify the causal variant. Families 8, 10 and 11 demonstrate intra-sibship locus heterogeneity. For family 11, the parents are known to be related (double bars), but the exact relationship is unknown.

**Table 3** LOD scores and mapped loci for four NSHI families with linkage data

Pedigrees and subpedigrees	Genotyping array	MLOD <sup>a</sup>	Observed LOD <sup>b</sup>	Mapped region <sup>c</sup>	ARHI gene in region	Gene region
<i>Family 7</i>	388 STR	5.73	1.48	7q21.11-q21.3.	<i>HGF</i>	7q21.11
Branch 1:		3.59	3.59	7q21.11-q22.2	<i>HGF</i>	7q21.11
Branch 2:		2.53	2.53	9q21.12-q21.13	<i>TMC1</i>	9q21.13
<i>Family 8</i>	5,913 SNP	4.19	2.36	10q26.3-qter	NA	NA
Branches 1 and 2:		2.18	2.08	7p14.1	NA	NA
			1.75	8p21.3-p21.2	NA	NA
			2.17	10p14-p12.31	NA	NA
			2.16	10p12.1-q11.21	NA	NA
			2.13	10q26.3-qter	NA	NA
			2.16	12q24.32-q24.33	NA	NA
			1.32 <sup>d</sup>	13pter-q12.11	<i>GJB2/GJB6</i>	13q12.11
			1.36	14q31.3-q32.12	NA	NA
			1.78	16p13.3	<i>TBC1D24<sup>e</sup></i>	16p13.3
Branch 3: 3 sibs affected		2.41	2.41	4p16.1-p15.2	NA	NA
Branch 3: 34 unknown <sup>f</sup>		1.81	1.81	4p16.1-p15.1	NA	NA
			1.80	5p15.33-p15.31	NA	NA
			1.81	7q21.3-q31.2	<i>SLC26A4</i>	7q22.3
			1.81	11q22.1-q22.3	<i>RD<sup>g</sup></i>	11q22.3
			1.80	12q23.1-q24.11	NA	NA
<i>Family 9</i>	5,913 SNP	6.27	1.65	4q21.21	NA	NA
Branch 1:		2.53	2.16	2p16.1-p14	NA	NA
			1.60	7q22.3-q31.1	<i>SLC26A4</i>	7q22.3
			1.79	12q24.21-q24.23	NA	NA
Branch 2:		2.53	2.21	2q37.3	NA	NA
			2.39	4q13.3-q21.23	NA	NA
			1.89	5q31.3-q32	NA	NA
			2.41	15q24.1-q26.1	<i>CIB2</i>	15q25.1
<i>Family 10</i>	6,090 SNP	3.01	3.01	7q31.1-q32.2	NA	NA
9, 11 unknown:		1.81	1.80	4q13.3-q21.1	NA	NA
			1.80	5q31.1-q31.3	<i>HARS<sup>h</sup></i>	5q31.3
			1.81 <sup>g</sup>	6p24.3-p12.3	<i>LHFPL5</i>	6p21.31
			1.80	7q31.1-q32.2	NA	NA
			1.80	8p12	NA	NA
			1.80	9pter-p22.3	NA	NA
			1.80	17q25.3	NA	NA
10 unknown:		2.41	2.40 <sup>g</sup>	7p12.3-q32.2	<i>SLC26A4</i>	7q22.3

<sup>a</sup>Maximum LOD scores ( $\theta=0$ ) from simulation using MSIM from the SLINK package.<sup>10</sup> <sup>b</sup>LOD scores are from linkage analysis using MERLIN<sup>14</sup> and Superlink.<sup>13</sup> <sup>c</sup>Mapped regions are based upon the regions of homozygosity. <sup>d</sup>For the SNP array, the markers adjacent to the *GJB2* region are uninformative, hence the low LOD score. <sup>e</sup>For genes that lie within the mapped regions and that are involved in autosomal recessive hearing impairment, namely *TBC1D24<sup>f</sup>* (MIM 613577), *RD<sup>g</sup>* (MIM 179410) or *HARS<sup>h</sup>* (MIM 142810), no potentially causal variants were identified by exome NGS.<sup>1</sup>For branch 3 of family 8, only two of the three siblings are homozygous for *SLC26A4* c.1337A>G (p.(Gln446Arg)). LOD scores for branch 3 were obtained with individual 34 having unknown affection status. <sup>f</sup>Family 10 has only one branch. The LOD score at chromosome 6 was obtained with individuals 9 and 11 having unknown affection status and at chromosome 7 with individual 10 having unknown affection status.

