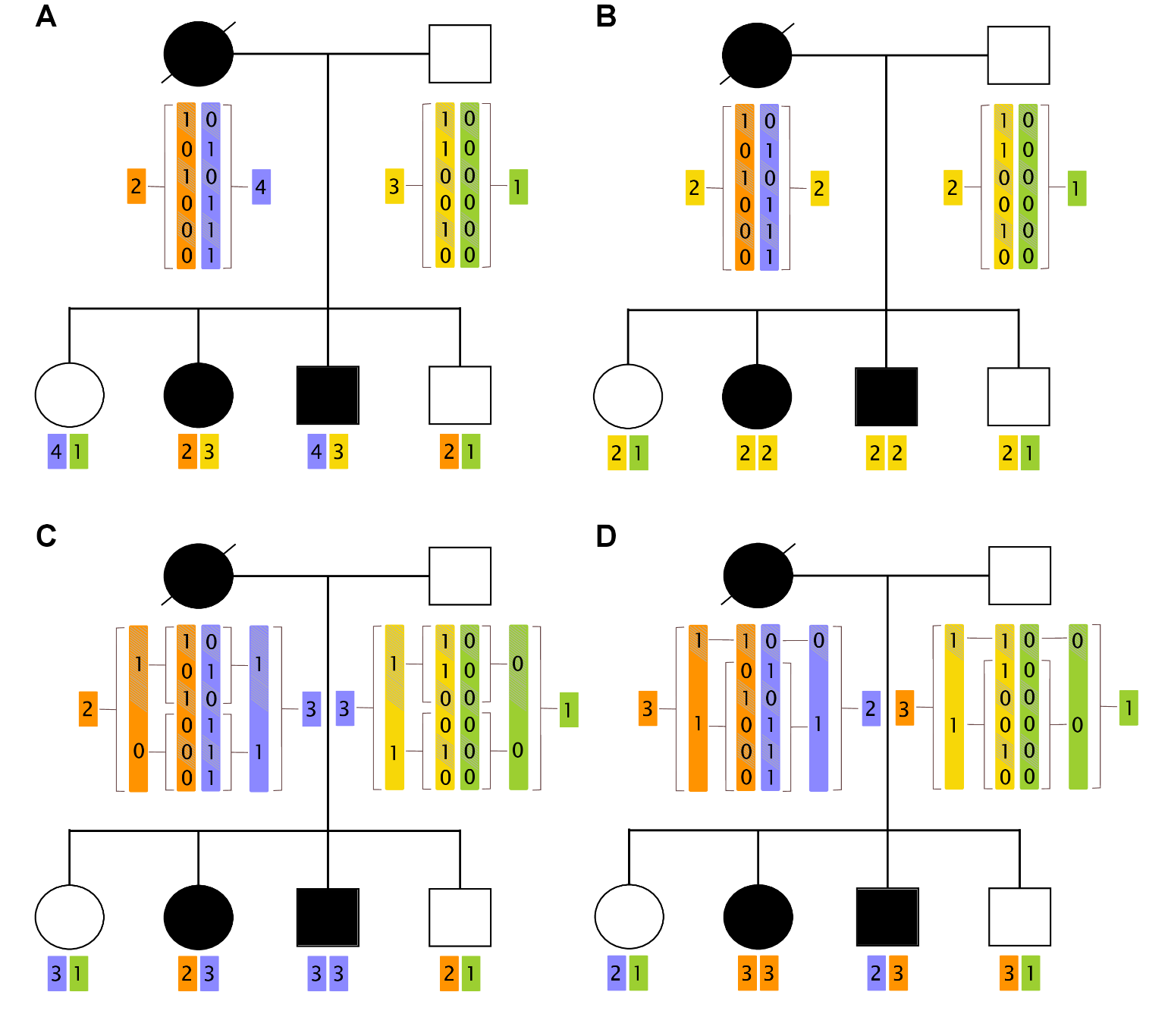
**Supplemental Methods**

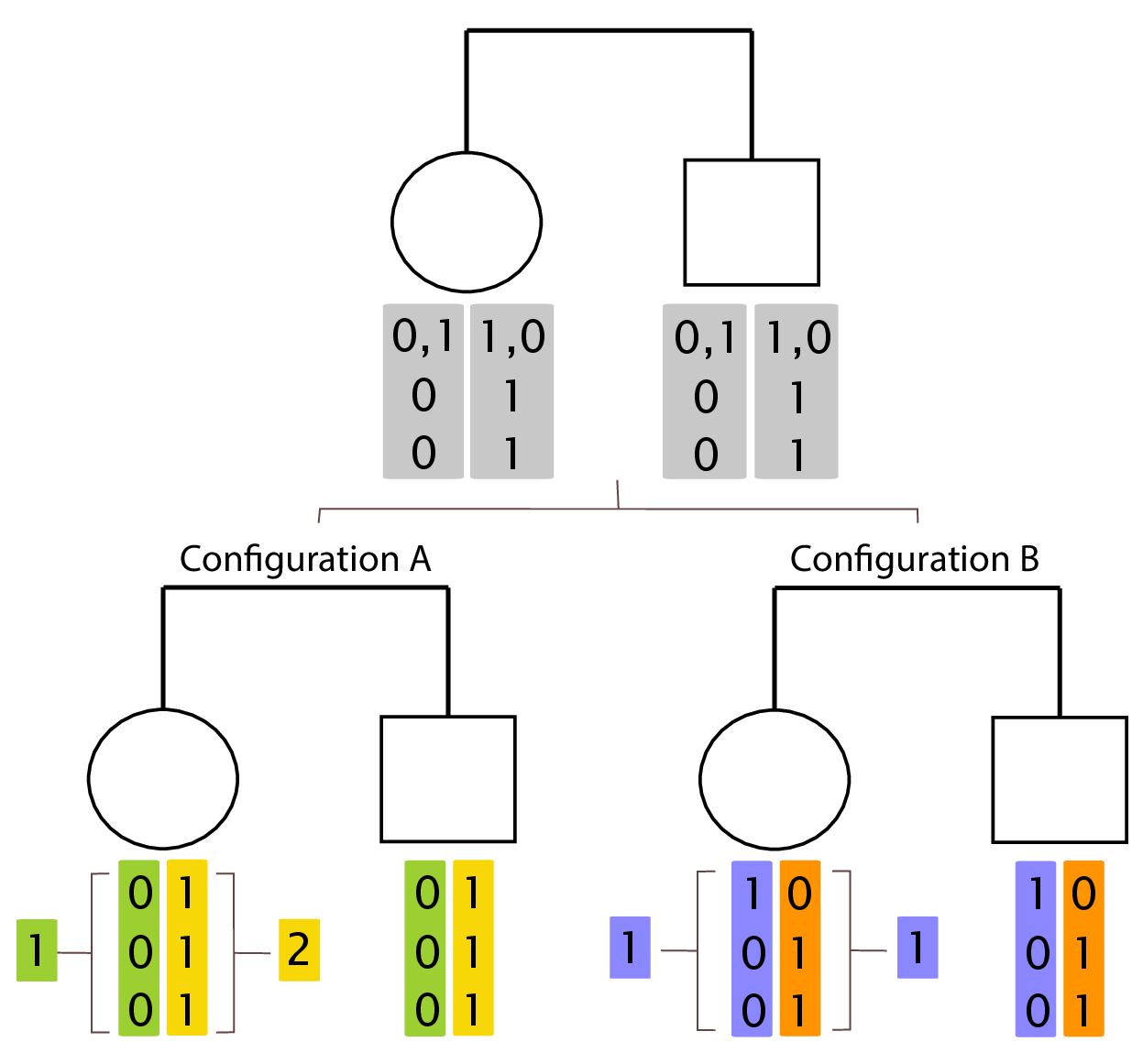
*Simulation studies*

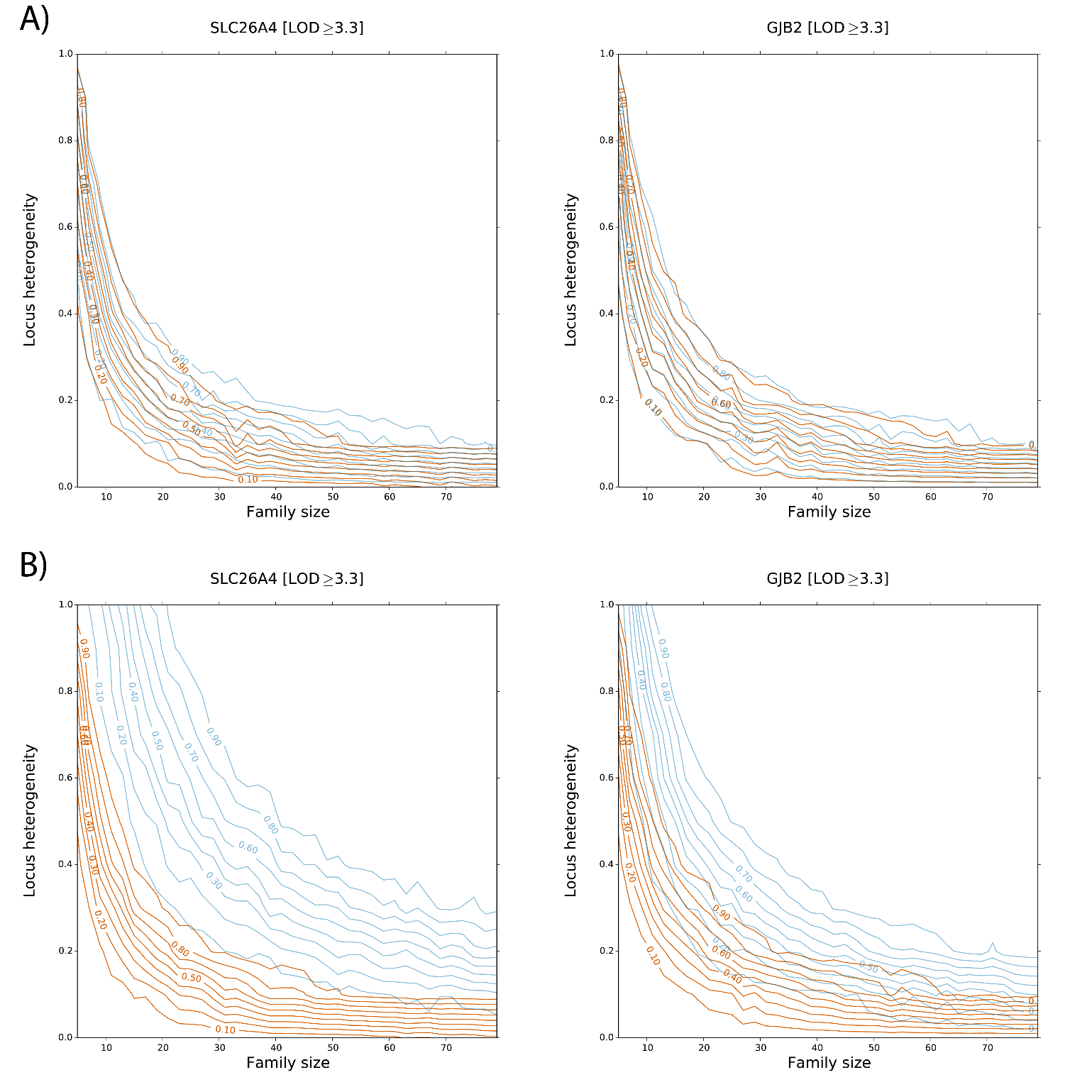
We performed empirical power calculations of two-point linkage analysis on several known non-syndromic hearing impairment (NSHI) genes, including two autosomal recessive genes *GJB2* and *SLC26A4*, and two autosomal dominant genes *MYO7A* and *MYH9*. We simulated 2-generational pedigrees, allowing for 3 up to 8 offspring in the last generation with proportions determined by the distribution of number of children per family in the United States in 2012, rescaled such that these proportions add up to 100% (3 children: 69.34%, 4 children: 20.52%, 5 children: 6.84%, 6 children: 2.28%, 7 children 0.76%, 8 children 0.26%). Genotypes are