Published in final edited form as:

Genet Epidemiol. 2008 November ; 32(7): 627–637. doi:10.1002/gepi.20340.

Ordered-Subset Analysis (OSA) for Family-Based Association Mapping of Complex Traits

Ren-Hua Chung¹, Silke Schmidt¹, Eden R. Martin², and Elizabeth R. Hauser^{1,*}

¹Center for Human Genetics, Duke University Medical Center, Durham, North Carolina

²Center for Genetic Epidemiology and Statistical Genetics, Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida

Abstract

Association analysis provides a powerful tool for complex disease gene mapping. However, in the presence of genetic heterogeneity, the power for association analysis can be low since only a fraction of the collected families may carry a specific disease susceptibility allele. Ordered-subset analysis (OSA) is a linkage test that can be powerful in the presence of genetic heterogeneity. OSA uses trait-related covariates to identify a subset of families that provide the most evidence for linkage. A similar strategy applied to genetic association analysis would likely result in increased power to detect association. Association in the presence of linkage (APL) is a family-based association test (FBAT) for nuclear families with multiple affected siblings that properly infers missing parental genotypes when linkage is present. We propose here APL-OSA, which applies the OSA method to the APL statistic to identify a subset of families that provide the most evidence for association. A permutation procedure is used to approximate the distribution of the APL-OSA statistic under the null hypothesis that there is no relationship between the family-specific covariate and the family-specific evidence for allelic association. We performed a comprehensive simulation study to verify that APL-OSA has the correct type I error rate under the null hypothesis. This simulation study also showed that APL-OSA can increase power relative to other commonly used association tests (APL, FBAT and FBAT with covariate adjustment) in the presence of genetic heterogeneity. Finally, we applied APL-OSA to a family study of age-related macular degeneration, where cigarette smoking was used as a covariate.

Keywords

family-based association analysis; linkage; ordered-subset analysis; covariate; genetic heterogeneity

INTRODUCTION

Genetic heterogeneity in complex human diseases may be defined in many different ways. For the purpose of this study, we define genetic heterogeneity as the presence of different susceptibility variants in different subsets of families, while the same variant increases disease risk within a given family. Thus, only a subset of the collected families may carry a specific disease susceptibility allele, while the disease phenotype in the other families may be caused by other genes or environmental factors. We are interested in the situation where a trait-related measured covariate may help distinguish between the different subsets of families. For example, in the original identification of the BRCA1 gene on chromosome 17, the age-of-onset

*Correspondence to: Elizabeth R. Hauser, Ph.D., Center for Human Genetics, Duke University Medical Center, Box 3445, Durham, NC 27710. E-mail: Elizabeth.Hauser@duke.edu

distribution differed substantially between families in which the BRCA1 variant was or was not segregating [Hall et al., 1990]. If a statistical method for complex disease gene mapping considers only a single susceptibility variant within a linked region, it may have limited power in the presence of genetic heterogeneity.

Several linkage analysis methods have been developed to account for genetic heterogeneity [Ott, 1983; Falk, 1997; Bhat et al., 1999]. Ordered-subset analysis (OSA) is a powerful linkage analysis tool that considers genetic heterogeneity based on a trait-related covariate [Hauser et al., 2004]. The covariate may help identify a genetically more homogeneous subset of families; for example, those with an earlier age at disease onset. The basic idea of the OSA method is to use a family-specific covariate, such as the average covariate value in affected family members, to identify the subset of families that provide the strongest linkage signal. Families are ordered by their covariate values and the multipoint (parametric or nonparametric) logarithm of the odds (LOD) score for each subset of families is recalculated every time a family is added into the subset. The maximum LOD score may occur at different map positions for different subsets of families. If the covariate is related to the disease, families with high- or low-ranking covariate values may be genetically more homogeneous and thus provide stronger evidence for linkage to a particular disease locus. The OSA method has been applied successfully to many complex disease studies [Allingham et al., 2005; Shah et al., 2006; Schmidt et al., 2006; Woodroffe et al., 2006]. Association study designs based on OSA-identified covariate cutoff points have also been proposed [Schmidt et al., 2008].

Family-based association designs have become popular for fine-mapping disease susceptibility genes. Family-based association analysis has an advantage over population-based association analysis since it remains a valid test when population stratification is present in the data [Spielman et al., 1993]. The transmission/disequilibrium test (TDT) proposed by Spielman et al. [1993] for case-parent trios is a popular family-based association test (FBAT). Subsequently, the TDT has been generalized to different family structures [Martin et al., 1997, 2000; Clayton, 1999; Rabinowitz and Laird, 2000]. Association in the presence of linkage (APL) is an FBAT that can test for association even in regions linked to the disease locus [Martin et al., 2003]. APL can use families with multiple affected sibs and properly infers missing parental genotypes by accounting for linkage. APL was generalized to nuclear families by implementing a bootstrap procedure for estimation of the variance of the test statistic [Chung et al., 2006]. APL was shown to have higher power than other FBATs such as pedigree disequilibrium test (PDT) [Martin et al., 1997, 2000] or FBAT [Rabinowitz and Laird, 2000] in nuclear families when parental data are not available [Martin et al., 2003; Chung et al., 2006].

Family-based association methods that incorporate covariates have previously been proposed [Lunetta et al., 2000]. FBAT allows users to provide covariate-adjusted traits by fitting a regression model to the originally measured traits. The adjusted traits are the residuals from the regression equation that models the traits as a function of covariates. They are calculated by subtracting an offset from the original trait, where the offset is the expected trait value based on fitting a general linear model under the null hypothesis [Lunetta et al., 2000]. More extensive options for calculating an offset are provided by pedigree based association test (PBAT) [Lunetta et al., 2000; Lange and Laird, 2002]. These methods work well when the goal is an adjustment of the association analysis for confounding variables. However, as in the case of linkage analysis, genetic heterogeneity may result in low power to detect allelic association. These adjustment methods are suboptimal because they do not model the relationship between covariate values and family-specific association evidence. An alternative approach that has been used for family-based association analysis is to stratify families by measured covariate values using ad hoc cutoff points [Fanous et al., 2005]. For example, in studying diseases with a wide range of age at symptom onset, such as Alzheimer or Parkinson disease, families were

stratified by average age at onset into genetically more homogeneous subgroups and separate association tests were performed within each group [Nicodemus et al., 2004; Kang et al., 2006]. However, choosing the best cutoff value to define the strata is not trivial.

