Genomics & Informatics Vol. 5(2) 61-67, June 2007

Controlling Linkage Disequilibrium in Association Tests: Revisiting *APOE* Association in Alzheimer's Disease

Leeyoung Park*

Genomic Epidemiology Division, Center for Genome Science, National Institute of Health, KCDC, Seoul 122-701, Korea

Abstract

The allele frequencies of markers as well as linkage disequilibrium (LD) can be changed in cases due to the LD between markers and the disease allele, exhibiting spurious associations of markers. To identify the true association, classical statistical tests for dealing with confounders have been applied to draw a conclusion as to whether the association of variants comes from LD with the known disease allele. However, a more direct test considering LD using estimated haplotype frequencies may be more efficient. The null hypothesis is that the different allele frequencies of a variant between cases and controls come solely from the increased disease allele frequency and the LD relationship with the disease allele. The haplotype frequencies of controls are estimated using the expectation maximization (EM) algorithm from the genotype data. The estimated frequencies are applied to calculate the expected haplotype frequencies in cases corresponding to the increase or decrease of the causative or protective alleles. The suggested method was applied to previously published data, and several *APOE* variants showed association with Alzheimer's disease independent from the *APOE* ε4 variant, rs429358, regardless of LD showing significant simulated p-values. The test results support the possibility that there may be more than one common disease variant in a locus.

Keywords: linkage disequilibrium, confounder, haplotype, population, common variant, low penetrance, *APOE*, Alzheimer's disease

Introduction

The inconsistent results of association studies relating to complex traits necessitate the development of precise approaches to the genetic studies of complex traits. Following the successes of the HapMap project (Altshuler *et al*., 2005), the recent focus in this field has been to find common low-penetrance variants which increase risk in population. This is achieved by using the linkage disequilibrium with the disease variant in case-control association studies. The advantage of linkage disequilibrium mapping using case-controls is better detection of common disease alleles in a population, rather than family specific disease alleles. With high linkage disequilibrium, the polymorphisms that are closely linked together in a population form regions that are termed linkage disequilibrium (LD) blocks. Within a LD block, several marker alleles in linkage disequilibrium with the risk allele are sufficient to identify the association of a complex trait with the disease polymorphisms in the region (Weiss and Clark, 2002). The usual measurement of LD is D' and r^2, and is generally considered as significant if D' is higher than .7 and r^2 is higher than .3 (Ardlie *et al*., 2002; Weiss and Clark, 2002). The former indicates scaled linkage disequilibrium, and the latter provides information about the correlation between two alleles.

Many genetic association studies of complex traits utilize a set of several haplotype-tag SNPs for testing the association of LD-block haplotypes with diseases. Generally, more sensitive detection in the association tests using haplotypes is expected due to the higher possibility of capturing the disease allele within haplotypes with well-defined tags rather than a single marker (Chapman *et al*., 2003; de Bakker *et al*., 2005; Johnson *et al*., 2001). It should be noted that the detection of association purely depends on the linkage disequilibrium state of the disease allele with the haplotypes, and the effective size of the disease allele (Chapman *et al*., 2003; De La Vega *et al*., 2005). There are many helpful studies regarding LD which rely on tagging methods and powerful calculations (de Bakker *et al*., 2005; Gordon *et al*., 2002). However, it should be noted that, due to the influence of Mendelian genetics, it tends to be underestimated that there can be several common disease variants in a gene region. In the case of common biallelic variants, if one allele is risky, the other has a protective effect. The existence of such common disease variants can cause a severe decrease in the detection power from the combination of disease alleles, since an individual will have a set of relevant disease alleles. Using the current association method, it is hard to distinguish which variants are the real causes of disease

*Corresponding author: E-mail ly_park@hanmail.net
Tel +82-2-380-2243, Fax +82-2-354-1069
Accepted 31 May 2007

among the variants in high LD especially using r^2 if the disease variant does not prove to be functional. Large sample sizes and replications can help with the interpretation of results, but the intrinsic problems still remain in situations where there are several disease variants in a LD block.