simulated for the four genes based on the variant sites and the corresponding minor allele frequencies in European Americans recorded in Exome Variant Server. We annotate these variants using Deafness Variation Database (DVD) and NCBI ClinVar, labelling variants as “pathogenic” if they present in both databases as pathogenic. Disease status for individuals are determined by genotypes on those pathogenic sites under dominant mode of inheritance for *MYO7A* and *MYH9*, and recessive & compound recessive for *GJB2* and *SLC26A4*, assuming complete penetrance. Additionally for each mode of inheritance we allow for allelic heterogeneity among families, i.e., the causal variant site in a gene may not be the same for different families. We ascertain simulated families having two or more affected offspring for linkage analysis. To introduce locus heterogeneity we mix families having pathogenic mutations in one gene but not others, so that each simulated gene contributes to etiology of only a proportion of families in the entire dataset. We simulate 500 replicates under each different settings of sample sizes, modes of inheritance, presence of allelic heterogeneity and proportions of locus heterogeneity. For each replicate we compute LOD and HLOD scores using regional markers generated by collapsed haplotype pattern markers (CHP) as well as the maximum LOD and HLOD scores using single variant markers (SVM). Power is estimated by  and where “#” stands for “number of times” and *N* is the total number of replicates. Power comparisons between CHP and SVM are summarized by contour plots as displayed in Figures S3 – S8.

