Ann. Hum. Genet. (1999), **63**, 351–367
Printed in Great Britain

European Mathematical Genetics Meeting

held at Loughborough, 9–11 April 1999

ORGANISED BY NUALA A SHEEHAN AT THE DEPARTMENT OF MATHEMATICAL SCIENCES, LOUGHBOROUGH UNIVERSITY

Abstracts

Risk Models for Familial Breast and Ovarian Cancer. A. C. ANTONIOU[1], D. F. EASTON[1], S. A. GAYTHER[2], J. F. STRATTON[2] and B. J. PONDER[2]. [1] *CRC Genetic Epidemiology Unit, Institute of Public Health and* [2] *CRC Human Cancer Genetics Research Group, University of Cambridge, UK.*

We investigated risk models for the inherited susceptibility of breast and ovarian cancer, using data from both high-risk families and a population based series of ovarian cancer. The first data set consisted of 112 families containing 2 or more relatives with epithelial ovarian cancer. BRCA1 and BRCA2 germline mutations were detected in 50% of these families. The second study involved 374 ovarian cancer cases, collected at the Royal Marsden Hospital, London, who had DNA samples analysed for BRCA1 mutations. 12 women were found to be carriers. We constructed genetic models for ovarian and breast cancer using the computer program MENDEL. In the first study we modelled the effects of BRCA1 and BRCA2 simultaneously and allowed for a third gene predisposing to ovarian cancer. None of the models fitted gave significant evidence for a third gene. Population frequencies of BRCA1 and BRCA2 mutations were estimated to be 0.13% and 0.17% respectively. Our results suggest that BRCA1 and BRCA2 may be sufficient to explain the majority of familial ovarian cancer and that families without mutations can be explained by sensitivity of the mutation testing and chance clusters of sporadic cases. Using data on the families of the 12 mutation carriers in the second study, we estimated age specific ovarian and breast cancer risks for BRCA1 mutation carriers. Under the best fitting model the cumulative ovarian cancer risk was 66% by age 70, and the corresponding breast cancer risk was 45%. The differences in penetrance estimates among studies suggest that modifying genetic or environmental factors may be important determinants of risk.

An example of complex segregation analysis of plant pedigree: reversion of cytoplasm type in Sugar Beet (*Beta vulgaris* L.) YU. S. AULCHENKO, S. G. VEPREV and T. I. AXENOVICH. *Institute of Cytology and Genetics, Novosibirsk, Russia.*

As a rule, the genetic analysis of traits in animals and plants is performed using 'intercross experiment'. This approach requires obtaining two genetically homogeneous strains contrasting by the trait under analysis; it is usually time and money-consuming procedure. At the same time there is a possibility to obtain useful information from data, which were not designed for genetic analysis, for example, from stock-maintenance pedigrees.

To extract information from pedigrees of arbitrary structure the likelihood-based technique of complex segregation analysis (CSA) might be applied (Elston & Stewart, 1971). This method is widely used in human genetics and it is applied for analysis of animal data sometimes. However, up

to date it was not used for analysis of traits in plants: although the technique of CSA allows analysis of plant pedigrees theoretically, practically the modifications are to be made because of specific features of plants' breeding structure and traits. We developed such a modification, which allows analysis of plant pedigree data.

The object under the study was Sugar Beet (*Beta vulgaris* L.). It is normally hermaphroditic, cross-pollinating plant. In some conditions, self-pollination also becomes possible. *The trait under the study* was gene-cytoplasmic Male Sterility (MS), which is characterized by production of corrupted and sterile pollen. MS plants have 'sterile' cytoplasm (S-plasmotype), which contains so-called 'S-mitochondria'. In the simplest case, there is a direct interrelationship: S-plasmotype $<=>$ MS plant and N-plasmotype $<=>$ N plant (male-fertile). The plasmotype is maternally inherited: the progeny have the same plasmotype as the maternal plant has.

However, in real life the situation is more complicated: there are two nuclear genes (FR1 and FR2), which can recover male fertility at the background of S-plasmotype. It is well-known, that these genes are unlinked mendelian genes with additive effect on recovering fertility, but the comparative impact of these genes can change as an environment changes.

The reversion of N-plasmotype to the S-plasmotype might be observed in a population; the rate of this event is very low ($\sim 10^{-4}$). The back event (i. e. reversion of S to N-plasmotype) happens never.

The phenomenon observed was the following (Veprev *et al*. 1997): in comparatively small pedigree (477 plants) the reversion of plasmotype from N to S occurred independently several times. This extremely high reversion rate can not be explained by an ordinary mutation rate. It has been hypothesized, that the reversion might be due to an action of some nuclear mutator gene (MUT). We tested this hypothesis using the method of complex segregation analysis of pedigree data (Aul'chenko *et al*. 1997).

The pedigree analysed was obtained from the single progenitor. The pedigree contained 477 plants, of which 68 were MS. There were 6 generations in the pedigree; a generation, obtained by self-fertilization were followed by a generation obtained by cross-pollination of small groups of sibs. Six plants were self-fertilized and six groups of sibs were cross-pollinated.

The formal model applied was the following: the set of possible genotypes was formed by 3 diallelic genes: the fertility recovering genes FR1 and FR2, and the MUT gene, which was postulated to increase the rate of reversion of plasmotype from N to S. The penetrance function was defined as the probability of phenotype x_i given plants' genotype and the plasmotype, which it received from the maternal plant: $\Pr(x_i\,|g_{FR1}, g_{FR2}, g_{MUT}, c_M)$. The phenotype was assumed to be formed in two stages: the plasmotype inheritance/formation (the stage when MUT gene might act); then the MS or N phenotype is formed (the stage of interaction of cytoplasm and FR genes). Therefore the penetrance function was defined as the product of two terms:

(i) The probability that a plant has S-plasmotype, given its' genotype by the MUT gene (g_{MUT}) and maternal plasmotype (c_M): $\Pr(c_i = S|g_{MUT}, c_M) = \{1$ if $c_M = S$; w_{MUT} if $c_M = N\}$. The probability of alternative event is $\Pr(c_i = N|g_{MUT}, c_M) = 1 - \Pr(c_i = S|g_{MUT}, c_M)$.