Statistically, these problems can be considered as confounding factors and effect modifiers, and can be dealt with using conventional statistical methods. Confounding factors create confusion in data sets, so need to be removed by controlling the factors responsible, but the effect modifier creates a modifying effect on the outcome due to the effects of itself on the main variable. There are several classical methods for dealing with these kinds of problems; randomization, stratification (specification), matching, multivariate regression, and propensity score. In the case of linkage disequilibrium, the problem is quite different from those, since the relationship between the confounding factors and the real associated factors are known. Unlike the usual statistical problems, the analytical method can be easily derived from the nature of the relationship, and will not reduce the detection power, underestimate or overestimate the interpretation. To test the independence of association from linked disease variants, a more direct way of controlling linkage disequilibrium is suggested.

Methods

Linkage Disequilibrium Properties: Changes due to a disease allele

Linkage disequilibrium is important in finding common disease variants from population data, but it can be a problem eliminating the confounding associations while finding the real disease variants. In this study, the basic behaviors of changes in LD and allele frequencies between cases and controls were studied depending on the existence of a disease variant in a locus. The assumption is that cases and controls are collected from the same population, and the LD in the controls represents the LD in the general population, which is often the case. For very common diseases (or traits), a correction of LD in controls considering the whole population may be necessary. Using the assumption, the changes in marker frequencies and LD are observed naturally during the sampling procedure of cases. Due to the LD between markers and the disease variant, the changes in allele frequencies as well as LD can be observed even when the r^2 or D' are not very high.

Here is an example of two markers, A and B, with minor alleles indicated in lower case and major alleles in upper case. Frequencies of all four possible haplotypes, AB, aB, Ab, and ab, can be expressed by one of the haplotype

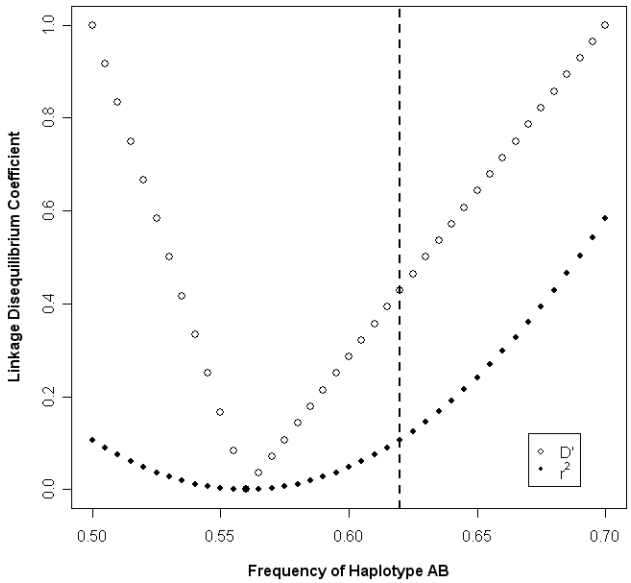

Fig. 1. LD changes due to frequencies of haplotype AB when minor allele frequencies of A and B are 0.3 and 0.2, respectively.

frequencies when the allele frequencies are fixed. The possible haplotype frequencies of AB are limited depending on their minor allele frequencies as shown in equation (1). The change in LD due to frequencies of haplotype AB is plotted in Fig. 1, where the minor allele frequencies of A and B are 0.2 and 0.3, respectively. The lowest LD is obtained when p(AB) is equal to p(A)p(B), in this case $0.8 \times 0.7 = 0.56$

$$1-p(a)-p(b) \le p(AB) \le 1-\max(p(a),\ p(b)) \qquad (1)$$

Table 1. Haplotype frequency changes between cases and controls due to a disease variant

Haplotypes	controls	cases	Changes
ABC	0.60	0.500	-0.100
aBC	0.15	0.125	-0.025
AbC	0.05	0.042	-0.008
ABc	0.02	0.050	0.030
abC	0.10	0.083	-0.017
aBc	0.03	0.075	0.045
Abc	0.03	0.075	0.045
abc	0.02	0.050	0.030
Total	1.00	1.000	0.000

Table 2. LD between variants considering a disease variant (C and c) in cases and controls; A and B are markers

LD	Controls			Cases		
	AB	AC	BC	AB	AC	BC
D'	0.429	0.286	0.375	0.298	0.250	0.375
r^2	0.107	0.021	0.062	0.059	0.042	0.111