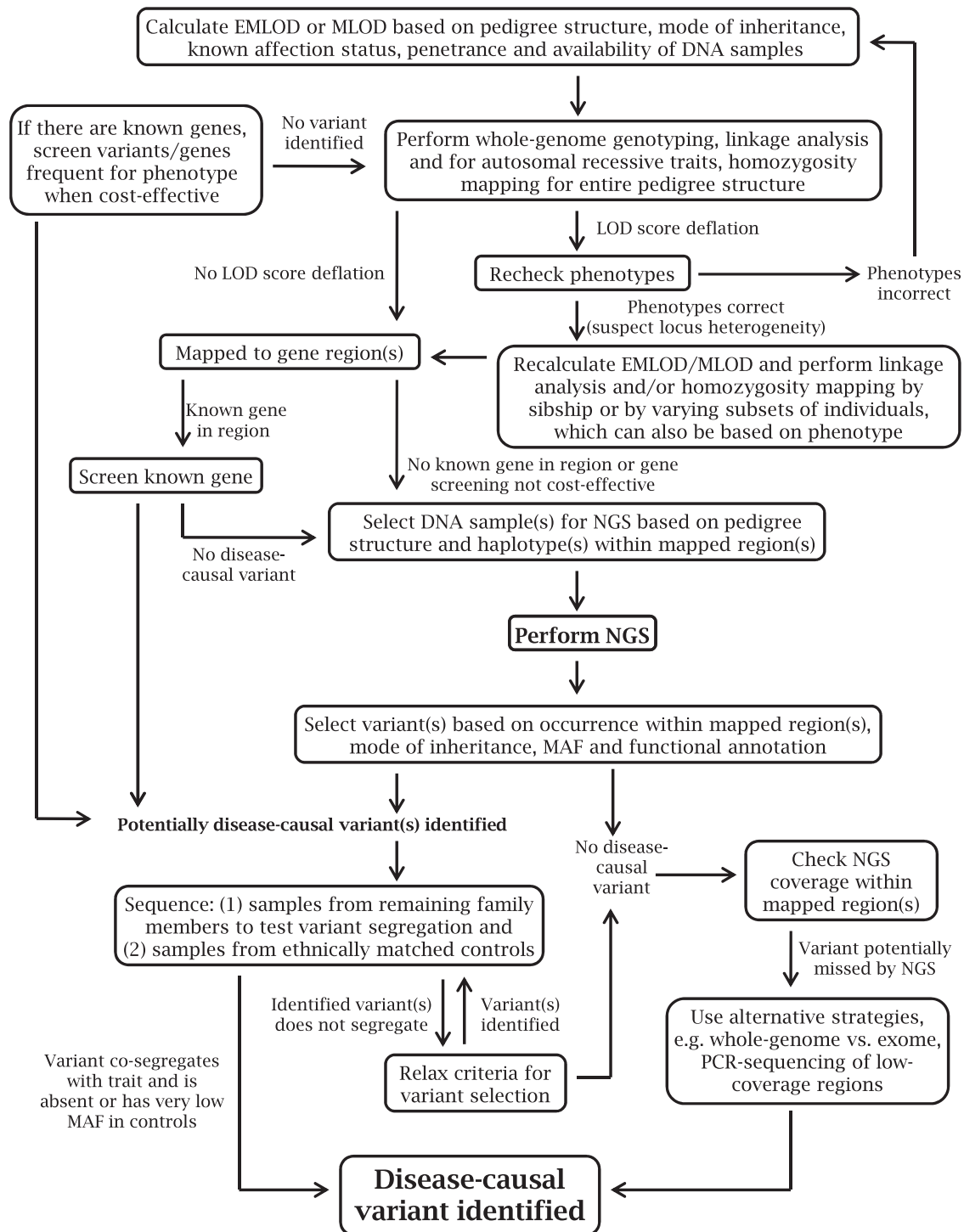
## DISCUSSION

Linkage analysis alone<sup>12,14,22,23</sup> or combined with homozygosity mapping<sup>11,24</sup> for consanguineous pedigrees segregating autosomal recessive traits can help elucidate if there is locus heterogeneity within a pedigree. Using information on the pedigree structure, mode of inheritance, affection status, penetrance and availability of genotype data, simulation studies can be used to obtain the expected MLOD (EMLOD) and MLOD scores for each pedigree. The MLOD score is the highest LOD score obtained for all replicates while the EMLOD score is the average for all replicates of the LOD score, which has been maximized over the recombination fraction ( $\theta$ ) for each replicate. For some replicates, the highest possible LOD score for the pedigree may not be obtained because: (1) marker loci are not fully informative; (2) there is reduced penetrance; or (3) there are phenocopies. If the

**Table 4** Number of families with or without locus heterogeneity for previously reported hearing impairment genes<sup>a</sup>

Classification based on variants identified and tests performed	Families with heterogeneity	Families without heterogeneity	Total
Variant identified in <i>GJB2</i> , <i>CIB2</i> or <i>HGF</i> <sup>b</sup>	19	98	117
<i>Variants identified in other known genes</i>			
Linkage analysis+Sanger sequencing	8	87	95
Linkage analysis+NGS	18	64	82
Total <sup>c</sup>	45	249	294

<sup>a</sup>Only families 7–10 from this report belong to the same cohort of 294 families with HI and are therefore included in these counts. <sup>b</sup>Of the 19 families with locus heterogeneity for *GJB2*, *CIB2* or *HGF*, all families were submitted for linkage analysis but only 4 families were submitted for NGS. Of the 98 families with locus homogeneity for *GJB2*, *CIB2* or *HGF*, 11 families were Sanger sequenced only, 83 families were submitted for linkage analysis and 4 families were submitted for both linkage analysis and NGS. <sup>c</sup>The proportion of families with locus heterogeneity is 15.3% (95%CI: 11.9%, 19.9%).



**Figure 4** Workflow for identification of causal variants in families with Mendelian traits.

phenotype under study is fully penetrant without phenocopies, it is reasonable to assume that fully informative marker loci, which are completely linked to the disease locus, are available for analysis and therefore the MLOD score should be obtained, otherwise the EMLOD score should be evaluated. For autosomal recessive HI, the disease is fully penetrant, and it is unlikely that there are phenocopies so it is permissible to evaluate the MLOD. If there is allelic heterogeneity or compound heterozygosity within a pedigree, linkage analysis will

