Genetic epidemiological approaches in complex neurological disorders

Hottenga, J.J.

CHAPTER 9

General discussion

JJ Hottenga

In this section the various methods that have been applied in this thesis for analyzing the correlation between genotype and phenotype in complex neurological disorders will be discussed together with some possible future perspectives.

Association & stratification

For complex neurological disorders association studies are a straightforward method to quickly assess the involvement of candidate genes. Furthermore, the design is powerful for testing polymorphisms with small effects, which is essential for studying the genetics of complex disorders[1]. However, association studies are often criticized because of the failure of replicating positive results[2]. In literature, the advantages, as well as the inherent problems of the study design have been thoroughly discussed[2-4]. One of the major concerns for association studies has been population stratification[1,5]. Spurious association due to population stratification occurs when both the disorder and gene frequency differs between two populations. If cases and controls are selected in different proportions from these two populations, association will occur without a causal relation between the tested polymorphism and the disorder.

In *chapter two* it was shown that population stratification surprisingly has a small effect, except for extreme situations; the unlikely situation that the gene frequency and/or the selection of two populations is extremely different in cases and controls. Although this may lead to the conclusion that one does not need to take population stratification into account, there are reasons why such stratification should be avoided. Population stratification may be relevant in case of studying very large samples of cases and controls, or searching for gene variants with limited effect size[6-8]. Only a small number of empirical studies is currently available, concerning the presence and use of testing population stratification. In addition, studies report contradictory results based on what is considered to be a substantial increase in false-positive findings[9-11]. To use an analogy, if one would perform a case-control study involving lung-

cancer and a given risk factor, confounding factors like smoking, gender and age should be taken into account when designing such a study. The confounding of population stratification can taken into account as well, as several tests and correction methods have been proposed[12-17].

If population stratification is not a major issue, the question remains why association studies are often false-positives and cannot easily be replicated. There is good reason to argue that major factors are the sample size of the study population, and statistical problems due to multiple testing. As shown indirectly in *chapters three* and *four*, sample size can largely influence the outcome of studies. Unfortunately, many studies suffer from too small sample sizes to evaluate genuine associations (and gene-gene interactions)[18]. Therefore, increasing the sample size (to several hundreds or even thousands), and increasing the statistical significance level (well below $\alpha = 0.05$) seem logical remedies to reduce false-positive association results. Replication studies, meta-analyses and more in-depth functional research should be applied to confirm initial association findings. Recently, it has been proposed that data for association studies should be made available online[19]. In this way other researchers can add data and check their results in combined data sets. Such an approach would make more effective use of resources and would help in avoiding publication bias towards false-positive findings.

The use of parametric linkage analysis in complex neurological disorders

Although parametric linkage analysis is often considered to be less efficient than non-parametric methods for localizing genes in complex traits, the results of this thesis show that it can be a useful approach given that efforts to study homogenous material are taken. In *chapters seven* and *eight* parametric linkage analysis was applied to study single Mendelian families affected with epilepsy. In *chapter seven* a Dutch family with familial cortical tremor and

epilepsy (FCTE) showed no evidence for linkage to the Japanese 8q23.3-24.1 locus, which indicated heterogeneity for this phenotype[20,21]. In *chapter eight* linkage analysis was performed in a single family for several loci involved in autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) (1p21, 15q24, 20q.13.3) and familial partial epilepsy with variable foci (FPEVF) (2q36, 22q11-q12)[22-26]. Linkage to chromosome 22q11-q12 favored the diagnosis of FPEVF, showing that linkage analysis can be used to support a diagnosis of rare familial epilepsy syndromes.

In addition to the epilepsy families, parametric linkage analysis was applied in seven Dutch migraine without aura (MO) families (*chapter six*). In order to increase the homogeneity, the MO families were selected based on the criterion that (nearly) all affected individuals should have MO. Branches in which a spouse was affected with migraine were not used. Despite these efforts, none of the individual families showed significant - or suggestive evidence for linkage. LOD scores were substantially lower than the expected simulated LOD scores, assuming a single segregating disorder gene per family. Allelic heterogeneity and/or presence of phenocopies were found, as a number of affected individuals in the large families did not carry the specific haplotype segregating with the disorder. Locus heterogeneity seemed to be present as well but could not be confirmed. Given the complexity and prevalence of migraine, the heterogeneity is not surprising. Interestingly, suggestive evidence for linkage was found at a locus on chromosome 4q21-q24, replicating the results of the Iceland and Finnish genome scans[27,28].