In some situations, covariates such as age at onset can only be observed for affected individuals, which makes it difficult to apply standard regression models. Macgregor et al. [2006] proposed the sequential addition (SA) method that applies OSA to population-based case-control analysis when only covariates for cases are available. A more genetically homogeneous subset of cases may be identified by the SA method, which improves the statistical power to detect allelic association. The OSA method can also be applied to family-based association analysis. When genetic heterogeneity is present in the data set, some families may provide more association information than others. Moreover, populations from different genetic backgrounds may be combined in a family-based analysis. When population stratification is present in the data set, some populations may provide more association information than others. The OSA method has been applied to TDT [Perdry et al., 2007]. However, the family type used in this approach is restricted to family trios. In practice, families can have missing parents and more than one affected and unaffected sibs. Therefore, a family-based association method that can identify the subset of families providing the strongest evidence for association and is flexible for general nuclear family structures is desirable.

Here, we present APL-OSA, which is a single-locus test that applies the OSA method to the APL test statistic. APL-OSA accepts bi-allelic markers such as single nucleotide polymorphisms. APL-OSA uses a trait-related covariate to identify a subset of families that have the most extreme APL statistic for a given marker. Sharing the same property as the original OSA method, only covariate values for affected siblings are required for APL-OSA and there is no need for a pre-defined covariate cutoff value. APL-OSA then tests the null hypothesis of no correlation between the family-specific contribution to the APL statistic and the family covariate. The null hypothesis is used to test if the covariate is informative for identifying a genetically more homogeneous subset of families that carry the specific disease susceptibility alleles. We performed a comprehensive simulation study to verify the validity of the APL-OSA test. We also compared the power of APL-OSA with other association analysis methods, including APL, FBAT and FBAT with covariate adjustment, as implemented in PBAT. We then applied APL-OSA to a family study of age-related macular degeneration (AMD). We tested a marker in the gene *LOC387715*, which was implicated as a candidate gene for AMD in previous studies [Schmidt et al., 2006; Rivera et al., 2005].

METHODS

THE APL STATISTIC

APL compares the difference in the observed number of alleles in affected siblings and the expected value conditioned on parental genotypes under the null hypothesis that there is no linkage or no association. The numerator for the APL statistic T_i for a nuclear family i is calculated as [Martin et al., 2003]

$$T_i = X_i - \sum_{j \in \Psi} P(G_{pj} | G_i, A) N_{pij}, \quad (1)$$

where X_i is the number of a specific allele in affected siblings, Ψ is the set of all possible parental mating-types based on the siblings' genotypes G_i , G_{pj} is one of the possible parental mating-types, A is the affection status for the siblings and N_{pij} is the number of the specific allele in the parental mating-type G_{pj} in family i . When parental genotypes are available, Ψ is simply the observed parental mating-type. When parental genotypes are missing, the probability $P(G_p | G, A)$ in Equation (1) is calculated by incorporating the identity-by-descent

(IBD) parameters to account for linkage [Martin et al., 2003]. The numerator T_s for the APL statistic is the sum of T_i over all n families. When parents are missing, there is variability in the estimates of the parental mating-type and IBD parameters. Hence, this variability should be considered when estimating the variance of T_s . Martin et al. [2003] showed that the variance for T_s under the null hypothesis can be calculated as

$$\widehat{\text{Var}}(T_s) = V_s - \Phi = \sum_{i=1}^n T_i^2 - \Phi, \quad (2)$$

where V_s is simply the sample variance of all T_i from n families and Φ accounts for the variability in the parental mating-type and IBD parameter estimation. Chung et al. [2006] used a bootstrap approach to estimate the variance of T_s , which allows APL to use general nuclear family structures. The bootstrap approach resamples the families with replacement k times. For each resampling, the APL statistic T_s can be calculated based on the resampled replicates. Then $\widehat{\text{Var}}(T_s)$ can be obtained as the sample variance of the k resampled T_s statistics. Finally, the APL statistic takes the form

$$Z = \frac{T_s}{\sqrt{\widehat{\text{Var}}(T_s)}}. \quad (3)$$

Under the null hypothesis of no linkage or no association, the APL statistic has an asymptotic normal distribution with a mean of 0 and a variance of 1.

THE GENERAL OSA METHOD

The OSA linkage method requires a disease-related covariate and an additive linkage statistic such as a parametric LOD score or nonparametric linkage score for each family. Assume there are n families. Families are ordered by their covariate values either in descending or ascending order. Starting from an empty set S , each family is added one by one into the set in the covariate-based order. Each time a family is added into the set, the LOD score L_k is recalculated for the set of families currently in set S . After all n families have been added into the set, the maximum LOD score L^* and the map position at which it occurs are identified, and the subset of families that contribute to L^* is the *maximum subset*. The null hypothesis for the OSA test is that there is no relationship between the family-specific covariate and the family-specific linkage statistic. A permutation procedure is used to approximate the distribution of L^* under the null hypothesis and determine the empirical significance level of the observed L^* . The permutation procedure randomly orders the families and then performs the OSA analysis exactly as described above.

THE APL-OSA METHOD

In order to apply the OSA method to APL, we substitute the linkage LOD scores with the APL statistics in the OSA method. The *maximum subset* is determined by the set of families giving the most extreme APL statistic. Considering nuclear families and a bi-allelic marker, the APL-OSA method is described as follows:

1. 1. Calculate covariate x_i for each family i . Assume covariates are available for affected siblings. The covariate x_i for a family i is determined by a function of the covariates for the affected siblings in the family. The functions can be, for example, the average, the minimum or the maximum of the covariates for the affected siblings.
2. 2. Sort the families according to their covariates (either in ascending or descending order).
3. 3. Start from a set S of c families with the smallest (or largest) covariates, where $c \geq 1$. Add one family at a time in the covariate-based order into S . If the covariate values

are tied, families with the same covariate value are added at the same time. Each time a family is added into S , recalculate the APL statistic Z using the naïve variance V_s in Equation (2) based on the samples in S .

4. 4. After each family has been added into the set, choose the subset S_m with the largest $|Z|$. We only compare the absolute value of Z s since Z can be significant in either positive or negative directions. Then calculate the APL-OSA statistic Z^* according to Equation (3) with the bootstrap variance estimator based on the samples in the maximum subset S_m .
5. 5. Randomly permute the covariates $\{x_i\}$. Then go back to step 2. Repeat step 5 k times. For each time j , we obtain a $Z^{(j)}$ from Z^* in step 4.
6. 6. After performing k permutations in step 5, we will obtain k $Z^{(j)}$ values. The P -value

for the APL-OSA statistic Z^* is calculated as $\sum_{j=1}^k I(Z^{(j)} \geq Z^*) / k$, where I is an indicator function such that $I(E) = 1$ if statement E is true and $I(E) = 0$ if E is false.