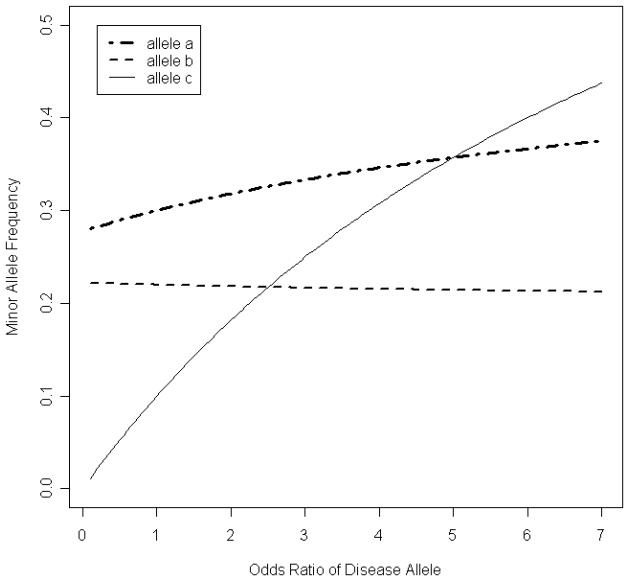

Fig. 2. Change of minor allele frequencies depending on odds ratio (c is the disease allele).

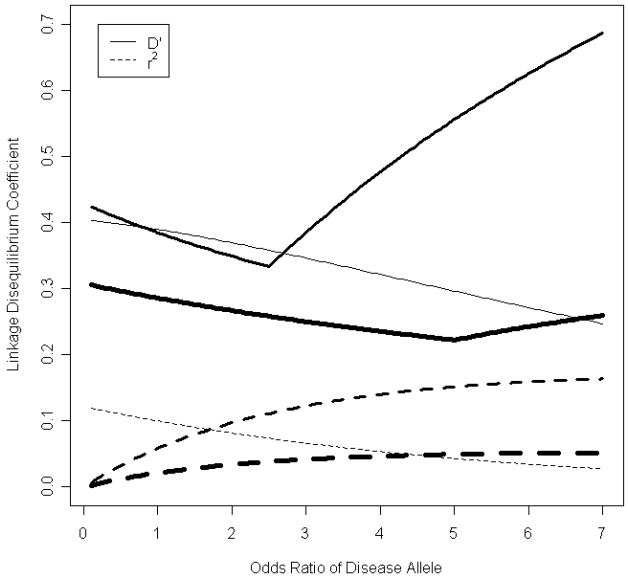

Fig. 3. Change of LDs between the three variants in cases depending on the odds ratio of the disease allele c (thin line: LD between A and B, regular line: LD between B and C, thick line: LD between A and C).

For a disease variant, C, with the odds ratio of 3 and the minor allele frequency of 0.1, the changes of haplotype frequencies and LD are calculated in Table 1 and 2, where p(AB) is 0.62, as indicated by a vertical dashed line in Fig. 1. The rest of the haplotype frequencies in the controls for all three variants are anonymously assigned in Table 1. In this example, the disease allele frequency in cases is 0.25 with an increment of 0.15 for minor disease alleles. The expected haplotype frequencies depend on the haplotype frequencies in controls, the proportion of the haplotype among the total haplotypes containing the same disease allele in controls (i.e. the disease allele frequency in controls), and the increment or decrement of disease allele frequencies in the cases. For 'pd', the minor allele frequency of the disease variant, 'RR', the relative risk of the disease allele, and 'OR', the odds ratio of the disease allele, the haplotype frequency in cases containing the disease allele can be derived using equation (2).

$$\begin{aligned} pH_{case} &= pH_{control} + pd \times (RR-1) \times \frac{pH_{control}}{pd} \\ &= pH_{control} + pd \times \left(\frac{OR}{1-pd+pd \times OR} - 1 \right) \times \frac{pH_{control}}{pd} \end{aligned} \quad (2)$$

As a result, the frequencies and LD are changed in Table 1 and 2, even with low LD between markers and the disease variant in controls. To examine the changes depending on the odds ratio of the disease allele, minor allele frequencies and LD are plotted in Fig. 2 and 3. Using the initial haplotype state set out in Table1, the minor allele of marker A is increased and the minor allele of marker B is marginally decreased as the odds ratio increases. Since the LD coefficient, r^2, indicates the correlation among variants, there is a natural decrease of LD between markers and an increase of LD between the marker and the disease allele, as shown in Fig. 3. The behavior of D' is different from the behavior of r^2. As the odds ratio is increased, the D' between markers decreases in a similar manner to r^2. However, as shown in Fig. 2 and 3, the D' between a marker and the disease variant becomes minimal at the point where the increment of the minor allele of the disease variant exceeds the marker's minor allele frequency, after which it increases.