(ii) The probability of MS or N phenotype, given plants' plasmotype and its' FR1 and FR2 genotype. Based on previous knowledge, this probability might be described via single nuisance parameter ω, that is comparative impact of FR1 to the fertility recovering: $\Pr(x_i = N|g_{FR1}, g_{FR2}, c_i) = \{1$ if $c_i = N$; $^1/_2[n(FR1)\omega + n(FR2)(1-\omega)]$ if $c_i = S\}$, where n(FRi) denotes number of' + ' alleles at the FRi locus. $\Pr(x_i = MS|g_{FR1}, g_{FR2}, c_i) = 1 - \Pr(x_i = N|g_{FR1}, g_{FR2}, c_i)$. And finally $\Pr(x_i\,|g_{FR1}, g_{FR2}, g_{MUT}, c_M) = \Sigma_{ci}\ \Pr(x_i\,|g_{FR1}, g_{FR2}, c_i)\ \Pr(c_i\,|g_{MUT}, c_M)$

The model also included transmission probabilities for the alleles of MUT gene $\Pr(g|g_m g_f)$, which are the probabilities that an offspring has the genotype g given parental genotypes are g_m and g_f. The

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

self-fertilization was introduced by imposing $g_m \equiv g_f$. The distribution of geno- and plasmotypes in the progeny of a cross-fertilized group of sibs followed from the panmixia assumption and selection imposed on groups (N phenotype). The complete model was described by seven parameters; only three penetrances (w_{MUT}) were of primary interest in the analysis.

Results. We have tested a number of genetical hypothesis. The hypothesis of random occurrence of reversion ($w_{MM} = w_{Mm} = w_{mm} = 0.24$, $\omega = 0.62$) was rejected versus more general mendelian model ($w_{MM} = 10^{-3}$, $w_{Mm} = 10^{-6}$, $w_{mm} = 0.40$, $\omega = 0.58$): $\chi^2 = 19.06$, df = 2, $p < 0.0001$. The hypothesis of recessive action of the MUT gene and completely additive action of FR-genes ($w_{MM} = w_{Mm} = [0]$, $w_{mm} = 0.36$, $\omega = [0.5]$) did not differ from the more general mendelian model: $\chi^2 = 0.76$, df = 3. The Elston–Stewart criterion showed, that mendelian segregation of the putative gene can be accepted: the hypothesis of mendelian transmission probabilities was accepted compared with that of arbitrary transmission ($\chi^2 = 1.06$, df = 3) and the 'environmental' hypothesis (equal transmission) was rejected ($\chi^2 = 22.66$, df = 2, $p < 0.0001$). These results suggested that in the pedigree under analysis the mendelian recessive mutator gene MUT was segregating.

Discussion. In this analysis, we confirmed previous hypothesis of existence of a MUT mutator gene, which controls reversion of N-plasmotype into S-plasmotype. Such nuclear mutator genes, which cause mutations with maternal-type inheritance, are known in plants (e. g. *iojap* gene in maize). It could be also hypothesized, that the founder plant was a heteroplasmon, i.e. it contained N and S-mitochondria in the cytoplasm. Then the pattern of the trait 'inheritance' in the pedigree could be explained to be due to the random sorting out of mitochondria in plants and generations. This hypothesis is hard to check formally, but from this hypothesis it follows, that heteroplasmical plants would be found in the pedigree. However, we were unable to find such plants (Veprev *et al*, 1997). The second reason in favor of the hypothesis of the MUT gene is that a line of Sugar Beet, with a constant conversion rate of about 0.2, had been arisen (Dudareva *et al*. 1990).

REFERENCES

Aul'chenko Yu. S., Veprev, S. G. & Aksenovich T. I. (1997). Conversion of cytoplasm type in Sugar Beet (*Beta Vulgaris* L.) during inbreeding: segregation pedigree analysis. *Genetika* (Moscow) **33**, 943–950.

Dudareva N. A., Veprev S. G., Popovsky A. V. *et al*. (1990). High-rate spontaneous reversion to cytoplasmic male sterility in sugar beet: a characterization of the mitochondrial genomes. *Theor. Appl. Genet*. **79**, 817–824.

Elston, R. C. & Stewart, J. A. (1971) General model for the genetic analysis of pedigree data. *Hum. Hered*. **21**, 523–542.

Veprev, S. G., Dikalova, A. E., Mglinets A. V. *et al*. (1997). Conversion of cytoplasm type in Sugar Beet (*Beta Vulgaris* L.) during inbreeding: genetic analysis and identification of mitochondrial DNA type. *Genetika* (Moscow) **33**, 934–942.

Modelling dominance hierarchies using multi-player game theory. M. BROOM[1], C. CANNINGS[2] and G. T. VICKERS[2]. [1] *University of Sussex*, [2] *University of Sheffield*.

Many animals spend important parts of their lives or their entire lives in groups where individuals do not have separate individual territories, but occupy a combined territory together. A common feature of such groups is the presence of dominance hierarchies, where animals sort themselves into

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

a preference order for feeding, mating etc. Such hierarchies may be linear, as is common in fowl, where bird A dominates all other birds, bird B dominates all others except bird A and so on, or there may be a more complex structure.

Two important questions relating to dominance hierarchies are;

(1) How are these hierarchies formed?

(2) Once formed, are they stable, and if so how are they maintained?

We are concerned with the first of these questions.

We present a game theoretic model of the formation of such a dominance hierarchy.

Most game theoretic models are based upon independent 2-player conflicts due to both their prevalence in nature and their relative mathematical simplicity. However, such models are not always sufficent. The authors have published a series of papers using multi-player models, either in 'pure' multi-player form or as a combination of 2-player contests with a dependence structure.

In modelling dominance hierarchy formation, we take the second approach. Although the contests between individuals are pairwise, the results of individual contests will have an influence upon which opponent is played next (winners will tend to play other winners, losers may play other losers or not fight at all). A structured multi-player game involving non-independent pairwise contests is introduced. We discuss the underlying mathematics of the model, the Evolutionarily Stable Strategies that are obtained and how such a structure can be used as a model for dominance hierarchy formation.

REFERENCES

Barnard, C. J. & Burk, T. E. (1979). Dominance hierarchies and the evolution of individual recognition *J. Theoret. Biol*, **81**, 65–73.

Broom, M., Cannings, C. & Vickers, G. T. (1997). A Sequential-arrivals model of territory acquisition *J. Theor. Biol*, **189**, 257–272.

Broom, M., Cannings, C. & Vickers, G. T. (1996). Choosing a Nest Site: Contests and Catalysts *Am. Nat.*, **147**, 1108–1114.

Broom, M., Cannings, C. & Vickers, G. T. (1997). Multi-player Matrix Games. *Bull. Math. Biol.* **59**, 931–952.