facilitate the mapping without a deflation in the LOD score compared with the MLOD/EMLOD score. However, when the underlying disease etiology is due to variants in multiple genes segregating in a single family, the MLOD/EMLOD score for the pedigree will usually not be achieved and low LOD scores are often observed. For autosomal recessive traits, if homozygosity mapping is performed in the presence of locus or allelic heterogeneity, no common region of homozygosity will be observed between affected individuals.

If a pedigree fails to provide a LOD score near the expected maximum values, analyses can be performed by breaking the pedigrees into branches to see whether or not each branch produces the MLOD/EMLOD score for that branch.<sup>25</sup> Homozygosity mapping can also be performed separately for each branch. However, if there is intra-sibship locus heterogeneity, this approach can likewise fail. In this case, linkage analysis and homozygosity mapping can be performed by analyzing subgroups of the affected individuals, by comparison of the observed maximum LOD score with the MLOD/EMLOD score for the pedigree subgroup and, when appropriate, identification of those affected individuals who share the same homozygous region. Because of multiple testing, significance levels that are usually used for LOD scores are no longer valid, for example, LOD of 3.3 is no longer equivalent to a genome-wide significance level of  $P=0.05$ .<sup>26</sup> Even if the observed maximum LOD scores that are obtained are equal or close to the MLOD/EMLOD score, due to the small pedigree size the disease locus may not map to a unique region. Additionally, disease loci may not necessarily lie within the regions with the highest LOD score (Table 2). For example, a region with a LOD score of 3.0 is only 1.6 times more likely to contain the gene of interest compared with a region with an observed LOD score of 2.8; the difference in the likelihood is so small the causal gene may lie in either region. Once NGS is completed, multiple linked regions can readily be followed up to identify putatively causal variants. Subgrouping affected individuals while performing linkage analysis can help to resolve which pedigree members are most likely to have the disorder or trait due to the same variant. However, if the family unit is very small, for example, a single affected individual with parents, it will not be informative for linkage, or the difference in likelihoods are so small that it is difficult to evaluate which subset of pedigree members are affected due to the same gene.