When examining the complete range of haplotype frequencies (indicated in equation 1), the r^2 between a marker and the disease variant is increased in every haplotype frequency of AB as the odds ratio increases. However, the D' becomes minimal until the increment of the minor allele of the disease variant exceeds the marker's minor allele in a given case, and thereafter increases, when the AB haplotype frequencies are higher than the multiplication between the marker major allele frequency and the disease major allele frequency. Otherwise, D' is decreased depending on the increment of odds ratios. Using the linkage disequilibrium changes between cases and controls, it is possible to elucidate whether the observed variants are real disease variants or just markers of which frequencies are changed solely due to real disease variants.

Testing Hypothesis

In many association studies, there are several variants in a locus that show associations, possibly through LD, with the disease variants. If the functional variants are not obvious, it is difficult to distinguish the true association from the linked markers. Here, a way of testing the independence in association from the known hypothetical disease variants is proposed, using the LD properties described earlier. The null hypothesis is that the increased or decreased marker allele frequencies arise solely from the LD relationship with the increased disease allele. The haplotype frequencies of the testing marker and disease variant(s) are estimated using the expectation maximization (EM) algorithm from the control genotype data, and applied to calculate the expected haplotype frequencies, considering LD with the increased allele frequency of the disease variant. This is related to the increment or decrement of haplotypes corresponding to the proportion of the haplotype frequency harbored in the risky (or protective) disease allele. In cases with more than one disease variant, the disease variant can be replaced by the disease haplotypes consisting of the disease variants, which can be considered as a multi-allelic disease variant. From known disease haplotypes, D ($d_{1,\ldots,n}$), consisting of disease variants with known estimated relative risks of haplotypes and known LD status between disease variants and the marker, the expected allele frequency of the testing variant is expressed as equation (3). Using the expected frequencies, it is possible to test the independence of association of the variant from the assumed disease variant.

$$E(pA_{case}|D,LD) = pA_{control} + \sum_{i=1}^{n} pd_i \times (RR_i - 1) \times \frac{pAd_i}{pd_i} \qquad (3)$$

Statistical Method

To test if the association of markers is independent from the known disease variants, the expected allele frequency in cases from equation (3) can be compared with the real allele frequency in the respective case. For estimating the control haplotype frequencies, the EM algorithm was applied. As indicated in the previous section, the null hypothesis is that the changed allele frequencies of the marker are solely due to the LD with the disease variant(s), which is estimated using equation (3), and the alternative is the distribution of real cases. The Pearson chi-square test was performed with the Monte Carlo method for the computation of simulated p-value using R 2.2.1.

Sample Descriptions

Using previously published data (Farrer *et al*., 1997; Kamboh *et al*., 1999; Nicodemus *et al*., 2004), the independence in association of all seven variants studied was tested for counting the LD with rs429358, which encodes ε4. In the case of the ε2 allele, coded by rs7412, high D' with the ε4 allele, rs429358, causes the haplotype frequency containing the minor allele of rs7412 to naturally decrease due to the increment of the minor allele of rs429358. In the meta-analysis study data (Farrer *et al*., 1997), the minor allele of rs7412 shows a minor protective effect. To verify that the decrement of the minor allele of rs7412 is due to the LD with rs429358, pooled Caucasian samples (Farrer *et al*., 1997) were used. For rs769452, the only study reported (Kamboh *et al*., 1999) was tested. For the remaining variants, data from Nicodemus *et al*. (2004) were applied to the test. The odds ratio of ε4 used in this test is 3.65218, calculated from the Caucasian data in the previous study, and used for the tests, since this provided the largest sample among the selected publications (Farrer *et al*., 1997).