In step 3, S initially has c families with the smallest (or largest) covariates. When all parental data are available so that no parameter needs to be estimated in APL, c can be a small positive integer such as 1. However, if parents are missing, a small number of c families may not be sufficient for APL to accurately estimate the allele frequency and IBD parameters. In that case, a larger number of c such as 20 is required. Moreover, identifying a very small subset of families with significant association may not be helpful for follow-up studies. Therefore, in practice, the parameter c may be chosen such that at least c families are included in the subset S_m . The same strategy was also used in Macgregor et al. [2006].

Unlike parametric LOD scores, which are additive across families, the APL statistic in Equation (3) is calculated based on a set of families and cannot be split into family-specific statistics. Therefore, in step 3 in APL-OSA, the variance for the APL statistic must be re-estimated each time a family is added into the set. We use the “naïve variance estimator” V_s in Equation (2) for step 3 in APL-OSA instead of the bootstrap variance estimator proposed in Chung et al. [2006]. Applying the bootstrap variance estimator in step 3 is computationally demanding since the estimate would have to be recalculated every time a family is added to the set. As an alternative, V_s can be estimated instantly so that it is used in step 3 in APL-OSA. After S_m is found based on V_s , the APL statistic with the bootstrap variance is calculated in step 4 so that the variance for the final APL statistic is properly estimated. In this way the desirable features of APL, i.e., use of a mixture of different nuclear family structures and use of the bootstrap variance estimator, are preserved without sacrificing the computational efficiency of APL-OSA. We use this procedure for the original covariate ordering as well as for the permutations.

Since the APL-OSA statistic is the most extreme APL statistic over the entire sample based on the covariate order, the distribution of the APL-OSA statistic cannot be the same as the distribution for the APL statistic in all families. Similarly to OSA, a permutation procedure in step 5 is used to test the significance of the APL-OSA statistic. The null hypothesis for the permutation test is that there is no relationship between the family-specific covariate values and the family-specific contributions to the APL statistic.

SIMULATION STUDY

We performed a comprehensive simulation study to demonstrate the validity of the APL-OSA statistic and to compare the power for APL-OSA with APL, FBAT and FBAT with covariate adjustment. The simulation software SIMLA [Schmidt et al., 2005] was used to generate replicates of family data. SIMLA allows users to specify different disease models and family structures. Since our primary association model is association maintained by linkage (linkage

disequilibrium or LD), all of our simulation models assume tight linkage between disease and marker loci. That is, there is no recombination between the loci.

Since the null hypothesis for APL-OSA is that there is no relationship between the family-specific contributions to the APL statistic and the family covariate values, we examined the type I error for APL-OSA under two null conditions: (1) *Null 1*: Only one population is in the sample. The covariates have no effect on the disease. However, the disease and the marker alleles are in complete LD ($r^2 = 1$). (2) *Null 2*: Two populations with different disease allele frequencies, prevalences and marker allele frequencies are mixed in the sample which results in population stratification. The covariate means are also different between the two populations. The covariates have no effect on the disease. There is no LD between the disease and the marker alleles in either population.

We simulated 500 families in each replicate of samples for *Null 1* and *Null 2* and each replicate has 30% of families with one affected and one unaffected siblings (AU), 60% of families with two affected and one unaffected siblings (AAU) and 10% of families with three affected siblings (AAA). All parents are assumed to be missing, which is commonly observed in genetic studies of late-onset diseases. This choice allows us to verify that the validity of the APL test is preserved and that missing parental genotypes can be inferred properly in APL-OSA. To simulate data under *Null 1*, the disease and the marker have the same allele frequencies and they are in complete LD. Covariates were simulated independently from a normal distribution with a mean of 40 and a standard deviation (SD) of 10. Covariates have no effect on the disease for both *Null 1* and *Null 2*. To simulate population stratification under *Null 2*, we combined one set of 100 families and the other set of 400 families. The disease allele has a main effect in one population and no effect in the other. Covariates were simulated independently from a normal distribution with a mean of 40 and SD of 10 for the first set of families and a mean of 20 and SD of 10 for the second set of families. The two sets of families have different disease allele frequencies and prevalences but have the same proportions of AU, AAU and AAA families (30, 60 and 10%, respectively). Marker allele frequencies are equal to the disease allele frequencies and there is no LD between disease and marker alleles in either population.

The disease models we used in the simulations under the null are given in Table I. We simulated dominant and recessive disease models with different disease allele frequencies and disease prevalences. The genotypic relative risk (GRR) is defined as the penetrance of the homozygous disease genotype divided by the penetrance of the homozygous wild-type genotype for a disease locus. The measure of LD between the disease and the marker alleles (r^2), sibling's recurrence-risk ratio and covariate mean and SD are also shown for each model in Table I. *Null 1* is simulated using *Models 1* and *2* in Table I. *Null 2* is simulated under *Models 3* and *4* in Table I.

To evaluate the power of APL-OSA, we simulated data under the alternative hypothesis that there is a relationship between family covariate values and family-specific contributions to the APL statistic. Genetic models used for simulations under the alternative are given in Table II. We considered the case of genetic heterogeneity, i.e., a mixture of two populations with the disease allele having a main effect in one population but no effect in the other. The disease in the population with no effect of this particular allele may be caused by other genes or environmental factors. The covariate means are different in the two populations. Hence, the covariate can distinguish between the two populations and is thus informative for the effect of the specific allele under study. We combined one set of 100 families where GRR is greater than 1 and the other set of 400 families where GRR is equal to 1. All families have two affected and one unaffected siblings (AAU). All parents are assumed to be missing in the simulated families. The two sets have different disease allele frequencies and prevalences. Covariates are simulated from a normal distribution with different means for the two sets of families. The

disease and the marker allele frequencies are the same and they are in complete LD. These conditions are simulated under *Models 5-10* in Table II with different disease models and GRR.

For age-related diseases such as Parkinson disease, the mean age of late-onset Parkinson disease was estimated as 65 with SD<10 and as 45 with SD<10 for early-onset Parkinson disease in several studies [Oliveri et al., 2001; Toft et al., 2007; Clark et al., 2004]. Hence, it is not uncommon that the covariate means for two heterogeneous populations are 2 SD apart, as our simulation setup in Table II. However, we also simulated the case that the covariate mean for population 1 is 30 instead of 40 for *Models 5-10* so that the covariate is less informative for distinguishing between the two populations. Different levels of LD between disease and marker alleles were simulated as well. *Models 11-14* in Table II have different LD measures r^2 between disease and marker alleles that are 0.75 and 0.5 for dominant and recessive models. We also simulated different proportions of families that are informative for the disease. For *Models 15* and *16* in Table II, 50% of families with GRR greater than 1 and 50% of families with GRR equal to 1 are simulated under both dominant and recessive models.