Linkage analysis and haplotype information can be used to guide which samples to select for NGS. For pedigrees with locus homogeneity selection of sample(s) for NGS is based upon the entire pedigree, whereas for families with locus heterogeneity samples are selected within each pedigree subgroup predicted to segregate the same causal variant. Selecting a sample for NGS from one affected individual is usually sufficient for autosomal recessive traits when the pedigree is consanguineous; however, for autosomal dominant traits it is advisable to select samples from two or more affected individuals. The selection of DNA samples for NGS can be based upon haplotypes within a pedigree or pedigree subgroup, where it is desirable to select individual(s) with the smallest shared haplotype region. Linkage information combined with NGS and filtering can reduce the number of variants for follow-up.<sup>5–6,27</sup> When there is strong evidence that a causal variant lies within a mapped region, variant selection can be easily performed even with relaxed filters (eg, occurrence in variant databases, population frequency  $>0.1\%$ ). On the other hand, when no causal variants are identified despite strong linkage evidence, this can give a hint that the variant may have been missed due to poor coverage by NGS or alternative strategies (eg, CNV calling) might be needed.

As the cost of NGS decreases, instead of sequencing DNA samples from a subset of family members, an affordable strategy will be to sequence all available pedigree members. Performing linkage analysis using NGS genotypes from the entire family will aid in identifying potentially causal genes/variants that segregate with disease.<sup>28,29</sup> The linkage analysis and homozygosity mapping approach described in this article can also be used with NGS data.

In summary, we demonstrate that linkage analysis and homozygosity mapping are still useful in the age of NGS. Although families

presented here have HI, the proposed strategies discussed here and outlined in Figure 4 should be applicable to any Mendelian trait. In the presence of intra-familial locus heterogeneity, the causal variants may be identified by: (1) initial screening of genes that are known to harbor disease-causal variants; (2) use of additional clinical information that may give phenotypic clues to genetic etiology; (3) when locus heterogeneity is suspected, performing simulation and linkage analysis or homozygosity mapping for branches or subsets of individuals in order to identify individuals most likely to share the same disease-causal variant; (4) using linkage and haplotype information to select DNA samples from affected individuals for NGS; and (5) considering locus heterogeneity when evaluating co-segregation of variant with phenotype within the family. We have shown that occurrence of intra-familial locus heterogeneity, including intra-sibship heterogeneity, is not rare and that taking intra-familial locus heterogeneity into account while analyzing pedigree data can increase the success rate in the identification of causal variants for Mendelian traits. Although we have concentrated on demonstrating the presence of locus heterogeneity in pedigrees segregating Mendelian disease, this problem is not unique to Mendelian traits but also occurs in families segregating complex traits particularly in those with bilineal inheritance.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank the family members who participated in this study and Dr Dennis Drayna and Dr Andrew Griffith for critical reviews of our manuscript. This study utilized the Biowulf Linux cluster at the National Institutes of Health (NIH), Bethesda, MD (<http://biowulf.nih.gov>). This work was supported by: the National Institute on Deafness and Other Communication Disorders (NIDCD)/NIH intramural program DC000039-18 (to TBF); the Higher Education Commission of Pakistan (to Sheikh Riazuddin and WA); NIH grants R01 DC003544 (to R.J.H.S.), R01 DC011803 and R01 DC012564 (to Saima Riazuddin and ZMA), R03 DC012434 (to ESW), R01 DC011651 and R01 DC003594 (to SML), and U54 HG006493 (to DAN); the International Center for Genetic Engineering and Biotechnology, Trieste, Italy CRP/PAK08-01 contract 08/009 (to Sheikh Riazuddin); and the Hearing Health Foundation (to ESW). Genome-wide genotyping of families (from WA and SML) was performed at the Center for Inherited Disease Research, which is funded through the NIH to The Johns Hopkins University, Contract Number N01-HG-65403.