Family selection and effects of heterogeneity in the MO linkage analysis

In *chapter six* several strategies were employed to account for the heterogeneity that is obviously a large problem in the genetics of migraine. The selection of specific phenotypes to reduce heterogeneity has often been a

successful approach for mapping genes, for example in early-onset cases of Alzheimer's disease and familial hemiplegic migraine[29,30]. However, the method has an effect only if the selection criterion contributes to a more homogenous sample. The possibility that multiple risk factors segregate in the studied families remains, which probably occurred in *chapter six* when selecting the families with MO. The selection of these highly loaded families may even have contributed to the heterogeneity because it becomes more probable to select families, in which multiple disorder genes segregate. Some evidence for this was found in the segregation analyses of *chapter five*, where the polygenic model fitted as good as the general dominant single locus model.

In contrast to the epilepsy families, the migraine families did not show strong evidence for linkage when analyzed individually. Two families were large enough to detect significant linkage. Ironically, the family size also increases the probability of heterogeneity, as the married-in spouses may contribute new risk factors. This was avoided as much as possible by excluding branches with two affected parents but probably combinations of risk factors still contributed to the migraine development in the remaining branches. An alternative approach is to analyze a number of (smaller) families in a single analysis but it will likely increase the probability of locus heterogeneity. In *chapter six* this approach was employed as well without a substantial change in conclusions. Analyzing a single large family or multiple smaller families remains a dilemma. In the field of migraine the use of both methods has led to positive results recently[27,31]. Given a substantial publication bias, it is difficult to determine, which method is optimal. Theoretically, there is only a limited number of genes involved in a disorder, as a limited number of biochemical pathways are affected. Therefore, studying a very large sample of smaller families should eventually have more power to detect the responsible gene variant(s), as compared to a single family approach.

Another frequently used approach to account for heterogeneity is a test with an iterated mixture parameter added to the statistic, which determines the probability of a given family to be linked to the locus[32-34]. Also this approach was applied in *chapter six,* using the program HOMOG[33,34]. Although this method is increasing the power of locus detection under heterogeneity, it is far from perfect. For example, the linkage model parameters are assumed to be equal for the loci. Also, the estimations of the mixture parameter can be wrong, depending on the parametric linkage model used and the number of families that is tested. Furthermore, the likelihood statistic has a difficult distribution with one or two degrees of freedom[35,36]. The mixture method has been extended for the non-parametric analysis, dealing with some of the problems that are inherent to the use of a linkage model[37]. Whether parametric or non-parametric linkage analysis is the best approach to detect linkage is an issue of discussion[38,39].

Alternative approaches accounting for heterogeneity

To test linkage in migraine without aura, alternative strategies could have been employed to reduce heterogeneity as well. These include study design changes like testing association or sib-pair analysis instead of (parametric) linkage[1]. Another way to cope with heterogeneity using a linkage approach is to divide larger families into smaller nuclear families and analyze them as being independent, using sib-pair analysis or non-parametric methods[39]. When the mode of inheritance is specified as dominant for parametric linkage, while the true mode of inheritance is recessive, this method will increase the detection probability of recessive loci[27,40,41]. In case the mode of inheritance was specified correctly, some power is probably lost because pedigrees have been split into nuclear families[42]. Some genome scans analyzing the data with both methods show that the LOD scores and detected locations are often very similar[27,43,44].

In addition to increasing the homogeneity of the phenotype, the homogeneity of the whole genome in a studied sample can be increased as well. Families can be selected from an isolated population, in which it is assumed that genetic drift, disease bottlenecks and founder effects have reduced the heterogeneity of the genetic risk factors[45]. Preferably, the genealogy of the population is known as well, so that selected families or persons can be related to each other[45-47]. Currently, a number of isolate studies applying different design- and statistical approaches have been published with optimistic results[47-49]. It should be noted, however, that the heterogeneity may not be reduced for some disorders of interest. In addition, the found loci may be unimportant risk factors in other populations.