The covariate value for each family is calculated as the average of the covariate values for the affected siblings. Since all parents were assumed to be missing in the simulated data, the number c in step 3 in APL-OSA was set to 20 so that the allele frequency and IBD parameters can be properly estimated in APL. The number of bootstrap replicates in step 4 in APL-OSA was set to 200 since 200 replicates are usually enough to estimate the standard error [Efron and Tibshirani, 1993]. The number of permutations in step 5 was set to 500. In practice, if a higher precision is desired for the P -value, the number of permutations needs to be increased.

Since the application of OSA to TDT by Perdry et al. [2007] is restricted to trio families with parents, their method cannot use our simulated data of families with multiple affected siblings and missing parents. Therefore, we do not include their method in the power comparison. We are not aware of other methods that test the relationship between the family-specific covariate and family-specific contribution to the association statistic. Hence, we compared the power of APL-OSA with APL and FBAT, which test association for the overall sample, and FBAT with covariate adjustment as implemented in PBAT. The FBAT option “-e” for the empirical variance estimator was used to account for linkage. We used PBAT with the option “standard phenotypic residuals” for the offset choice, and the trait was regressed on the covariate alone to obtain these residuals.

RESULTS

TYPE I ERROR AND POWER

Table III gives the type I error rates of APL-OSA for different models. The type I error results for APL-OSA with covariates sorted in descending and ascending orders are both presented. We can see that APL-OSA has the correct type I error rates close to the 0.05 nominal level under both *Null 1* and *Null 2* for *Models 1-4* under both the low-to-high and high-to-low ordering. In particular, it is crucial for APL-OSA to maintain a correct type I error rate under *Model 3* and *Model 4* when population stratification that is not informative for allele-specific effects is present. This demonstrates that APL-OSA as an FBAT is robust to a fairly extreme extent of population stratification.

Figures 1-4 show the power comparison between APL-OSA, APL, FBAT and PBAT under different models. Since families with GRR greater than 1 were simulated with higher covariate values, APL-OSA with covariates sorted from high to low is expected to have more power than APL-OSA with covariates sorted from low to high. Therefore, only the APL-OSA results for covariates sorted from high to low are shown. Power calculated at the 0.05 significance level corresponds to the situation where there is a priori knowledge of the appropriate covariate

ordering. Power calculated at the 0.025 significance level based on the Bonferroni correction for multiple testing for the high-to-low and low-to-high covariate orderings is shown as well. This definition of power is more relevant when there is no prior information about the covariate ordering so that both the high-to-low and low-to-high orderings are evaluated. Figure 1 shows the power comparison with different GRRs for both recessive and dominant disease models (*Models 5-10*). We can see that under *Models 5-10*, APL-OSA consistently has more power than other methods even when a conservative Bonferroni correction is applied. Figure 2 shows the power comparison when the covariate mean for population 1 under *Models 5-10* is changed to 30, which is only one SD apart from the covariate mean for population 2. We can see that when the covariate means for the two populations are close, APL-OSA has similar or lower power than other methods. Figure 3 shows the power comparison with different degrees of LD between disease and marker alleles (*Models 11-14*) and illustrates that, even though the power of all methods decreases with decreasing LD, APL-OSA still maintains a higher power than the other methods.

Figure 4 shows the comparison for different proportions of families with GRR greater than 1 and equal to 1 combined into a data set for study. We can see that when there are 20% of families with GRR greater than 1 and 80% of families with GRR equal to 1 in the data, which is the case we simulated in Figures 1 and 2, APL-OSA has higher power than other methods. However, if there are 50% of families with GRR greater than 1 and also 50% of families with GRR equal to 1, APL, FBAT and PBAT have higher power than APL-OSA (*Models 15 and 16*). Based on these simulations, APL-OSA is well suited to a model of genetic heterogeneity in which association may be missed in the overall sample but detected when considering the trait-related covariate.

In Figures 1-4, we can see that APL consistently has higher power than FBAT. This agrees with the observation in Chung et al. [2006] that APL is more powerful than FBAT using nuclear families with missing parents. The power difference between APL and FBAT is especially large for the dominant model. We can also see that FBAT with the covariate adjustment implemented in PBAT can have more power than FBAT for the dominant model. However, the gain in power by covariate adjustment for FBAT is not remarkable under our simulation models and in fact covariate adjustment could obscure the covariate-related association. Therefore, when covariate levels are informative about the association evidence from a subset of families, a covariate adjustment of the traits as implemented in PBAT provides a limited power increase. On the other hand, APL-OSA can be powerful in this case since it identifies the most informative family subset and explicitly tests the relationship between the family-specific association evidence and covariate values.

ANALYSIS OF THE LOC387715 GENE FOR AMD

Previous studies identified the marker *rs10490924* in the *LOC387715* gene as significantly associated with AMD [Schmidt et al., 2006; Rivera et al., 2005]. Cigarette smoking is the most consistently identified environmental risk factor for AMD. It was previously suggested that smoking may be responsible for linkage heterogeneity on chromosome 10q26 [Weeks et al., 2004], the region in which the *LOC387715* gene is located. In a family-based study [Schmidt et al., 2006], linkage OSA was applied to multiplex families with at least two affected sampled relatives, using the high-to-low ordering of pack-years of cigarette smoking as a covariate. It was found that the majority of the linkage evidence was contributed by 40 families in which affected relatives had an average of ≥ 44 pack-years of smoking, with an OSA *P*-value of 0.048 near marker *rs10490924*. An APL analysis of *rs10490924* in all 200 families (140 multiplex, 60 singleton) was not significant. However, when the OSA-identified covariate cutoff value was used to select a subset of families with heavy smokers (46 families with an average of ≥ 44 pack-years), the result was significant at *rs10490924* ($P = 0.02$). We re-analyzed this marker

in the same family data set used in the previous study. APL and FBAT used all genotyped families, while APL-OSA and PBAT used 115 families (86 multiplex, 29 singleton) with available covariate data. Since most of the families had missing parents, the c value in step 3 in APL-OSA was set to 20. The test was only performed based on the high-to-low covariate ordering because a previous study demonstrated evidence for linkage heterogeneity on the basis of that ordering [Weeks et al., 2004]. The results are given in Table IV. APL-OSA was significant for marker *rs10490924* ($P = 0.028$) and identified a subset of 50 families with a covariate cutoff value of 41 pack-years, which is very similar to the previous result [Schmidt et al., 2006]. PBAT, APL and FBAT did not show significant ($P < 0.05$) results for this marker, despite the larger sample size contributing to the APL and FBAT test statistics. Our re-analysis shows that APL-OSA can detect genetic association in a heterogeneous data set and recapitulates the results of the original stratified association analysis.