Mesterton-Gibbons, M. & Dugatkin, L. A. (1995). Towards a theory of dominance hierarchies: effect of assessment, group size, and variation in fighting ability. *Behav. Ecol.* **6**, 416–423.

Modelling Dominance Hierarchy formation as a Multi-player game. M. BROOM[1] and C. CANNINGS[2]. [1] *University of Sussex, U.K.*, [2] *University of Sheffield, U.K.*

Animals which live in groups often establish a dominance hierarchy (pecking order) which is known by all members of that group. Disputes are then more easily resolved since individuals 'know their place' in the hierarchy, and the lower individual will usually back-down.

We address the issue of how such a hierarchy may be established within a small group of animals. One possibility is that there is an all-play-all (round robin) 'tournament', so that every pair of individuals know which is dominant; though cycles can occur. Such a tournament requires that n players have $(n(n-1)/2$ 'fights' (e.g. 16 players have 120 fights). A more economical approach is through a knockout tournament of the usual type (such as the Wimbledon Tennis Championship),

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

which requires $n = 2^m$ players have $n-1$ fights (everyone, except the winner loses once). This leaves large groups of individuals with only a crude position in the hierarchy; for a random individual the average number of individuals with an equal position is $(4^m - 3.2^m + 2)/(3.2^m)$, and a proportion $(4^m - 3.2^m + 2)/[3.2^m\ (2^m - 1)]$ of encounters are between individuals of equal standing (e.g. 16 players have 15 fights and the average number of equal individuals is 35/8, and 7/24 of encounters are between equals).

We consider a less familiar (except to chess players) type of tournament; a Swiss tournament. In this (in the simplest case) we have $n = 2^m$ individuals, who play m rounds. In the first round the players are paired at random, resulting in 2^{m-1} winners, and 2^{m-1} losers. Winners are paired, losers are paired. In each successive round all individuals play an individual who has won the same number of contests. At the end of the m rounds there have been $m.2^m$ fights, and there are ${}_mC_r/2^m$ individuals with score r, the average number of opponents with an equal score is now ${}_{2m}C_m/(4^m) - 1$ (e.g. for 16 players there are 64 fights and the average number of equal individuals is 27/8, and 27/120 of encounters are between equals).

We base our Swiss tournament on the classical Hawk–Dove contest, so that every individual fight is such a conflict. The payoff in the tournament is some V_i where i is the number of wins in the n rounds, and $V_i < V_j$, if and only if, $i < j$. We show how to find the Nash strategies, and demonstrate certain features of these. The special cases of 2 and 3 rounds are considered.

Estimating racial and geographical variation in Down syndrome incidence.
A. D. CAROTHERS[1], C. A. HECHT[2] and E. B. HOOK[2]. [1]*MRC Human Genetics Unit, Edinburgh EH4 2XU, U.K.*, [2]*School of Public Health, Berkeley, CA 94720–7360, U.S.A.*

Reported livebirth prevalence of Down syndrome (DS) may be affected by the maternal age distribution of the population, completeness of ascertainment, accuracy of diagnosis, extent of selective prenatal termination of affected pregnancies, and as yet unidentified genetic and environmental factors. To search for evidence of the latter, we reviewed all published reports in which it was possible to adjust both for effects of maternal age and for selective termination (where relevant). We constructed indices that allowed direct comparisons of prevalence rates after standardizing for maternal age. Reference rates were derived from studies previously identified as having near-complete ascertainment. An index value significantly different from one may result from random fluctuations, as well as from variations in the factors listed above. To reduce the effect of the former, we applied James–Stein shrinkage to the entire set of index values obtained.

We found 49 population groups for which an index could be calculated. Methodological descriptions suggested that low values could often be attributed to under-ascertainment. A possible exception concerned African-American groups, though even among these most acceptable studies were compatible with an index value of one. As we have reported elsewhere, there was also a suggestive increase in rates among U.S. residents of Mexican or Central American origin (Hook *et al.* 1999). Nevertheless, our results suggest that 'real' variation between population groups reported to date probably amounts to no more than $\pm 25\%$, and that it is not even possible to exclude the hypothesis of no variation at all. However, reliable data in many human populations is lacking including, surprisingly, some jurisdictions with relatively advanced health-care systems.

The use of indices standardized for maternal age is a simple and easily understood way to compare rates in populations with widely differing maternal age distributions. We suggest that future reports

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

of DS livebirth prevalence should routinely present data that allow calculation of an index standardized for maternal-age and adjusted for elective prenatal terminations. A full description of the study is given by Carothers *et al.* (1999).

REFERENCES

Carothers, A. D., Hecht, C. A. & Hook, E. B. (1999). International variation in reported livebirth prevalence rates of Down syndrome, adjusted for maternal age. *J. Med. Genet.* **36**, 386–393.

Hook E. B., Carothers A. D. & Hecht C. A. (1999). Elevated maternal age specific rates of Down syndrome liveborn offspring of women of Mexican and Central American origin in California. *Prenatal Diagnosis* **19**, 245–251.

An extended tTDT test for uncertain haplotype transmission. D. CLAYTON, *MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge, CB2 2SR, UK.*

The statistical theory of the transmission disequilibrium test for allelic association was briefly reviewed and conditional and full likelihood approaches contrasted. The TDT is based on the former approach and its popularity amongst geneticists is largely due to its robustness against unmeasured population stratification.

A new TDT test statistic was proposed for situations in which transmission is uncertain. Such situations arise when transmission of a multilocus marker haplotype is considered, since haplotype phase is often unknown in a substantial number of instances. Even for single locus markers, transmission is uncertain if one or both parents are missing. In both these situations, uncertainty may be reduced by typing further siblings, whose disease status may be unaffected or unknown.

A natural approach is to use methods based on the 'missing data' likelihood, but this loses the robustness of the conditional likelihood approach. The proposed test is a score test based on a partial score function which omits the terms most influenced by hidden population stratification.

Although the approach is asymptotic, some preliminary attempts to improve the small sample properties by use of Monte Carlo methods were described.

(This paper has been accepted for publication in the *American Journal of Human Genetics*.)

How powerful and robust are the incongruence length differences tests? P. DARLU[1] and G. LECOINTRE[2]. [1]*INSERM U155*, [2]*MNHN, Paris, France.*

The incongruence length differences test (ILD) was proposed by Farris *et al.* (1994) in an attempt to quantify the conflicts that can occur between set of characters from different sources of data used to infer the phylogeny of a given set of taxa. This test is based on the observed difference between the number of steps required by separate and by combined analyses of the data. This difference is compared to the distribution of the differences found for a series of randomized partitions of the characters into matrices of the original sizes. The null hypothesis of congruence is rejected at the 5% level when 95 % of the randomized partitions show an ILD less than the observed ILD.