Results

The described method of controlling LD in association tests was applied to the actual association results of *APOE* variants with Alzheimer's disease. Considering inconsistent associations in complex traits, it is noteworthy that ApoE ε4 association in Alzheimer's disease has been replicated in most studies, especially in Caucasians, indicating that a reasonably large effect would not be overlooked in an association study. Although this disease allele in *APOE* merely contributes to the partial genetic etiology of Alzheimer's disease, the odds ratio of homozygotes is 14.9 as derived by meta-analysis, and the allelic effect is approximately multiplicative (Farrer *et al*., 1997). However, there are other functional variants in *APOE* that may contribute to Alzheimer's disease such as other non-synonymous single nucleotide polymorphisms (SNPs) and transcriptional variants. Previously, the genetic effects of potential transcriptional variants were tested taking into consideration the linkage disequilibrium with rs429358 coding ε4 using regression with adjustment and stratification depending on their genotypes. However, only rs1081106 (+5361) showed marginal significance among the five SNPs tested (Nicodemus *et al*., 2004). As indicated previously, with the known relationship between the testing markers and the disease variant, the conventional statistical tests used in this study may represent less valid results for counting the LD appropriately. Therefore, the genetic effect of the studied *APOE* variants was re-evaluated using the method for counting the LD suggested here.

Regarding the variants in *APOE*, a total of eight variants

Table 3. Testing independence of associations from ApoE ε4

Variants	Control MAF	Expected MAF	Case MAF	major allele count	minor allele count	Individual Associations	Testing Independence
rs449647 (-491)	0.191	0.154	0.122	960	134	0.0001	0.0045
rs769446 (-427)	0.101	0.065	0.068	1020	74	0.0004	0.7622
rs405509 (-219)	0.476	0.561	0.579	461	633	0.0001	0.2505
rs440446 (+113)	0.362	0.325	0.239	832	262	0.0001	0.0001
rs769452 (ε4*P)	0.0018	0.005	0.024	1091	27	0.0001	0.0001
rs429358 (ε4)	0.137	-	0.367	6465	3749	0.0002*	-
rs7412 (ε2)	0.084	0.062	0.039	9816	398	0.0002*	0.0002*
rs1081106 (+5361)	0.099	0.062	0.048	1042	52	0.0001	0.0504

Simulated p-value using 10000 replicates; *:simulated p-value using 5000 replicates; MAF: minor allele frequency.

near the gene have been studied for their associations with Alzheimer's disease, but the results were inconsistent in many cases except in the case of ε4. In this gene region, there are many rare non-synonymous variants with minor allele frequency (MAF) < 0.01, which are associated with certain diseases, and three, ε2, 3, and 4, are known to be common protein variants (de Knijff *et al*., 1994). The common protein variants are defined by two sequence variants, rs429358 (coding ε4 T/C C130R carrier%= 6.4~36.8) and rs7412 (coding ε2, C/T R176C carrier%= 0~14.5), which are known to be associated with Alzheimer's disease. The haplotype TC defines ε3, CC defines ε4, and TT defines ε2. Among the rare variants, one variant, Leu28Pro (ε4*P), which is always observed with Cys130Arg coded by rs429358, has been shown to be associated with AD (heterozygosity in controls = .18%; heterozygosity in cases=2.42%) (Kamboh *et al*., 1999).

As indicated in Table 3, several variants have been shown to be associated with Alzheimer's disease independent from the *APOE* ε4 variant, rs429358, regardless of LD, and showing significant simulated p-values. These tests show that the associations of rs449647, rs440446, rs769452, and rs7412 are independent from rs429358. However, the results merely represent the independence in association from rs429358, and do not confirm the actual association of each variant. Therefore, in order to find the real effect of each variant, further studies are necessary. Due to the unavailability of data, testing for multiple putative disease variants, particularly using non-synonymous variants, could not be tested further.

Discussion

Using the basic property of the change in linkage disequilibrium between cases and controls depending on the real disease variant, we have proposed a new test for the independence of association for counting LD. The independence in association from several disease haplotypes can be tested for the existence of several disease variants using equation (3). However, it should be noted that this test does not provide a concrete result for the number of disease variants or their actual odds ratio. As seen in the APOE association result, if there are several common disease variants in a locus, the consideration of number of disease variants as well as their LD relationship is critical in detecting associations. By adding the probable disease variants until finding the best fitting model, the model with real disease variants can be found with the highest likelihood. The actual association of those disease variants may be confirmed by likelihood ratio tests using haplotypes consisting of the disease variants. Further study on relevant issues would be necessary.