DISCUSSION

In the presence of genetic heterogeneity, association analysis may have low power to detect any one disease allele since its effect may vary across the collected families. Identifying a subset of families in which this allele has the strongest effect can increase the power of the association test. We have developed APL-OSA, which can detect a subset of families that provide the most association information on the basis of a measured trait-related covariate. APL-OSA requires covariates only for affected siblings. Therefore, it is applicable to situations when covariate values, such as age at onset, are available for only affected individuals. Like the original APL method, APL-OSA can use nuclear families with multiple affected siblings and can correctly infer missing parental genotypes when linkage is present. Our simulation results showed that APL-OSA has the correct type I error rate under the null hypothesis of no correlation between covariate values and family-specific contributions to the APL statistics. Our power analysis showed that APL-OSA can have more power than other association methods, such as APL, FBAT and FBAT with covariate adjustment, in the presence of genetic heterogeneity.

Common association tests such as TDT and APL assume no association or no linkage under the null and rejecting the null implies both linkage and association. The null hypothesis for APL-OSA is that there is no relationship between covariates and association test statistics. Rejecting the null in APL-OSA implies that there is correlation between the family-specific covariates and association statistics. There are different alternative hypotheses that may lead to a rejection of this null hypothesis, including genetic heterogeneity as defined here, gene-environment interaction and quantitative trait locus models, as discussed previously [Schmidt et al., 2007]. In our simulation models, a rejection implies that the APL covariate is informative for identifying the subset of families in which the marker allele is in LD with the disease susceptibility allele of interest. As demonstrated in our simulation studies under *Null 1* with random covariates, even though the disease and the marker alleles are in complete LD, APL-OSA appropriately does *not* detect the covariate-APL relationship. Under *Null 2* that the covariate is informative for genetic heterogeneity, APL-OSA still maintains the correct type I error rate because there is no LD between disease and marker alleles. Thus, APL-OSA retains the desirable property of robustness to population stratification, even when the covariate itself is different between the populations.

The power of APL-OSA is determined by several factors. Since APL-OSA tests the relationship between family-specific covariates and association test statistics, choosing meaningful covariates is very important. When the covariate levels are not informative enough to identify genetic heterogeneity in the data, APL-OSA has similar or lower power than APL and FBAT (Fig. 2). However, APL-OSA provides more information than other methods such as the covariate cutoff value and the subset of more genetically homogeneous families that will be

helpful for follow-up studies. On the other hand, as seen in Figure 4, if the disease allele does increase susceptibility in the majority of families, these families can provide sufficient association information for tests like APL and FBAT that do not consider genetic heterogeneity. Therefore, methods that do not consider genetic heterogeneity can have more power than APL-OSA when the disease allele is present in a large fraction of the families. The same property also holds for OSA [Hauser et al., 2004] and was also observed in theoretical studies of genetic heterogeneity [Leal and Ott, 2000]. In practice, association testing in the full data set along with APL-OSA may need to be applied at the same time. Moreover, since the relationship between the covariate and the APL statistics is typically unknown in real data sets, both ascending and descending covariate orders should be evaluated and an adjustment for multiple testing for both covariate orders may be required [Hauser et al., 2004; Schmidt et al., 2007].

In the AMD example, APL-OSA was significant for marker *rs1049024* using smoking pack-years as a family covariate, while APL and FBAT applied to the entire sample of families were not significant. This illustrates that APL-OSA can have more power due to identifying a more homogeneous subset of families with heavy smokers. In this example, a covariate adjustment with PBAT did not generate a significant result. Consistent with our simulation study, the covariate adjustment implemented in PBAT may not be the optimal method for modeling the type of genetic heterogeneity that appears to exist in the AMD data set.

Unlike linkage OSA, which identifies the maximum family subset starting from an empty set, a larger number of families may be required for APL-OSA in the initial subset so that the allele frequency and IBD parameters can be properly estimated in APL when parental genotypes are missing. Our simulation study showed that when all parents are missing, using c equal to 1 in step 3 in APL-OSA can cause a high type I error rate for APL-OSA (data not shown). According to our simulation studies such as *Models 1-4* where all parents are missing, using 20 families as an initial subset is sufficient for parameter estimation in APL.

Continuous disease-related covariates are most suitable for use in APL-OSA. Discrete covariates may not be appropriate if there are few (<10) permutable categories. Statistics calculated based on the trait can also be used as a covariate for APL-OSA. For example, when a follow-up association test is performed on the same data set previously used for linkage analysis, the LOD scores may be used as covariates for APL-OSA. This will not affect the validity of APL-OSA since linkage and FBAT statistics are not correlated when there is either no linkage or no association [Chung et al., 2007]. Association test statistics for another marker can be used as covariates for APL-OSA to test for gene-gene interactions [Shah et al., 2005]. Furthermore, if there is correlation in covariate values within a family, APL-OSA remains valid since the family-specific covariate is calculated as a function of the covariate values in all affected siblings. Extensive simulation studies demonstrated the validity of APL-OSA for different covariate correlation patterns within nuclear family structures (data not shown). In our simulation studies, the covariate value for each family is calculated as the average of the covariate values for the affected siblings. Other strategies for calculating the covariate for each family to achieve the maximum power will be investigated in future studies.