In this work, we have tested the efficiency of the ILD test to detect incongruence between data when they are simulated under contrasted evolutionary models. We have simulated DNA sequences

under constant (CR) and variable (VR) evolutionary rate (branch lengths being short or long, in a ratio 1:2), trees being balanced (B) or unbalanced (U), with various sequence lengths (L = 100 or 1000 sites), mutation rate being low or high (the expected number of substitution by site, s, being between 0.01 and 0.2) and the substitution rates (between) among sites being homogenous or heterogeneous (expressed in term of the a parameter of the gamma distribution and varying from $\alpha = 0.01$ to large values) (Rambaut & Grassly, 1996).

First, we have compared two set of data generated along the same tree and with the same evolutionary conditions. The probability to falsely reject the true hypothesis of congruence (type I error) turns out to be less than the expected 5%, meaning that the 5% level is rather conservative. Then we have compared two set of data, still generated along the same tree, but with contrasted evolutionary conditions (CR versus VR, low versus large values of mutation rate, homogeneity versus heterogeneity rate of change). We have found that the cases where one falsely rejects the hypothesis of congruence significantly occur when the expected number of changes by site is large and/or when the among-site rate variation is important. As expected, the degree of homoplasy measured by the Consistency Index or the Retention Index seems to have no clear effect on the ILD test.

We have also investigated the cases where the two set of data are generated with the same evolutionary conditions but along two different trees (balanced and unbalanced), in order to evaluate the probability to accept the false hypothesis of congruence (Type II error). It appears that the hypothesis of congruence is wrongly accepted more often than expected, particularly when the length of the sequences is short and the heterogeneity of change among-site rate variation is large. For example, the false hypothesis of congruence is accepted in 48% (α large) and in 91% ($\alpha = 0.06$) of the simulations with L = 100, while these proportions decrease to almost 0% and 52% respectively with L = 1000. The other parameters, as s and CR or VR, seem to be less influential.

Finally, we can conclude that this test has to be cautiously used, because of its lack of power and because it can lead to erroneous conclusions when short sequences are used and/or when the among site rate variation is high.

REFERENCES

Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. (1994). Testing significance of congruence. *Cladistics* **10**, 315–319.

Rambaut, A. & Grassly, N. C. (1996). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.*

Yet more on the Fundamental Theorem of Natural Selection. A. W. F. EDWARDS. *Gonville and Caius College, Cambridge CB2 1TA, U.K.*

An understanding of Fisher's Fundamental Theorem has been inhibited by inappropriate graphical metaphors such as Wright's adaptive surfaces. According to Svirezhev (1972), however, if the gene-frequency space is transformed using Fisher's square-root transformation then the trajectories of change in the continuous-time model are in the direction of steepest ascent of the mean fitness, with a rate corresponding to the square root of the genic variance in fitness. This underlines the different roles of the mean fitness (determining the direction) and the genic variance (determining the rate) and makes clear why an adaptive surface alone cannot in general determine the trajectory.

For two and three alleles and discrete generations both the mean fitness and the genic standard deviation were graphed in the transformed space. The diallelic case turned out to be special because the under the transformation the gradient of the mean fitness is exactly equal to $2\sqrt{2}$ times the genic standard deviation and hence there is thus a simple approximate relationship between the change in the transformed gene frequency and the gradient of the mean fitness in the transformed space. These relationships are lost with three or more alleles and the mean fitness can only be regarded as an adaptive surface (after transformation) for the atypical diallelic case.

REFERENCE

Svirezhev, Y. M. (1972). [*In Russian*] Optimum principles in population genetics. *Studies on Theoretical Genetics*, Novosibirsk, 86–102.

Tell me your HLA, I will tell you whom to marry. E. GENIN[1], C. OBER[2] and G. THOMSON[3]. [1] *INSERM U155, Paris, France*, [2] *Chicago University, USA*, [3] *University of California, Berkeley, USA*.

Mate choice has been the subject of many investigations and debates among population geneticists. Experiments in mice have shown that odours and genes from the Major Histocompatibility Complex (MHC) may influence mating. Female mice generally mate with males different from themselves for genes in the H2 region and the recognition could involve odours. Attempts to generalise these findings in humans however have been difficult for various reasons. Confounding factors such as ethnic preferences and population stratification are one problem, the difficulty to plan an experiment in humans is another problem and the high level of polymorphism in HLA is a last problem.

An experiment performed in a Swiss University by Wedekind *et al.* used sweaty T-shirts to establish that people can sniff out HLA differences and reached the conclusion that females prefer the odour of HLA different males. Another evidence for the possible role of HLA in mate choice comes from a study in the Hutterite populations by Ober et al where it has been shown that the couples shared on average less HLA haplotypes than expected under random mating.

In this study, we have confirmed the findings in Hutterites and outlined a strong sex dependence effect. A preferential choice for mates with HLA -A, -B, -C, -DR, -DQ haplotypes different from the mother's haplotypes was found. To study mating patterns, a new test was implemented that is robust to population stratification and inbreeding. The statistical properties of this test are discussed and compared to the properties of a classical test for assortative mating.

How to deal with model equivalence in the MASC method? F. GUYON[1], P. MARGARITTE-JEANNIN[2], H. QUESNEVILLE[3] and F. CLERGET-DARPOUX[2]. [1] *GIS Infobiogen*, [2] *INSERM U155*, [3] *Institut Jacques Monod, Paris, France*.

The MASC method (Clerget-Darpoux *et al.*, 1988) is designed to model the role of a candidate gene in the etiology of a disease. Unrelated affected individuals (called index cases) are categorized in three nested steps according to

1. the status of their parents and sibs (familial configurations);

2. their marker genotypes;
3. the numbers of parental marker alleles they share with a specific sib.

In each step, the categorization is made conditionally to the one of the previous step. The information on the model parameters differs according to the three steps. The first one corresponds to the familial disease segregation, the second one to the correlation between the marker and the disease at population level and in the last one to the correlation between the marker and the disease at familial level.

Genetic parameters (penetrances and coupling frequencies between the marker and the functional alleles) are estimated by minimising a χ^2 function to fit the observed values in each category to the one expected under the genetic model considered.

Index cases with several affected relatives are usually over-represented. The information on disease segregation (step 1) is no longer available and only steps 2 and 3 are used for parameter estimation. It may result in equivalence between sets of parameters or lack of identifiability of parameters sensitive to this information.