**Supplemental Figures**

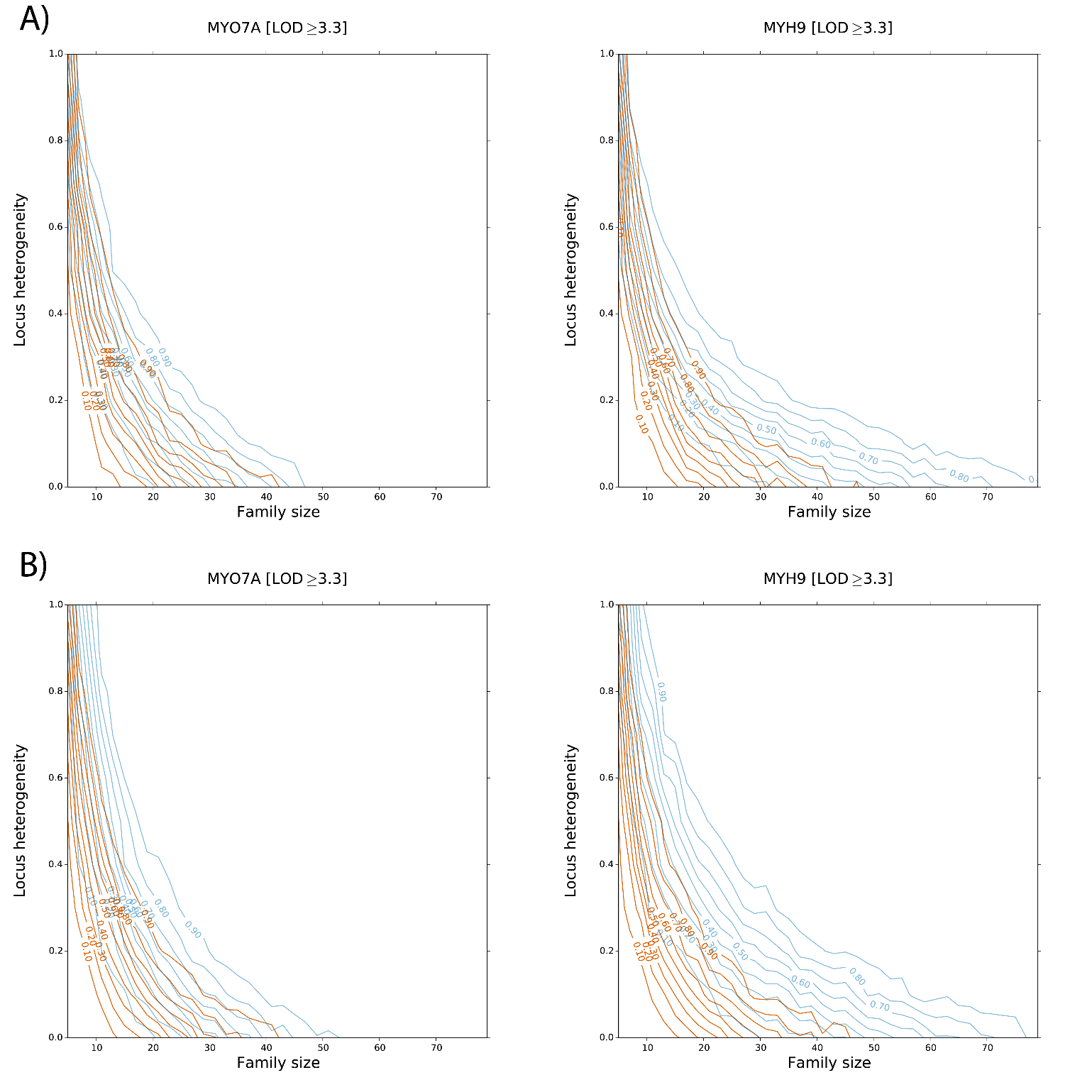
****

**Figure S1: The Collapsed Haplotype Pattern (CHP) method**. This figure illustrates the creation of regional markers for a group of 6 variants in pedigree with autosomal compound recessive disorder. Panel A displays the use of original haplotype patterns as regional markers, a special case of the CHP method using bins of size 1; panel B displays the *complete collapsing* theme with bin size equaling the length of the region; panel C displays a collapsing theme with bins of size 3 and panel D displays the *LD based collapsing* theme assuming the 2nd to 6th variant loci are in LD with each other. Despite the use of different collapsing themes and the resulting regional marker coding, LOD scores of two-point linkage analysis for all panels are the same (LOD = 0.903) and is greater than the maximum of the six LOD scores when analyzing each locus separately (max. LOD = 0.301).