Both bootstrap and permutation approaches are used in APL-OSA. These algorithms can be time consuming. Some optimizations, as discussed in Hauser et al. [2004], can be used to reduce the number of permutations. To improve the computational efficiency of our algorithm, we used the APL statistics calculated from the naïve variance estimator V_s to select the *maximum subset* S_m . V_s can be larger than the true variance according to Equation (2). Then the standardized APL statistic using V_s will have a lower variance. However, using the true variance estimated from the bootstrap procedure to select S_m is not practical because of the computational burden. For example, if there are n families in the sample, the total number of bootstrap iterations in the APL-OSA algorithm can be as high as $n \times p \times k$, where p is the

number of bootstrap iterations for each variance estimate and k is the number of permutations. Since p and k are usually required to be relatively large, $n \times k \times p$ can be a huge number in practice. V_s can be calculated instantly and the number of bootstrap iterations is reduced to $k \times p$. According to our simulation studies, V_s provides a reasonable initial estimate to select S_m . Moreover, once S_m is selected, the APL statistic based on the bootstrap variance estimate is calculated in step 4 of the APL-OSA algorithm so that its variance is properly estimated. Therefore, using V_s in selecting S_m does not affect the validity of the APL-OSA statistic. The power of APL-OSA may be reduced since S_m identified by using the APL statistic with V_s may not be the same as the subset identified by using the APL statistic with the bootstrap variance estimate. However, according to our simulation studies, even when all parents are missing, the subset of families selected using either V_s or the bootstrap variance estimator are typically very similar (data not shown). It is also worth noting that although V_s can be used to identify the subset, it cannot replace the bootstrap variance estimator in step 4 of the APL-OSA algorithm to calculate the APL-OSA statistic. Simulation results showed that the type I error rates for APL-OSA can be inflated using V_s in step 4 (data not shown).

Currently, APL-OSA can only use one covariate at a time to identify S_m . It is possible to extend APL-OSA to use multiple covariates. An APL-OSA statistic can be obtained for each covariate. A global test statistic can be developed to test the significance of the APL-OSA statistics obtained from the set of covariates. Moreover, APL-OSA considers one marker at a time for testing association. It is also possible to extend APL-OSA to select the most significant set of markers within a given genomic region. Several association tests for multiple markers have been proposed [Chung et al., 2006; Horvath et al., 2004; Liang and Chiu, 2005; Xu et al., 2006; Rakovski et al., 2007]. A multimarker test for APL can also be developed. Then a strategy similar to that used for the original OSA method can be used to rank markers and select markers with the strongest association signal. Extending APL-OSA to multiple covariates and multiple markers will be the focus of our future work.

In summary, APL-OSA provides a powerful tool for association analysis in the presence of genetic heterogeneity. It provides a valid and robust test for the correlation between family-specific covariate and family-specific contribution to the association statistic. We have implemented APL-OSA as computer software using C++ for Windows, Linux and Solaris platforms. The input for the software consists of a standard .dat file and a modified .ped file with covariates appended for each individual. The output consists of a results file for S_m and a file listing the families included in S_m . The covariate cutoff point for the subset of families is reported in the results file. This cutoff point may be used to stratify samples for a replication study [Schmidt et al., 2007]. The APL-OSA program can be freely downloaded at the Duke Center for Human Genetics Web site.

WEB RESOURCES

Duke Center for Human Genetics, <http://www.chg.duke.edu/research/software.html> (for APL and APL-OSA software packages).

FBAT, <http://biosun1.harvard.edu/~fbat/fbat.htm> (for FBAT).

PBAT, <http://www.biostat.harvard.edu/~clange/default.htm> (for PBAT).

ACKNOWLEDGMENTS

We gratefully acknowledge our colleagues Xuemei Lou and Dr. Xuejun Qin for helpful discussions and software testing. The AMD study was made possible by a grant (NEI R01 EY12118) to Drs. Margaret Pericak-Vance (University of Miami) and Jonathan Haines (Vanderbilt University), with additional support for methodological research (NEI R03 EY015216 to S.S.).

Contract grant sponsor: National Institutes of Health (NIMH); Contract grant number: R01 MH595228; Contract grant sponsor: Neurosciences Education and Research Foundation.

REFERENCES

- Allingham RR, Wiggs JL, Hauser ER, Larocque-Abramson KR, Santiago-Turla C, Broome B, Del Bono EA, Graham FL, Haines JL, Pericak-Vance MA, Hauser MA. Early adult-onset POAG linked to 15q11-13 using ordered subset analysis. *Invest Ophthalmol Vis Sci* 2005;46:2002–2005. [PubMed: 15914615]
- Bhat A, Heath SC, Ott J. Heterogeneity for multiple disease loci in linkage analysis. *Hum Hered* 1999;49:229–231. [PubMed: 10436385]
- Chung RH, Hauser ER, Martin ER. The APL test: extension to general nuclear families and haplotypes and examination of its robustness. *Hum Hered* 2006;61:189–199. [PubMed: 16877866]
- Chung RH, Hauser ER, Martin ER. Interpretation of simultaneous linkage and family-based association tests in genome screens. *Genet Epidemiol* 2007;31:134–142. [PubMed: 17123303]
- Clark LN, Afridi S, Mejia-Santana H, Harris J, Louis ED, Cote LJ, Andrews H, Singleton A, De Vrieze FW, Hardy J, Mayeux R, Fahn S, Waters C, Ford B, Frucht S, Ottman R, Marder K. Analysis of an early-onset Parkinson's disease cohort for DJ-1 mutations. *Mov Disord* 2004;19:796–800. [PubMed: 15254937]
- Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999;65:1170–1177. [PubMed: 10486336]
- Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*. 57th edition. Chapman & Hall; New York: 1993.
- Falk C. Effect of genetic heterogeneity and assortative mating on linkage analysis: a simulation study. *Am J Hum Genet* 1997;61:1169–1178. [PubMed: 9345086]
- Fanous AH, van den Oord EJ, Riley BP, Aggen SH, Neale MC, O'Neill FA, Walsh D, Kendler KS. Relationship between a high-risk haplotype in the DTNBP1 (dysbindin) gene and clinical features of schizophrenia. *Am J Psychiatry* 2005;162:1824–1832. [PubMed: 16199828]
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990;250:1684–1689.
- Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* 2004;27:53–63. [PubMed: 15185403]
- Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 2004;26:61–69. [PubMed: 14691957]
- Kang SJ, Scott WK, Li YJ, Hauser MA, van der Walt JM, Fujiwara K, Mayhew GM, West SG, Vance JM, Martin ER. Family-based case-control study of MAOA and MAOB polymorphisms in Parkinson disease. *Mov Disord* 2006;21:2175–2180. [PubMed: 17044053]
- Lange C, Laird NM. On a general class of conditional tests for family-based association studies in genetics: the asymptotic distribution, the conditional power, and optimality considerations. *Genet Epidemiol* 2002;23:165–180. [PubMed: 12214309]
- Leal SM, Ott J. Effects of stratification in the analysis of affected sib-pair data: benefits and costs. *Am J Hum Genet* 2000;66:567–575. [PubMed: 10677317]
- Liang KY, Chiu YF. Multipoint linkage disequilibrium mapping using case-control designs. *Genet Epidemiol* 2005;29:365–376. [PubMed: 16294298]
- Lunetta KL, Faraone SV, Biederman J, Laird NM. Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions. *Am J Hum Genet* 2000;66:605–614. [PubMed: 10677320]
- Macgregor S, Craddock N, Holmans PA. Use of phenotypic covariates in association analysis by sequential addition of cases. *Eur J Hum Genet* 2006;14:529–534. [PubMed: 16538225]
- Martin ER, Kaplan NL, Weir BS. Tests for linkage and association in nuclear families. *Am J Hum Genet* 1997;61:439–448. [PubMed: 9311750]
- Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000;67:146–154. [PubMed: 10825280]