Nowadays, research should be aimed at developing more specific methods to detect linkage under heterogeneity. Correcting for linkage evidence at other loci may be such an option and various methods to employ this strategy have been developed[50-52]. The use of ordered subset analysis, in which families are rank-ordered based on a covariate (phenotype) and then permuted until the maximum LOD score of a given subset is found, may be extended for heterogeneity as well[53].

The selection of a proper family-based association test

With the current availability of dense single nucleotide polymorphisms maps and the increased number of mapped susceptibility loci, the emphasis of future genetic research for complex diseases will more often be focused on fine mapping genes with family-based association studies. A simple question, though sometimes difficult to answer, is how to select the proper statistical method used to analyze the data. Of course, a limitation is the study sample characteristics but the possibilities are large when a sample of sib-pairs for a dichotomous trait with unaffected, affected siblings and parents has been genotyped. If parent-affected child trios are selected from these data, then the haplotype relative risk (HRR) method, transmission disequilibrium test (TDT),

reconstruction combined-transmission disequilibrium test (RC-TDT) or TRANSMIT test can be applied[54-57]. In addition, tests like the sib transmission disequilibrium test (S-TDT), the discordant alleles test (DAT) and discordant sibship test (SDT) that make use of the known allele sharing / transmission in siblings can be used as well[58-60]. Furthermore, complete families can be analyzed, using a weighting function for the familial relations in tests like the pedigree disequilibrium test (PDT) and family based association test (FBAT)[61,62]. These examples were taken from a much larger list of family-based association tests that was thoroughly reviewed in Schulze and McMahon in 2002[63]. Since then, the number of tests has still increased[64,65]. Finally, other possibilities can be taken under consideration, such as the use of covariates in the analysis, the use of multiple markers or (tag) haplotypes and the use of quantitative traits[66-68].

In *chapter four* three statistics (TDT, Mantel-Haenszel extension, Z'score TDT/S-TDT) were applied to study if there was an association between the HVR haplotype, causing retinopathy, and co-morbid migraine and Raynaud phenomenon in a Dutch family[55,58.,69]. The data were generated using only one large pedigree, which caused problems in several of the proposed tests (PDT and FBAT) because the statistics are based on the weighting of multiple families[61,62]. However, splitting the family into multiple nuclear families resolved this issue, and some family-based statistics could be employed. Differences that were observed between the use of trios and sib-pairs may be explained by the increase of sample size using sibling-based approaches. With the different approaches, changing sub-samples from a single family are studied, therefore, the results may differ in outcome based on the selection. If the relation between the genotype and phenotype is strong and sample size is relatively large, these effects will probably not alter the outcome. In smaller samples, however, this might not be true. It is therefore important to know the properties of a given test. Here, literature becomes less extensive: many tests are developed but are tested only for a limited number of situations. In

addition, a limited number of tests is compared and mainly for the power to detect association. Experience in implementation and support for many of these tests is difficult to obtain, which may lead to a reduced use of the most optimal method. Future research should be aimed at more careful comparison of tests with empirical and simulated data. In addition, a more user-friendly program combining multiple tests, like SPSS for example, would likely be helpful for many (epidemiological) geneticists.

Conclusion

Different complex neurological disorders require different mapping strategies and study design to successfully locate genes involved in the disorder. Linkage analysis and association analysis are applied to contribute to these findings. In this thesis the results were often dependent on a selection either in samples or statistics used for the analysis. In case-control studies the selection of cases and controls may sometimes lead to confounding. In linkage analysis the family selection and method of analysis can be the difference between failure and success of a study. Essential is to know the effects and limitations within a study, even more when other possibilities of research are limited, which is often the case in complex neurological disorders.

References

1. Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994;265:2037-48.
2. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet. 2003;8;361:865-72.
3. Little J, Bradley L, Bray MS et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol. 2002;15;156:300-10.
4. Ioannidis JP. Genetic associations: false or true? Trends Mol Med. 2003;9:135-8.
5. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet. 2003;15;361:598-604.
6. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nat Genet. 2004;36:512-7.
7. Freedman ML, Reich D, Penney KL et al. Assessing the impact of population stratification on genetic association studies. Nat Genet. 2004;36:388-93.
8. Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol. 2004;27:14-20.