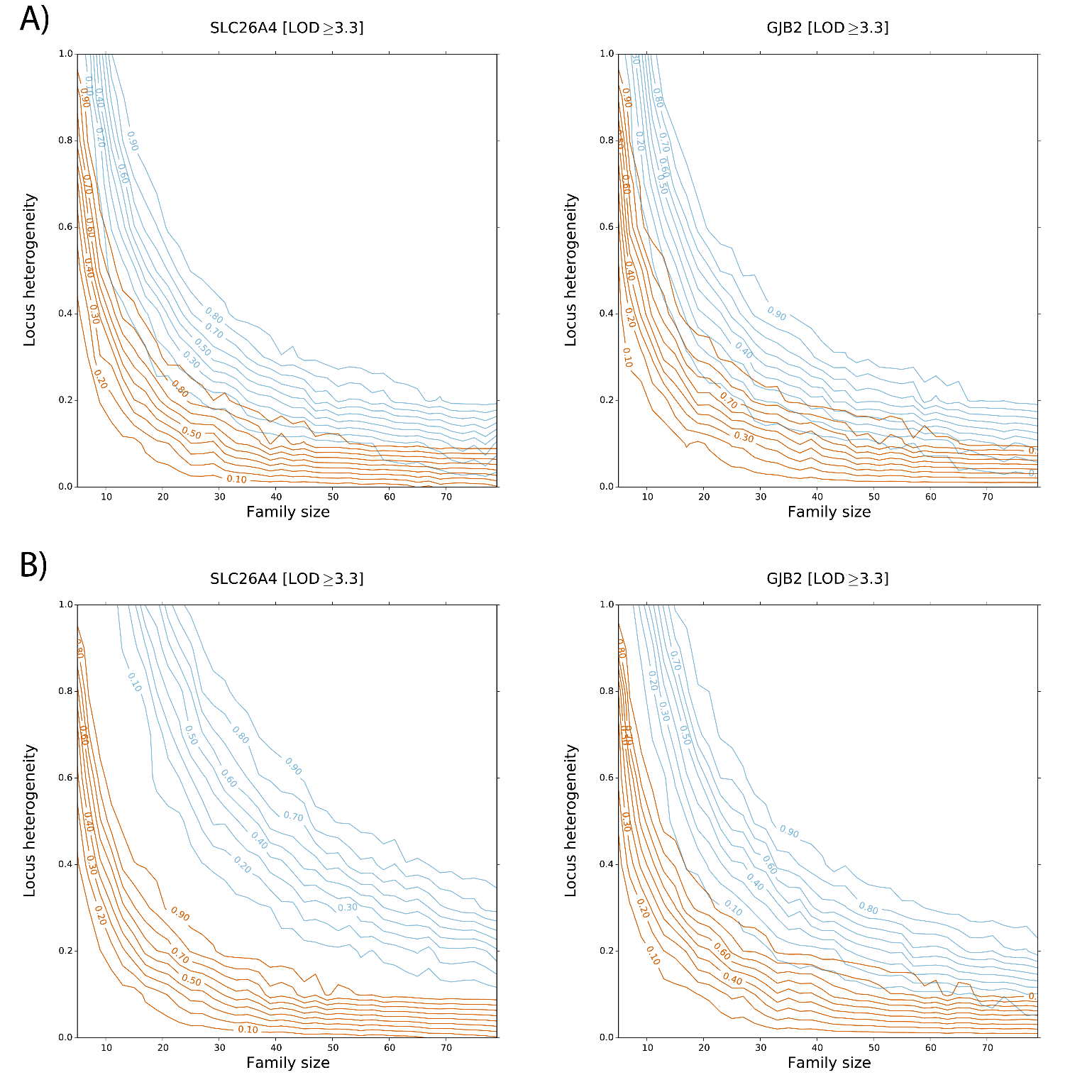


**Figure S2: Coding of haplotypes with ambiguous phase.** For the three variants displayed in the figure, suppose that the haplotypes cannot be uniquely determined due to two equally likely configurations on the 1st variant. Consequently two configurations A (lower left) and B (lower right) are both valid haplotyping result for the dataset. However under the complete collapsing theme for example, Configuration A will result in heterozygous regional genotypes which makes it potentially possible to track transmission of haplotypes to offspring, thus is more informative for linkage analysis compare to the homozygous regional genotypes resulted in Configuration B. In this example Configuration A is preferred over Configuration B. In general under such circumstances, CHP method uses whichever configuration that is potentially more informative for linkage analysis.

**Figure S3: Power comparisons for LOD statistic of collapsed haplotype pattern markers (CHP) vs. single variant markers (SVM) under recessive model.** X-axis is family size (each with 3 up to 8 offspring), Y-axis is proportion of locus heterogeneity, i.e. the proportion of families with non-syndromic hearing impairment (NSHI) caused by detrimental mutations on the gene under investigation. Contour curves on the graphs are power estimates, dark orange lines for CHP and light blue lines for SVM. Panel A displays power comparison for the two recessive genes under the assumption of no allelic heterogeneity, i.e. causal variant is the same among families; panel B displays power comparison under the assumption of allelic heterogeneity, i.e. causal variants are different among families.



**Figure S4: Power comparisons for LOD statistic of collapsed haplotype pattern markers (CHP) vs. single variant markers (SVM) under dominant model.