- Martin ER, Bass MP, Hauser ER, Kaplan NL. Accounting for linkage in family-based tests of association with missing parental genotypes. *Am J Hum Genet* 2003;73:1016–1026. [PubMed: 14551902]
- Nicodemus KK, Stenger JE, Schmechel DE, Welsh-Bohmer KA, Saunders AM, Roses AD, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA, Martin ER. Comprehensive association analysis of APOE regulatory region polymorphisms in Alzheimer disease. *Neurogenetics* 2004;5:201–208. [PubMed: 15455263]
- Oliveri RL, Zappia M, Annesi G, Bosco D, Annesi F, Spadafora P, Pasqua AA, Tomaino C, Nicoletti G, Pirritano D, Labate A, Gambardella A, Logroscino G, Manobianca G, Epifanio A, Morgante L, Savettieri G, Quattrone A. The parkin gene is not involved in late-onset Parkinson's disease. *Neurology* 2001;57:359–362. [PubMed: 11468333]
- Ott J. Linkage analysis and family classification under heterogeneity. *Ann Hum Genet* 1983;47:311–320. [PubMed: 6651220]
- Perdry H, Maher B, Babron M-C, McHenry T, Clerget-Darpoux F, Marazita ML. An ordered subset approach to including covariates in transmission disequilibrium. *BMC Proceedings* 2007;1:S77. [PubMed: 18466579]
- Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 2000;50:211–223. [PubMed: 10782012]
- Rakovski CS, Xu X, Lazarus R, Blacker D, Laird NM. A new multimarker test for family-based association studies. *Genet Epidemiol* 2007;31:9–17. [PubMed: 17086514]
- Rivera A, Fisher SA, Fritsche LG, Keilhauer CN, Lichtner P, Meitinger T, Weber BH. Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum Mol Genet* 2005;14:3227–3236. [PubMed: 16174643]
- Schmidt M, Hauser E, Martin E, Schmidt S. Extension of the SIMLA package for generating pedigrees with complex inheritance patterns: environmental covariates, gene-gene and gene-environment interaction. *Stat Appl Genet Mol Biol* 2005;4[http://www.be-press.com/sagmb/vol4/iss1/]
- Schmidt S, Hauser MA, Scott WK, Postel EA, Agarwal A, Gallins P, Wong F, Chen YS, Spencer K, Schnetz-Boutaud N, Haines JL, Pericak-Vance MA. Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am J Hum Genet* 2006;78:852–864. [PubMed: 16642439]
- Schmidt S, Qin X, Schmidt MA, Martin ER, Hauser ER. Interpreting analyses of continuous covariates in affected sibling pair linkage studies. *Genet Epidemiol* 2007;31:541–552. [PubMed: 17410529]
- Schmidt S, Schmidt MA, Qin XJ, Martin ER, Hauser ER. Increased efficiency of case-control association analysis by using allele-sharing and covariate information. *Hum Hered* 2008;65:154–165. [PubMed: 17934318]
- Shah SH, Schmidt MA, Mei H, Scott WK, Hauser ER, Schmidt S. Searching for epistatic interactions in nuclear families using conditional linkage analysis. *BMC Genet* 2005;6:S148. [PubMed: 16451608]
- Shah SH, Kraus WE, Crossman DC, Granger CB, Haines JL, Jones CJ, Mooser V, Huang L, Haynes C, Dowdy E, Vega GL, Grundy SM, Vance JM, Hauser ER. Serum lipids in the GENECARD study of coronary artery disease identify quantitative trait loci and phenotypic subsets on chromosomes 3q and 5q. *Ann Hum Genet* 2006;70:738–748. [PubMed: 17044848]
- Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506–516. [PubMed: 8447318]
- Toft M, Myhre R, Pielsticker L, White LR, Aasly JO, Farrer MJ. PINK1 mutation heterozygosity and the risk of Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2007;78:82–84. [PubMed: 17172567]
- Weeks DE, Conley YP, Tsai HJ, Mah TS, Schmidt S, Postel EA, Agarwal A, Haines JL, Pericak-Vance MA, Rosenfeld PJ, Paul TO, Eller AW, Morse LS, Dailey JP, Ferrell RE, Gorin MB. Age-related maculopathy: a genomewide scan with continued evidence of susceptibility loci within the 1q31, 10q26, and 17q25 regions. *Am J Hum Genet* 2004;75:174–189. [PubMed: 15168325]
- Woodroffe A, Krafchak CM, Fuse N, Lichter PR, Moroi SE, Schertzer R, Downs CA, Duren WL, Boehnke M, Richards JE. Ordered subset analysis supports a glaucoma locus at GLC1I on

chromosome 15 in families with earlier adult age at diagnosis. *Exp Eye Res* 2006;82:1068–1074. [PubMed: 16298363]

Xu X, Rakovski C, Xu XP, Laird N. An efficient family-based association test using multiple markers. *Genet Epidemiol* 2006;30:620–626. [PubMed: 16868964]

Fig. 1. Power comparison between APL-OSA, APL, FBAT and PBAT over 500 replicates. Power is calculated for significance level of 0.05 (APL-OSA_0.05, APL, FBAT and PBAT) and 0.025 (APL-OSA_0.025) for recessive (*Models 5-7*) and dominant (*Models 8-10*) models with different GRRs. APL-OSA, association in the presence of linkage-ordered-subset analysis; FBAT, family-based association test; GRR, genotypic relative risk.

Fig. 2.

Power comparison between APL-OSA, APL, FBAT and PBAT over 500 replicates. Power is calculated for significance level of 0.05 (APL-OSA_0.05, APL, FBAT and PBAT) and 0.025 (APL-OSA_0.025) for recessive (*Models 5-7*) and dominant (*Models 8-10*) models with different GRRs. The covariate means and standard deviations for the population 1 in *Models 5-10* are set to 30 and 10, respectively. APL-OSA, association in the presence of linkage-ordered-subset analysis; FBAT, family-based association test; GRR, genotypic relative risk.

Fig. 3.