#### WEB RESOURCES

The URLs for data presented herein are as follows: 1000 Genomes Project, <http://www.1000genomes.org>; HomozygosityMapper, <http://www.homozygositymapper.org>; MutationTaster, <http://www.mutationtaster.org/>; NHLBI Exome Sequencing Project (ESP) Exome Variant Server (EVS), <http://evs.gs.washington.edu/EVS/>; and Primer3, <http://frodo.wi.mit.edu/>.

- 1 Yasunaga S, Grati M, Cohen-Salmon M *et al*: A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. *Nat Genet* 1999; **21**: 363–369.
- 2 Schultz JM, Khan SN, Ahmed ZM *et al*: Noncoding mutations of HGF are associated with nonsyndromic hearing loss, DFNB39. *Am J Hum Genet* 2009; **85**: 25–39.
- 3 Riazuddin S, Belyantseva IA, Giese AP *et al*: Alterations of the CIB2 calcium- and integrin-binding protein cause Usher syndrome type 1J and nonsyndromic deafness DFNB48. *Nat Genet* 2012; **44**: 1265–1271.
- 4 Ng SB, Buckingham KJ, Lee C *et al*: Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30–U41.
- 5 Rehman AU, Santos-Cortez RLP, Morell RJ *et al*: Mutations in TBC1D24, a gene associated with epilepsy, also cause nonsyndromic deafness DFNB86. *Am J Hum Genet* 2014; **94**: 144–152.