9. Koller DL, Peacock M, Lai D et al. False positive rates in association studies as a function of degree of stratification. J Bone Miner Res. 2004;19:1291-5.
10. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev. 2002;11:505-12.
11. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev. 2002;11:513-20.
12. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55:997-1004.
13. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet. 1999;65:220-8.
14. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945-59.
15. Satten GA, Flanders WD, Yang Q. Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. Am J Hum Genet. 2001;68:466-77.
16. Ripatti S, Pitkaniemi J, Sillanpaa MJ. Joint modeling of genetic association and population stratification using latent-class models. Genet Epidemiol. 2001; 21 Suppl 1:S409-14.
17. Pfaff CL. Adjusting for population structure in admixed populations. Genetic Epidemiology. 2002;1. 22:196–201.
18. Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. Lancet. 2003;15;361:567-71.
19. Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR; International Meta-analysis of HIV Host Genetics. Commentary: meta-analysis of individual participants' data in genetic epidemiology. Am J Epidemiol. 2002;156:204-10.
20. Plaster NM, Uyama E, Uchino M et al. Genetic localization of the familial adult myoclonic epilepsy (FAME) gene to chromosome 8q24. Neurology. 1999;53:1180-3.
21. Mikami M, Yasuda T, Terao A et al. Localization of a gene for benign adult familial myoclonic epilepsy to chromosome 8q23.3-q24.1. Am J Hum Genet. 1999;65:745-51.
22. De Fusco M, Becchetti A, Patrignani A et al. The nicotinic receptor beta 2 subunit is mutant in nocturnal frontal lobe epilepsy. Nat Genet. 2000 Nov;26:275-6.
23. Scheffer IE, Phillips HA, O'Brien CE et al. Familial partial epilepsy with variable foci: a new partial epilepsy syndrome with suggestion of linkage to chromosome 2. Ann Neurol. 1998;44:890-9.
24. Phillips HA, Scheffer IE, Crossland KM et al. Autosomal dominant nocturnal frontal-lobe epilepsy: genetic heterogeneity and evidence for a second locus at 15q24. Am J Hum Genet. 1998;63:1108-16.
25. Phillips HA, Scheffer IE, Berkovic SF et al. Localization of a gene for autosomal dominant nocturnal frontal lobe epilepsy to chromosome 20q13.2. Nat Genet. 1995;10:117-8.
26. Xiong L, Labuda M, Li DS et al. Mapping of a gene determining familial partial epilepsy with variable foci to chromosome 22q11-q12. Am J Hum Genet. 1999;65:1698-710.
27. Wessman M, Kallela M, Kaunisto MA et al. A susceptibility locus for migraine with aura, on chromosome 4q24. Am J Hum Genet. 2002;70:652-62.
28. Bjornsson A, Gudmundsson G, Gudfinnsson E et al. Localization of a gene for migraine without aura to chromosome 4q21. Am J Hum Genet. 2003;73:986-93.