Power comparison between APL-OSA, APL, FBAT and PBAT over 500 replicates. Power is calculated for significance level of 0.05 (APL-OSA_0.05, APL, FBAT and PBAT) and 0.025 (APL-OSA_0.025) for recessive (*Models 7, 11 and 12*) and dominant (*Models 10, 13 and 14*) models with different levels of LD. APL-OSA, association in the presence of linkage-ordered-subset analysis; FBAT, family-based association test; LD, linkage disequilibrium.

Fig. 4.

Power comparison between APL-OSA, APL, FBAT and PBAT over 500 replicates. Power is calculated for significance level of 0.05 (APL-OSA_0.05, APL, FBAT and PBAT) and 0.025 (APL-OSA_0.025) for recessive (*Models 5 and 15*) and dominant (*Models 8 and 16*) models with different proportions of population admixture. APL-OSA, association in the presence of linkage-ordered-subset analysis; FBAT, family-based association test.

TABLE 1

Genetic models for simulations under the null

Model number	Populations	Disease model	Disease allele frequency	Genotypic relative risk (GRR)	Prevalence	r^{2a}	Recurrence-risk ratio	Covariate mean, SD ^b
1	Pop. 1	Recessive	0.25	5	0.05	1	1.153	40, 10
2	Pop. 1	Dominant	0.15	5	0.05	1	1.306	40, 10
3	Pop. 1 (20%) ^c	Recessive	0.25	5	0.05	0	1.153	40, 10
	Pop. 2 (80%)	Recessive	0.35	1	0.1	0	1	20, 10
4	Pop. 1 (20%)	Dominant	0.15	5	0.05	0	1.306	40, 10
	Pop. 2 (80%)	Dominant	0.35	1	0.1	0	1	20, 10

LD, linkage disequilibrium.

^aLD measure between disease and marker alleles.

^bCovariates are simulated from a normal distribution with mean and standard deviation (SD) as specified.

^cProportions of populations in the sample.

TABLE II

Genetic models for power comparison

Model number	Populations	Disease model	Disease allele frequency	Genotypic relative risk (GRR)	Prevalence	r^{2a}	Recurrence risk ratio	Covariate mean, SD ^b			
5	Pop. 1 (20%) ^c	Recessive	0.25	3	1.4 ^d	0.05	0.09 ^e	1	1.051	1.01 ^f	40, 10
	Pop. 2 (80%)	Recessive	0.35	1		0.1			1		20, 10
6	Pop. 1 (20%)	Recessive	0.25	4	1.6	0.05	0.09	1	1.098	1.02	40, 10
	Pop. 2 (80%)	Recessive	0.35	1		0.1			1		20, 10
7	Pop. 1 (20%)	Recessive	0.25	5	1.8	0.05	0.09	1	1.153	1.03	40, 10
	Pop. 2 (80%)	Recessive	0.35	1		0.1			1		20, 10
8	Pop. 1 (20%)	Dominant	0.15	3	1.4	0.05	0.09	1	1.139	1.03	40, 10
	Pop. 2 (80%)	Dominant	0.35	1		0.1			1		20, 10
9	Pop. 1 (20%)	Dominant	0.15	4	1.6	0.05	0.09	1	1.227	1.04	40, 10
	Pop. 2 (80%)	Dominant	0.35	1		0.1			1		20, 10
10	Pop. 1 (20%)	Dominant	0.15	5	1.8	0.05	0.09	1	1.306	1.06	40, 10
	Pop. 2 (80%)	Dominant	0.35	1		0.1			1		20, 10
11	Pop. 1 (20%)	Recessive	0.25	5	1.8	0.05	0.09	0.75	1.153	1.03	40, 10
	Pop. 2 (80%)	Recessive	0.35	1		0.1			1		20, 10
12	Pop. 1 (20%)	Recessive	0.25	5	1.8	0.05	0.09	0.5	1.153	1.03	40, 10
	Pop. 2 (80%)	Recessive	0.35	1		0.1			1		20, 10
13	Pop. 1 (20%)	Recessive	0.15	5	1.8	0.05	0.09	0.75	1.306	1.06	40, 10
	Pop. 2 (80%)	Dominant	0.35	1		0.1			1		20, 10
14	Pop. 1 (20%)	Dominant	0.15	5	1.8	0.05	0.09	0.5	1.306	1.06	40, 10
	Pop. 2 (80%)	Dominant	0.35	1		0.1			1		20, 10
15	Pop. 1 (50%)	Recessive	0.25	3	2	0.05	0.075	1	1.051	1.03	40, 10
	Pop. 2 (50%)	Recessive	0.35	1		0.1			1		20, 10
16	Pop. 1 (50%)	Dominant	0.15	3	2	0.05	0.075	1	1.139	1.07	40, 10
	Pop. 2 (50%)	Dominant	0.35	1		0.1			1		20, 10

LD, linkage disequilibrium.

^aLD measure between disease and marker alleles.

^bCovariates are simulated from a normal distribution with mean and standard deviation (SD) as specified.

^cProportions of populations in the sample.

^dRelative risk for the overall sample.

^ePrevalence for the overall sample.

^fRecurrence risk ratio for the overall sample.

TABLE III

Type I error for APL-OSA

Models	Number of families	Hypothesis	Type I error (L→H) ^a $\alpha = 0.05$	Type I error (H→L) ^b $\alpha = 0.05$
1	500 ^c	Null 1	0.055	0.048
2	500 ^c	Null 1	0.048	0.048
3	500 ^d	Null 2	0.054	0.051
4	500 ^d	Null 2	0.051	0.049

Three thousand replicates were simulated for each model. APL-OSA, association in the presence of linkage-ordered subset analysis.

^a Covariates are sorted from low to high.

^b Covariates are sorted from high to low.

^c 150 AUU, 300 AAU, 50 AAA. All samples have covariate mean = 40, SD = 10.

^d 150 AUU, 300 AAU, 50 AAA. Twenty percent of the families have covariate mean = 40, SD = 10. Eighty percent of the families have covariate mean = 20, SD = 10.

TABLE IVAPL-OSA, PBAT, APL and FBAT results for *LOC387715* analysis

Marker <i>rs10490924</i>	<i>P</i> -value
APL-OSA ^a	0.028 ^b
PBAT	0.122
APL	0.147
FBAT	0.063
APL-OSA, association in the presence of linkage-ordered subset analysis; FBAT, family-based association test.	

^a Covariate is sorted from high to low. Cutoff value for the covariate is 41 and there are 50 families in the subset.

^b *P*-value was calculated based on 10,000 permutations.