- 6 Santos-Cortez RLP, Lee K, Azeem Z *et al*: Mutations in KARS, encoding lysyl-tRNA synthetase, cause autosomal-recessive nonsyndromic hearing impairment DFNB89. *Am J Hum Genet* 2013; **93**: 132–140.
- 7 Rehman AU, Morell RJ, Belyantseva IA *et al*: Targeted capture and next-generation sequencing identifies C9orf75, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79. *Am J Hum Genet* 2010; **86**: 378–388.
- 8 Yang YP, Muzny DM, Reid JG *et al*: Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N Engl J Med* 2013; **369**: 1502–1511.
- 9 Gilissen C, Hoischen A, Brunner HG, Veltman JA: Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 2012; **20**: 490–497.
- 10 Ott J: Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 1989; **86**: 4175–4178.
- 11 Seelow D, Schuelke M, Hildebrandt F, Nurnberg P: HomozygosityMapper—an interactive approach to homozygosity mapping. *Nucleic Acids Res* 2009; **37**: W593–W599.
- 12 Ott J: *Analysis of human genetic linkage*, 3rd edn. Baltimore, MD, USA, 1999.
- 13 Fishelson M, Geiger D: Exact genetic linkage computations for general pedigrees. *Bioinformatics* 2002; **18** (Suppl 1): S189–S198.
- 14 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 15 Anwar S, Riazuddin S, Ahmed ZM *et al*: SLC26A4 mutation spectrum associated with DFNB4 deafness and Pendred's syndrome in Pakistanis. *J Hum Genet* 2009; **54**: 266–270.
- 16 Morell RJ, Kim HJ, Hood LJ *et al*: Mutations in the connexin 26 gene (GJB2) among Ashkenazi Jews with nonsyndromic recessive deafness. *N Engl J Med* 1998; **339**: 1500–1505.
- 17 Snoeckx RL, Huygen PLM, Feldmann D *et al*: GJB2 mutations and degree of hearing loss: A multicenter study. *Am J Hum Genet* 2005; **77**: 945–957.
- 18 Riazuddin S, Nazli S, Ahmed ZM *et al*: Mutation spectrum of MYO7A and evaluation of a novel nonsyndromic deafness DFNB2 allele with residual function. *Hum Mutat* 2008; **29**: 502–511.
- 19 Choi BY, Stewart AK, Madeo AC *et al*: Hypo-functional SLC26A4 variants associated with nonsyndromic hearing loss and enlargement of the vestibular aqueduct: genotype-phenotype correlation or coincidental polymorphisms? *Hum Mutat* 2009; **30**: 599–608.
- 20 Dekok YJM, Vandermaarel SM, Bitnerglindzic M *et al*: Association between X-linked mixed deafness and mutations in the Pou domain gene Pou3f4. *Science* 1995; **267**: 685–688.
- 21 Santos RL, Wajid M, Khan MN *et al*: Novel sequence variants in the TMC1 gene in Pakistani families with autosomal recessive hearing impairment. *Hum Mutat* 2005; **26**: 396.
- 22 Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984; **81**: 3443–3446.
- 23 Gudbjartsson DF, Jonasson K, Frigge ML, Kong A: Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000; **25**: 12–13.
- 24 Friedman TB, Liang Y, Weber JL *et al*: A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nat Genet* 1995; **9**: 86–91.
- 25 Raza MH, Gertz EM, Mundorff J *et al*: Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance. *Hum Genet* 2013; **132**: 385–396.
- 26 Lander E, Kruglyak L: Genetic dissection of complex traits - guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- 27 Santos-Cortez RL, Lee K, Giese AP *et al*: Adenylate cyclase 1 (ADCY1) mutations cause recessive hearing impairment in humans and defects in hair cell function and hearing in zebrafish. *Hum Mol Genet* 2014; **23**: 3289–3298.
- 28 Furman I, Rieder MJ, Da Ponte S *et al*: Sequence-based linkage analysis. *Am J Hum Genet* 2004; **75**: 647–653.
- 29 Smith KR, Bromhead CJ, Hildebrand MS *et al*: Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. *Genome Biol* 2011; **12**: R85.
- 30 Shearer AE, DeLuca AP, Hildebrand MS *et al*: Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci USA* 2010; **107**: 21104–21109.
- 31 Wilch E, Azaiez H, Fisher RA *et al*: A novel DFNB1 deletion allele supports the existence of a distant cis-regulatory region that controls GJB2 and GJB6 expression. *Clin Genet* 2010; **78**: 267–274.
- 32 Kalay E, Li Y, Uzumcu A *et al*: Mutations in the lipoma HMGIC fusion partner-like 5 (LHFP5) gene cause autosomal recessive nonsyndromic hearing loss. *Hum Mutat* 2006; **27**: 633–639.
- 33 Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; **7**: 575–576.
- 34 Khan SY, Ahmed ZM, Shabbir MI *et al*: Mutations of the RDX gene cause nonsyndromic hearing loss at the DFNB24 locus. *Hum Mutat* 2007; **28**: 417–423.
- 35 Puffenberger EG, Jinks RN, Sougnez C *et al*: Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One* 2012; **7**: e28936.

## APPENDIX

### University of Washington Center for Mendelian Genomics:

University of Washington: Michael J Bamshad, Jay Shendure, Deborah A Nickerson, Peter Anderson, Marcus Annable, Mallory Beightol, Brian L Browning, Kati J Buckingham, Christina Chen, Jennifer Chin, Jessica X Chong, Colleen Davis, Lindsay Felker, Christopher Frazar, David Hanna, Gail P Jarvik, Eric Johanson, Martin Kircher, Tom Kolar, Daniel Luksic, Margaret J McMillin, Sean McGee, Brenton Munson, Brian J O'Roak, Bryan Paepfer, Karynne Patterson, Eric Phillips,

Jessica Pijoan, Christa Poel, Peggy D Robertson, Tristan Shaffer, Cindy Shephard, Deborah L Siegel, Joshua D Smith, Jeffrey C Staples, Holly K Tabor, Monica Tackett, and Qian Yi.

Seattle Children's Hospital: Michael J Bamshad, and Holly K Tabor.

Baylor College of Medicine: Zongxiao He, Suzanne M Leal, Regie Santos-Cortez, and Gao Wang.

University of Michigan: Gonçalo R Abecasis, and Goo Jun.

HudsonAlpha Institute of Technology: Gregory M Cooper, and Preti Jain.