** X-axis is family size (each with 3 up to 8 offspring), Y-axis is proportion of locus heterogeneity, i.e. the proportion of families with non-syndromic hearing impairment (NSHI) caused by detrimental mutations on the gene under investigation. Contour curves on the graphs are power estimates, dark orange lines for CHP and light blue lines for SVM. Panel A displays power comparison for the two dominant genes under the assumption of no allelic heterogeneity, i.e. causal variant is the same among families; panel B displays power comparison under the assumption of allelic heterogeneity, i.e. causal variants are different among families.



**Figure S5: Power comparisons for LOD statistic of collapsed haplotype pattern markers (CHP) vs. single variant markers (SVM) under compound recessive model.** X-axis is family size (each with 3 up to 8 offspring), Y-axis is proportion of locus heterogeneity, i.e. the proportion of families with non-syndromic hearing impairment (NSHI) caused by detrimental mutations on the gene under investigation. Contour curves on the graphs are power estimates, dark orange lines for CHP and light blue lines for SVM. Panel A displays power comparison for the two recessive genes under the assumption of no allelic heterogeneity, i.e. causal variant is the same among families; panel B displays power comparison under the assumption of allelic heterogeneity, i.e. causal variants are different among families.

**Web Resources**

America’s Families and Living Arrangements, https://www.census.gov/prod/2013pubs/p20-570.pdf

Exome Variant Server (EVS), http://evs.gs.washington.edu/EVS

Deafness Variation Database (DVD), http://deafnessvariationdatabase.com

NCBI ClinVar, https://www.ncbi.nlm.nih.gov/clinvar