From the test example, several *APOE* variants turned out to be associated with Alzheimer's disease independent from the association of the *APOE* ε4 variant, rs429358. For obtaining the allelic odds ratio of rs429358 and testing rs7412, the pooled Caucasian samples in each study may cause inaccuracy in the results possibly due to the heterogeneity. However, the results counting heterogeneity are not expected to be critically different from the results in here. These tests indicate the existence of more disease variants than rs429358 in *APOE*, which has been doubtful for associations with Alzheimer's disease due to their LD with *APOE* rs429358. Analyses with full genotype data would give a better idea regarding the real disease variants. However, it could not be concluded without further replications of these results and confirmation by functional evidence.

Regarding the transcriptional activity of the *APOE* variants, among the five putative regulatory SNPs near *APOE* studied for their association with Alzheimer's disease, the -491 variant, rs449647, has been studied most extensively. Transcriptional activity of this variant was verified for haplotypes with another SNP, rs769446, using luciferase-based transient cell transfection of astrocytoma cells (Artiga *et al.*, 1998; Bullido *et al.*, 1998). Allele A of rs449647 increases the transcriptional activity *in vitro* (Artiga *et al.*, 1998; Bullido *et al.*, 1998) and leads to an increase in brain ApoE level in humans (Laws *et al*., 2002), which is consistent with the hypothesis that an increased ApoE level increases the risk of Alzheimer's disease. The combined analysis using both rs449647 and rs769446 results in significantly altered promoter activity among the haplotypes, TT, AT, and AC. The AC haplotype, consisting of the A allele of rs449647 and the C allele of rs769446, shows the highest promoter activity, and the TT haplotype shows the least promoter activity. In the meta-analysis of Alzforum (http://www.alzforum.org/res/com/gen/alzgene/meta.asp?geneID=85), the minor alleles, T allele of rs449647 and C allele of rs769446 are the protective alleles. The protective effect of the C allele of rs769446 is opposite from expectations considering the increased transcriptional activity, possibly due to the LD with rs429358. The risk associated with the AC haplotype is the highest compared to the other two haplotypes among non-ApoE ε4 subpopulation, supporting the notion that the transcriptional effect of rs769446 may still be a weak causal factor. Moreover, a recent study on regulatory elements of the *APOE* promoter showed clear cell specificity of these elements, in which rs405509 showed differential transcriptional activity depending on the associated allele and cell type (Ramos *et al*., 2005). Although the functionality of rs449647 matches the causality even with the result of significant association after the LD correction, sometimes, the functional evidence may not correlate with the diseases as seen in the negative associations of both rs769446 and rs405509 in Table3. Otherwise, their weak causal effect may be hidden in this test due to the LD with disease variants other than rs429358.

As stated previously, among the equivocal association studies, *APOE* ε4 association is noteworthy for its robust replications especially in Caucasians. This robustness may come from rs429358 as well as other disease variants in this gene, and/or from the unusual LD pattern due to the selection pressure on rs429358, which can be supported by the sequence comparison with chimpanzees and the slightly increased LD among common variants near rs429358 from the LD pattern using the re-sequenced data (http://droog.gs.washington.edu/ mdecode/data/apoe/) (Nickerson *et al*., 2000). In either case, the association studies need to be conducted more carefully with consideration to the LD properties described in this study.

This study provides the possible basis for a better interpretation of haplotype associations in a region. Moreover, the LD property studied here may assist improved tagging and power calculation in relation to several common disease variants by providing the associated relationship between the disease variants and markers in association tests. Depending on the described LD property, the association can be detected even with the low LD if the markers are close enough. This knowledge could be applied to design an efficient study for finding new disease variants. The LD property described here could also be used together with the publicly available LD data for the elucidation of further disease variants.

Acknowledgements

The initiation of this work was supported by Dr. Steven G. Younkin. Thank for his passion in studying the genetic etiology of Alzheimer's disease which inspired this work. I also appreciate the kindness of Dr. Young J. Kim for processing the fast publication of this paper.