The aim of this study is to reintroduce the disease segregation information (generally available from epidemiological studies): the probability for the parent of an index to be affected (P_0); the risk for a sib of an index to be affected given the two parents are unaffected (r_0), or given one parent is affected (r'_0).

Two different approaches are proposed. In the first one, pseudo observations corresponding to step 1 are generated for a given index sample size N as a function of P_0, r_0 and r'_0. This corresponds to adding a penalty term to the minimization where this term is the χ^2 between the pseudo observations and the expected numbers under the tested genetic model.

Another approach is to add constraints to the χ^2 minimization problem in order to respect bounds on known recurrences risks: $r_0 \pm \delta_r$, $r'_0 \pm \delta_r$, $P_0 \pm \delta_p$. The constrained optimization problem is solved using a Sequential Quadratic Programming DONLP2 (Spelluci, 1994).

Both approaches are not computationally equivalent. The first one uses a non constrained minimization algorithm (based on a Quasi-Newton algorithm Gemini). It changes the minimization function by convexifying it. The second approach doesn't change the minimization function but reduces the set of admissible parameters.

The gain of information is measured through three identifiability and sensivity criteria: the condition number of the Hessian of the minimization function, the size of the support region, and the sensitivity matrix. The condition number of H is the ratio of its highest eigenvalue to its smallest eigenvalue. The support region is the region of accepted parameters given a level of significance (Edwards, 1972). The sensivity matrix is the matrix of the derivatives of observed data with respect to parameters. These criteria give information on the dependance between observed variables and model parameters, and measure the stability of the minimum to perturbation of the data.

The interest of these approaches is illustrated on data simulated with Genoom (Quesneville & Anxolabehere, 1997). A model with two alleles at the functional locus and two alleles at the marker locus is considered. The parameters of this model are: the penetrance matrix whose coefficients are $(f, \lambda_1, f, \lambda_2, f)$ (with f denoting the global penetrance) and the coupling matrix C. Only index cases with at least one affected sib are selected. We analyse penetrance identifiability: just using information relative to steps 2 and 3; using additional information corresponding to step 1 with the two approaches described above.

The identifiability criteria clearly show that without level 1 information, global penetrance is not identifiable. Using disease segregation information improves penetrance identifiability and provides accurate estimates of this parameter. We intend to extend this parameter identifiability study to other parameters and to more complex genetic models.

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

REFERENCES

Clerget-Darpoux, F., Babron, M. C., Prum, B., Lathrop, G. M., Deschamps, I., & Hors, J. (1988). A new method to test genetic models in HLA associated diseases: the MASC method. *Ann. Hum. Genet.* **52**, 247–258.

Edwards, A. W. F. (1972). *Likelihood. An account of the statistical concept of likelihood and its application to scientific inference.* Cambridge: University Press.

Quesneville, H. & Anxolabehere, D. (1997). GENOOM: a simulation package for GENetic Object-Oriented Modeling. *Ann. Hum. Genet.* **61**, 531–550.

Spelluci, P. (1994). A SQP method for general nonlinear programs using only equality constrained subproblems. http://plato. la.

Association analysis of non-normal, quantitative twin data. C. HINDSBERGER. *Department of Biostatistics, Copenhagen, Denmark*

If familial aggregation of a quantitative trait is found, the next logical step is to determine whether or not the trait is affected by genes, and if this is the case, to quantify the degree of this genetic effect. A classical twin study of dizygotic and monozygotic twin pairs makes it possible to divide the variance of the trait Y into components describing the genetic variance, the variance due to shared environment as well as the variance due to individual, non-shared environment.

Under an assumption of normality these parameters can be estimated by the method of maximum likelihood (MLE). If this assumption is not full-filled the MLE-approach may result in poor estimating efficiency and incorrect standard error estimates of the parameters. A more robust approach is to use generalized estimating equations as suggested by Prentice & Zhao (1991).

Non-normality of a continuous trait may however be caused by an unobservable discrete factor due to genes and/or environment. The situation where the discrete factor is a major gene can be modeled by the use of a classical segregation analysis model.

A variance component model that makes it possible to quantify the effect of genes and environment on the discrete factor will be described, and two estimation methods, the maximum likelihood method and a method based on generalized estimating equations will be discussed.

These methods will be illustrated on data from a population of young, Danish twins registered in the Danish Twin Register.

REFERENCE

Prentice, R. L. & Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–839.

Marker assisted selection for multiple traits. C. LANGE, J. C. WHITTAKER and M. C. DENHAM, *Department of Applied Statistics, University of Reading,*

In this talk we propose two different methods for marker assisted selection for multiple traits. While the first method employs generalised estimating equations the second method is an extension of ridge regression for multivariate data. The advantages and disadvantage of both methods are

compared and discussed. Furthermore the differing influence of the marker interval length on the efficiency of both methods is illustrated by a simulation experiment. Additionally we show possible extensions of both methods for non-normally distributed traits.

REFERENCES

Brown, P. J. & Zidek, J. V. (1980). Adaptive multivariate ridge regression. *The Annals of Statistics* **8**, 64–74.

Liang, K-Y & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Whittaker, J. C., Haley, C. S. & Thompson R. (1997). Optimal weighting of information in marker-assisted selection. *Genet. Res.* **69**, 137–144.

Distance to the Truth: Classifying Results of Simulated Genome Scans. S. LOESGEN. *Institute of Epidemiology, GSF-National Research Centre for Environment and Health, Neuherberg, Germany.*

Simulation studies are widely used to characterise and compare the performance of statistical methods for identifying susceptibility genes. Even whole genome simulations are performed and analysed to validate multipoint methods. Here arises a special problem: the definition of true and false peaks. The variety of definitions used in the literature is remarkable.

Peaks in the range of 1–10 cM next to the disease locus for a 1 cM scan (Terwilliger *et al.* 1997) up to any peak on a linked chromosome (Holmans & Craddock, 1997) are accepted as true positives. Boehnke's (1994) theoretical limits of resolution in linkage studies could be taken into account. Further, to characterise a peak, the percentage of values exceeding the threshold in the chosen interval is given, or only the extreme – picked for height or some definition incorporating width (Terwilliger *et al.* 1997). Flanking markers outside the interval have to be dealt with. Many authors use different definitions for significance and power calculations, e. g. percentage of values exceeding the threshold on unlinked chromosomes as false positives while only counting peaks next to the disease locus as true.