29. Van Broeckhoven C, Backhovens H, Cruts M et al. Mapping of a gene predisposing to early-onset Alzheimer's disease to chromosome 14q24.3. Nat Genet. 1992;2:335-9.
30. Joutel A, Bousser MG, Biousse V et al. A gene for familial hemiplegic migraine maps to chromosome 19. Nat Genet. 1993;5:40-5.
31. Soragna D, Vettori A, Carraro G et al. A locus for migraine without aura maps on chromosome 14q21.2-q22.3. Am J Hum Genet. 2003;72:161-7.
32. Smith CAB. Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet. 1963;27:175–182.
33. Ott J. Linkage analysis and family classification under heterogeneity. Ann Hum Genet. 1983;47:311–20.
34. Bhat A, Heath SC, Ott J. Heterogeneity for multiple disease loci in linkage analysis. Hum Hered. 1999;49:229-31.
35. Whittemore AS, Halpern J. Problems in the definition, interpretation, and evaluation of genetic heterogeneity. Am J Hum Genet. 2001;68:457-65.
36. Hodge SE, Vieland VJ, Greenberg DA. HLODs remain powerful tools for detection of linkage in the presence of genetic heterogeneity. Am J Hum Genet. 2002;70:556-9.
37. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996;58:1347–63.
38. Abreu PC, Greenberg DA, Hodge SE. Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. Am J Hum Genet. 1999;65:847-57.
39. Goring HH, Terwilliger JD. Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. Am J Hum Genet. 2000;66:1310-27.
40. Knapp M, Seuchter SA, Baur MP. Linkage analysis in nuclear families. 1: Optimality criteria for affected sib-pair tests. Hum Hered. 1994;44:37-43.
41. Knapp M, Seuchter SA, Baur MP. Linkage analysis in nuclear families. 2: Relationship between affected sib-pair tests and lod score analysis. Hum Hered. 1994;44:44-51.
42. Huang J, Vieland VJ. Comparison of 'model-free' and 'model-based' linkage statistics in the presence of locus heterogeneity: single data set and multiple data set applications. Hum Hered. 2001;51:217-25.
43. Das SK, Hasstedt SJ, Zhang Z, Elbein SC. Linkage and association mapping of a chromosome 1q21-q24 type 2 diabetes susceptibility locus in northern European Caucasians. Diabetes. 2004;53:492-9.
44. Angius A, Petretto E, Maestrale GB et al. A new essential hypertension susceptibility locus on chromosome 2p24-p25, detected by genomewide search. Am J Hum Genet. 2002;71:893-905.
45. Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet. 2000;1:182-90.
46. Enserink M. Physicians wary of scheme to pool Icelanders' genetic data. Science. 1998;281:890-891.
47. Vaessen N, Heutink P, Houwing-Duistermaat JJ, et al. A genome-wide search for linkage-disequilibrium with type 1 diabetes in a recent genetically isolated population from the Netherlands. Diabetes. 2002;51:856-9.
48. Freimer NB, Reus VI, Escamilla MA et al. Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet. 1996;12:436-41.

49. Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH et al. Localization of a susceptibility gene for common forms of stroke to 5q12. Am J Hum Genet. 2002;70:593-603.
50. Schaid DJ, McDonnell SK, Thibodeau SN. Regression models for linkage heterogeneity applied to familial prostrate cancer. Am J Hum Genet. 2001;68:1189–96.
51. Cordell HJ, Jacobs KB, Wedig GC, Elston RC. Improving the power for disease locus detection in affected-sib-pair studies by using two-locus analysis and multiple regression methods. Genet Epidemiol. 1999;17 Suppl 1:S521-6.
52. Vieland VJ, Wang K, Huang J. Power to detect linkage based on multiple sets of data in the presence of locus heterogeneity: comparative evaluation of model-based linkage methods for affected sib pair data. Hum Hered. 2001;51:199–208.
53. Schmidt S, Scott WK, Postel EA et al. Ordered subset linkage analysis supports a susceptibility locus for age-related macular degeneration on chromosome 16p12. BMC Genet. 2004;5:18.
54. Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet. 1987;51:227-33.
55. 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;52:506-16.
56. Knapp M. Reconstructing parental genotypes when testing for linkage in the presence of association. Theor Popul Biol. 2001;60:141-8.
57. Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet. 1999;65:1170-7.
58. Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet. 1998;62:450-8.
59. Boehnke M, Langefeld CD. Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet. 1998;62:950-61.
60. Horvath S, Laird NM. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet. 1998;63:1886-97.
61. 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;67:146-54.
62. 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;50:211-23.
63. Schulze TG, McMahon FJ. Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. Am J Med Genet. 2002;114:1-11.
64. Lange C, Silverman EK, Xu X et al. A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics. 2003;4:195-206.
65. Qian D. Haplotype sharing correlation analysis using family data: a comparison with family-based association test in the presence of allelic heterogeneity. Genet Epidemiol. 2004;27:43-52.
66. Horvath S, Wei E, Xu X et al. Family-based association test method: age of onset traits and covariates. Genet Epidemiol. 2001;21 Suppl 1:S403-8.
67. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. Am J Hum Genet. 2000;66:279-92.
68. Clayton D, Jones H. Transmission/disequilibrium tests for extended marker haplotypes. Am J Hum Genet. 1999;65:1161-9.
69. Laird NM, Blacker D, Wilcox M. The sib transmission/disequilibrium test is a Mantel-Haenszel test. Am J Hum Genet 1998;62:450-8.