References

Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J., and Donnelly P. (2005). A haplotype map of the human genome. *Nature* 437, 1299-1320.

Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* 3, 299-309.

Artiga, M. J., Bullido, M. J., Frank, A., Sastre, I., Recuero, M., Garcia, M. A., Lendon, C. L., Han, S. W., Morris, J. C., Vazquez, J., Goate, A., and Valdivieso, F. (1998). Risk for Alzheimer's disease correlates with transcriptional activity of the APOE gene. *Hum. Mol. Genet.* 7, 1887-1892.

Bullido, M. J., Artiga, M. J., Recuero, M., Sastre, I., Garcia, M. A., Aldudo, J., Lendon, C., Han, S. W., Morris, J. C., Frank, A., Vazquez, J., Goate, A., and Valdivieso, F. (1998). A polymorphism in the regulatory region of APOE associated with risk for Alzheimer's dementia. *Nat. Genet.* 18, 69-71.

Chapman, J. M., Cooper, J. D., Todd, J. A., and Clayton, D. G. (2003). Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum. Hered.* 56, 18-31.

de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nat. Genet.* 37, 1217-1223.

genetic association studies. *Nat. Genet.* 37, 1217-1223.

de Knijff, P., van den Maagdenberg, A. M., Frants, R. R., and Havekes, L. M. (1994). Genetic heterogeneity of apolipoprotein E and its influence on plasma lipid and lipoprotein levels. *Hum. Mutat.* 4, 178-194.

De La Vega, F. M., Gordon, D., Su, X., Scafe, C., Isaac, H., Gilbert, D. A., and Spier, E. G. (2005). Power and sample size calculations for genetic case/control studies using gene-centric SNP maps: application to human chromosomes 6, 21, and 22 in three populations. *Hum. Hered.* 60, 43-60.

Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., Myers, R. H., Pericak-Vance, M. A., Risch, N., and van Duijn, C. M. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Jama.* 278, 1349-1356.

Gordon, D., Finch, S. J., Nothnagel, M., and Ott, J. (2002). Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum. Hered.* 54, 22-33.

Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di Genova, G., Ueda, H., Cordell, H. J., Eaves, I. A., Dudbridge, F., Twells, R. C., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S. C., Clayton, D. G., and Todd, J. A. (2001). Haplotype tagging for the identification of common disease genes. *Nat. Genet.* 29, 233-237.

Kamboh, M. I., Aston, C. E., Perez-Tur, J., Kokmen, E., Ferrell, R. E., Hardy, J., and DeKosky, S. T. (1999). A novel mutation in the apolipoprotein E gene (APOE*4 Pittsburgh) is associated with the risk of late-onset Alzheimer's disease. *Neurosci. Lett.* 263, 129-132.

Laws, S. M., Hone, E., Taddei, K., Harper, C., Dean, B., McClean, C., Masters, C., Lautenschlager, N., Gandy, S. E., and Martins, R. N. (2002). Variation at the APOE -491 promoter locus is associated with altered brain levels of apolipoprotein E. *Mol. Psychiatry.* 7, 886-890.

Nickerson, D. A., Taylor, S. L., Fullerton, S. M., Weiss, K. M., Clark, A. G., Stengard, J. H., Salomaa, V., Boerwinkle, E., and Sing, C. F. (2000). Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. *Genome Res.* 10, 1532-1545.

Nicodemus, K. K., Stenger, J. E., Schmechel, D. E., Welsh-Bohmer, K. A., Saunders, A. M., Roses, A. D., Gilbert, J. R., Vance, J. M., Haines, J. L., Pericak-Vance, M. A., and Martin, E. R. (2004). Comprehensive association analysis of APOE regulatory region polymorphisms in Alzheimer disease. *Neurogenetics* 5, 201-208.

Ramos, M. C., Matias, S., Artiga, M. J., Pozueta, J., Sastre, I., Valdivieso, F., and Bullido, M. J. (2005). Neuronal specific regulatory elements in apolipoprotein E gene proximal promoter. *Neuroreport* 16, 1027-1030.

Weiss, K. M. and Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex human traits. *Trends Genet.* 18, 19-24.