Table 1. *True and false positives in dependence on distance tolerated to disease locus*

		'Power'			'Significance'		
Count	Threshold	⩽ 5 cM	⩽ 30 cM	linked	⩽ 5 cM	⩽ 30 cM	unlinked
One	< 2.2E-5	8	11	11	6	0	0[a]
per	< 7.4E-4	27	32	32	20	0	2[a]
interval	< 0.01	63	70	71	63	19	21[a]
One	< 2.2E-5	7	11	11	4	0	0
per	< 7.4E-4	23	32	32	9	0	2
interval	< 0.01	41	69	71	30	2	13
Any peak	< 2.2E-5	8	11	11[b]	3	0	0[b]
gaps <	< 7.4E-4	27	32	32[b]	6	0	2[b]
interval	< 0.01	63	70	73[b]	31	3	15[b]

Counts of 100 simulations of 100 affected sibpair families are given, following a single locus multiplicative disease model with a genotypic relative risk of 4. Chromosome length is 150 cM.

[a] Total number of marker exceeding threshold on unlinked chromosome,

[b] Values exceeding threshold are counted as one when gaps are < 30 cM.

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

A fast and flexible simulation program GESIMS is presented which was used to generate 100 samples of 100 affected sibpair families following a single locus multiplicative disease model with a genotypic relative risk of 4. These datasets were analysed with GENEHUNTER. Table 1 gives the number of true and false positives obtained by some selected criteria. Differences are large, especially for the number of false positives observed. A generally accepted application-oriented definition is called for, closing the gap between unlinked and closely linked for realistic significance and power considerations.

REFERENCES

Boehnke, M. (1994). Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *Am. J. Hum. Genet.* **55**, 379–390.

Craddock, N. & Holmans P. (1997). Efficient strategies for genome scanning using maximum likelihood affected-sib-pair analysis. *Am. J. Hum. Genet.* **60,** 657–666.

Terwilliger, J. D., Shannon, W. D., Lathrop, G. M., Nolan, J. P., Goldin, L. R., Case, G. A. & Weeks, D. E. (1997). True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. *Am. J. Hum. Genet.* **61**, 430–438.

A multi-marker approach to fine scale association mapping of disease loci. A. P. MORRIS and J. C. WHITTAKER. *Department of Applied Statistics, University of Reading, Reading, UK.*

The traditional linkage based approach to mapping disease genes has proved efficient in locating genes contributing a major effect to human diseases to within 1 cM (Devlin & Risch, 1995). In such a small region of the genome, recombination is unlikely to occur so that further refinement of disease gene location will require infeasibly large pedigrees (Boehnke, 1994). As a result, it will not be unusual for linkage based mapping of disease loci to leave 1Mb of DNA to be searched which will be expensive and time consuming.

An alternative approach to disease gene mapping on a fine scale is association analysis. We would expect that association of alleles at the disease locus with alleles at flanking markers will be maintained over many generations whilst the association will be dissipated at more distant markers. The rate of dissipation is related to the recombination fraction between the disease locus and the marker locus. In effect, disease-marker association provides information from many more generations than can be observed in even the largest scale pedigrees.

Here, we present a new approach to fine scale association mapping of a monogenic disease locus, based on the observation of transmissions of alleles from parents to affected offspring in samples of nuclear families at a number of closely spaced markers in a candidate region. The method is developed as a multi-marker generalisation of the allele transmission model of Bickeböller & Clerget-Darpoux (1995), assuming that a high risk disease allele has resulted from a single mutation of the normal allele at the disease locus. We present simulations to illustrate the method, based on the data of Kerem *et al.* (1991) concerned with the location of the ΔF508 mutation for cystic fibrosis on chromosome 7.

REFERENCES

Bickeböller, H. & Clerget-Darpoux, F. (1995). Statistical properties of the allelic and genotypic transmission disequilibrium test for multiallelic markers. *Genet. Epidemiol.* **12**, 865–870.

Boehnke, M. (1994). Limits of resolution of genetic linkage studies: implications for the positional cloning of human genetic diseases. *Am. J. Hum. Genet.* **55**, 379–390.

Devlin, B & Risch, N. (1995). A comparison of linkage disequilibrium measures for fine scale mapping. *Genomics* **29**, 311–322.

Kerem, B. *et al.* (1991). Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073–1080.

Address for correspondence: A. P. Morris, University of Reading, Department of Applied Statistics, Harry Pitt Building, PO Box 240, Whiteknights Road, Reading RG6 6FN, U.K.
e-mail: sns98am@reading.ac.uk.

Complex segregation analysis of nonsyndromic congenital dysplasia of the hip (CDH).

C. SCAPOLI[1], G. BERTOLANI[2], G. ATTI[3] and V. SOLLAZZO[2]. [1]*Dip. di Biologia*, [2]*Dip. Scienze biomediche e terapie avanzate* (*sez. cl. ortopedica*), [3]*Div. Pediatrica*, *Univ. di Ferrara*, *Italy*.

Congenital dysplasia of the hip (CDH) is one of the most common skeletal congenital anomalies. Often bilateral, is more frequent in females than in males, with a 5:1 ratio[1]. A great discrepancy in the incidence and prevalence of CDH can be observed according to geographic and ethnographic situations. For Caucasian populations an incidence of CDH of 1 per 1, 000 can be assumed. In Italy, the incidence is about 10 per 1, 000 and in Ferrara is still higher, equal to 18.4 per 1, 000 live births[2]. Several examples of familiar transmission of isolated CDH are well known; however, only very scarce literature about the definition of the hereditary transmission model of CDH is available, and the cause of CDH is not yet completely clear, even if the importance of genetic and environmental factors is evident[3,4].

To clarify the inheritance pattern of CDH in the population of Ferrara, among all patients treated at the Ferrara's Center for the study of CDH in the period 1991–96, a sample of 171 patients with positive familiar recurrence of CDH was collected. The following criteria for inclusion in the study were used: a) nonsyndromic CDH cases; b) clinical and echographic neonatal diagnosis; c) evaluation of CDH severity according to Graf's echographic classification[5].

The 171 available pedigrees have been partitioned into 507 nuclear families, comprising a total of 1937 individuals, among which 499 affected by CDH. Each person, whether affected or not, has been assigned to a liability class based on sex. Liability frequencies have been taken from sex-specific birth prevalence, as indicated by Atti *et al.*[2]. For nuclear families including the proband, the index case was the true proband of the family with an ascertainment probability (p) assumed to be close to one (= 0.9). In all other families, in order to statistically check the ascertainment of highly selected pedigrees, besides the internal proband a pointer has been added, chosen among the closest affected relatives[6].

Table 1. *POINTER analysis of the 171 total pedigrees*

Model	d[a]	t	q[a]	Z[a]	H[a]	2lnL+C	AIC	χ^2	p
(1) Sporadic	—	—	(0)	—	(0)	942.164	942.2		
(2) Polygenic	—	—	(0)	(1)	0.616	812.308	814.3		
(3) Multifactorial	—	—	(0)	0.205	0.687	806.535	810.5		
(4) ML Recessive	(0)	3.059	0.121	—	—	794.050	798.0	0.048	0.827
(5) ML Codominant	(0.5)	3.126	0.126	—	—	800.558	804.6	6.557	0.011
(6) ML Dominant	(1)	1.741	0.017	—	—	814.643	818.6	20.65	< 0.001
(7) GSL model	0.036	3.111	0.119	—	—	794.002	800.0	0.019	0.892

[a] Parameters fixed by hypotheses in parentheses.

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

Table 2. *COMDS analysis of the 171 total pedigrees*

Model	d[a]	t	q	S[a]	d_{ma}	t_m	q_{ma}	S_{ma}	$-2\ln L+C$	χ^2	*p*
(1) GSL model, $\hat{s}$	0.01	20.6	0.05	0.28	—	—	—	—	1319.44	26.2	
(2) GTL model, S = 1	0[b]	3.36	0.20	(1)	0[b]	2.95	0.24	(1)	1296.89	3.61	0.46
(3) GTL model, $\hat{s}$	0[b]	3.30	0.24	0.63	0[b]	3.06	0.21	1.33	1293.28		
(4) Rec. – Rec.	(0)	3.36	0.20	(1)	(0)	2.95	0.24	(1)	1296.89	3.61	0.46
(5) Rec. – Codom.	(0)	1.45	0.20	(1)	(0.5)	3.16	0.23	(1)	1298.14	4.86	0.30
(6) Rec. – Dom.	(0)	3.37	0.17	(1)	(1)	2.82	0.04	(1)	1298.58	5.30	0.26

[a] Parameters fixed by hypotheses in parentheses.
[b] Moved to a bound.

Nuclear families classified according to these rules have been analysed by complex segregation analysis, using the mixed model of inheritance expanded as the unified model[7], as implemented in the computer programs POINTER and COMDS[8]. The parameters used to test the hypotheses are estimated by maximising the likelihood (L) of the family phenotypes. To choose the most appropriate hypothesis, nested models have been compared by using the LRT. The Akaike information criterion (AIC) has been used for non-nested models.

The results of complex segregation analyses under the conditional likelihood are presented in Tables 1 and 2. Using POINTER (table 1), the hypotheses of sporadic, multifactorial and polygenic transmission of CDH have been strongly rejected (models 1 to 3 vs 7). Among models postulating a major locus only the recessive is accepted (model 4 vs. 7).

Since the CDH diagnosis was made according to Graf's grading system of severity, each affected individual has been assigned to one of four severity classes coinciding with the four pathological levels of Graf's method. Then, when in the COMDS analysis ultrasonographic level is taken into consideration, the estimate of the parameter S leads to a significant improvement of the likelihood ($\chi^2_{[1]} = 50.547$; $P < .001$). Moreover, when the hypothesis of a major plus a modifier locus has been tested, the model strictly fits the data, giving clear evidence for the presence of at least a second locus (table 2, model 6 vs. 5, $\chi^2_{[4]} = 26.16$, $P < 0.001$). Amongst the various two-locus models, the one assuming a recessive transmission for the major gene shows the best likelihood, even if the results for the modifier locus are not as clear as those for the major locus. In fact, for the modifier locus, all the genetic hypotheses (recessive, codominant, dominant) are accepted on the basis of the chi square test (table 2); however, since the recessive-recessive model is the most parsimonious, we consider it to be the appropriate two-locus model for our data.

REFERENCES

Atti, G., De Sanctis, V. & Vigi, V. (1998). Screening della displasia evolutiva dell'anca a Ferrara. *Ital. J. Ped.* **24**, 568–573.

Graf, R. (1995). *Sonography of the infant hip and its therapeutic implications* (eds. R. Graf & B. Wilson) p. 69. Chapman and Hall.

Lalouel, J. M., Rao, D. C., Morton, N. E. *et al.* (1983). A unified model for complex segregation analysis. *Am. J. Hum. Genet.* **35**, 816–826.

Morton, N. E., Rao, D. C. & Lalouel, J. M. (1983). *Methods in Genetic Epidemiology.* Karger (eds), Basel, ch. 5, pp 62 – 98.

Morton, N. E., Shields, D. C. & Collins, A. (1991). Genetic epidemiology of complex phenotypes. *Ann. Hum. Genet.* **55**, 301–314.

Weinstein, S. (1987). Natural history of congenital hip dislocation (CDH) and hip dysplasia. *Clin. Orthop. and Rel. Res.* 225, 62–76.

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

Wynne-Davies, R. (1970). A family study of neonatal and late-diagnosis congenital dislocation of the hip. *J. Med. Genet.* **7**, 315–333.

Zervas, H. Z., Constantopoulos, C. & Theodorou, S. D. *et al.* (1983). HLA antigens and congenital dislocation of the hip. *Tissue Antigens* **22**, 295–297.

Permutation Likelihoods for Analysing BRCA2 Genotype-Phenotype Correlations

D. THOMPSON[1], D. EASTON[1] and Breast Cancer Linkage Consortium members, [1]*CRC Genetic Epidemiology Unit, University of Cambridge.*

Mutations in the BRCA2 gene, located on chromosome 13q, confer susceptibility to breast and ovarian cancer (Wooster *et al.* 1995). To date there have been nearly 600 distinct BRCA2 mutations reported (Breast Cancer Information Core), distributed throughout the gene. In 1997 Gayther et al. identified a 3.3 kb region of exon 11 in which mutations appeared to confer a higher risk of ovarian cancer relative to breast cancer than mutations elsewhere in the gene. This section was named the Ovarian Cancer Cluster Region, or OCCR.

Here we present a study using a much larger set of families with BRCA2 mutations to test and evaluate the OCCR effect. The 163 families included in the analysis were collected by members of the Breast Cancer Linkage Consortium from centres across Western Europe and North America, and all have at least one tested carrier of a protein-truncating BRCA2 mutation.

The odds ratio for ovarian versus breast cancer in OCCR families relative to non-OCCR families was significantly greater than one (Odds Ratio = 3.86, $P < .0001$), confirming the initial observation that cancer risks are different in this part of gene. However, this does not differentiate between an increased ovarian risk, a reduced breast cancer risk relative to the rest of the gene, or a combination of both. To establish the size and significance of each effect, the cancer risks for non-OCCR mutations were estimated as a baseline, allowing the estimation of the relative risks for OCCR mutations relative to non-OCCR mutations. The population frequencies of each BRCA2 mutation are unknown, so the risks were estimated using the likelihood conditional on the set of BRCA2 mutations observed in the study. This involves calculating the likelihood under all permutations of the observed set of mutations seen at each centre. Likelihoods were computed using the program MENDEL (Lange & Weeks, 1988).

The maximum likelihood estimates of the relative risks were 0.65 for breast cancer (95% CI = 0.48–0.88) and 1.34 (95% CI = 0.86–2.09) for ovarian cancer. Thus there is significant evidence that the risk of breast cancer is significantly lower for mutations within the OCCR than for mutations outside it. Evidence that the OCCR effect is actually due to an increased risk of ovarian cancer is weaker, and may be a consequence of the ascertainment strategy. These parameter estimates give a cumulative risk of breast cancer by age 70 of 51.7% (95% CI = 39.6%–61.4%) for non-OCCR mutations, compared to 37.6% (95% CI = 25.9%–47.4%) for mutations in the OCCR. The OCCR definition coincides with the coding region for a sequence of internal repeats in the BRCA2 protein (Gayther & Ponder, 1998) which have been shown to interact with the RAD51 DNA repair gene (Wong *et al.* 1997) and are thought to play an important role in BRCA2's function.

REFERENCES

Breast Cancer Information Core at http://www.nchgr.nih.gov/Intramural_research/Lab_transfer/Bic

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press

Gaytner, S. A., Mangion, J., Russell, P., *et al.* (1997). Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. *Nature Genetics*. **15**, 103–105.

Gaytner, S. A. & Ponder, B. A. J. (1998). Clues to the Function of the Tumour Susceptibility Gene BRCA2. *Disease Markers* **14**, 1–8.

Lange, K. & Weeks, D. (1988). Programs for Pedigree Analysis: MENDEL, FISHER and dGENE. *Genetic Epidemiology* 5, 471–472.

Wong, A. K. C., Pero, R., Ormonde, P. A. *et al.* (1997). RAD51 Interacts with the Evolutionarily Conserved BRC Motifs in the Human Breast Cancer Susceptibility Gene BRCA2. *J. Biol. Chem.* **272**, 31941–31944.

Wooster, R., Neuhausen, S. L., Mangion, J. *et al.* (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792

Marker-assisted selection using ridge regression. J. C. WHITTAKER[1], R. THOMPSON[2] and M. C. DENHAM[1], [1] *Department of Applied Statistics, University of Reading*, [2] *IACR Rothamsted and Roslin Institute* (*Edinburgh*).

In crosses between inbred lines linear regression can be used to estimate the correlation of markers with a trait of interest; these regression coefficients then allow marker assisted selection (MAS) for quantitative traits. However, a subset of markers to include in the model must be selected, and no completely satisfactory method of doing this exists. We discuss some of the problems introduced by this model selection procedure and suggest some alternative approaches. In particular, we show that in simulation experiments replacing this model selection procedure by ridge regression can improve the mean response to selection and reduce the variability of selection response. We explain how ridge regression can be viewed as a Bayesian shrinkage procedures; this suggests a number of avenues which may lead to further progress. Finally, we note that this and related problems will become increasingly important as MAS becomes more commonly used and as marker densities increase.

Systematic reviews of genome screens using the Genome Screen Meta-Analysis method. L. H. WISE and C. M. LEWIS, *Division of Medical and Molecular Genetics, Guy's, King's and St. Thomas' Schools of Medicine, London.*

Genome search results are now available for many complex traits and several studies may be performed in a single disease. However, due to the lack of power of individual screens to detect loci of small effect, non replication of results across studies is a common problem. There is a clear need for a method of systematically reviewing the results of such studies that will assess evidence for linkage across studies.

The GSMA method (Wise *et al.*, in press) provides a systematic quantitative overview of the separate analyses whilst dealing with some of the problems specific to meta analysis of genome screens such as differences in family structures, markers genotyped and the original method of analysis. The GSMA uses genome wide results from each study (eg lod scores, p values) but does not require original genotype data.

Proof of principle and power properties of the GSMA have been investigated by applying the method to the GAW11 problem 2 data set (Greenberg *et al.*, in press). This is simulated genome search data for a complex disease exhibiting genetic heterogeneity in three separate populations, studied in four separate centres.

As proof of principle the GSMA method was used to systematically review the results of four simulated genome screens, one from each study centre. At an NPL level of 2 only one of the individual studies identified the three susceptibility loci common to all populations and the mean number of false positives per screen was 2.5. The GSMA method correctly identified the regions containing these three susceptibility loci at the 95% confidence limit with one false positive region detected, thus indicating the value of a meta-analysis of genome searches.

To investigate the power properties of the GSMA 25 systematic reviews were performed each using the results of 4 separate analyses and the results were compared with the results of a pooled analysis of the 4 contributory data sets. The power of the GSMA for a given type 1 error rate was always higher than that of the individual studies, and approached the power of a pooled analysis. For example the power of the GSMA to detect at least three loci at a type one error rate of 5% is 0.56 compared to 0.16 for their individual analyses and 0.68 for the pooled analysis.

We conclude that the GSMA is a valid and powerful tool for systematically reviewing the results of genome searches in complex diseases which deals with some of the problems specific to these reviews and does not require original genotype data.

REFERENCES

Greenberg, D. A., Maccluer, J. W., Spence, M. A., Falk, C. T. & Hodge, S. E. (1999). Genetic Analysis Workshop 11: Development of Problem 2. In: *Genetic Analysis Workshop 11: Analysis of genetic and environmental factors in common diseases*. (Eds Goldin L, Amos CI, Chase GA, Goldstein AM, Jarvik GP, Martinez MM, Suarez BK, Weeks DE, Wijsman EM, and MacCluer, JW) *Genetic Epidemiology* (In press).

Wise, L. H., Lanchbury, J. S. & Lewis, C. M. (1999). Meta-analysis of genome searches. *Ann. Hum. Genet.* **63**, 263–272.

https://doi.org/10.1111/j.1469-1809.1999.ahg634_0351_6.x Published online by Cambridge University Press