GENETIC TESTING STRATEGIES AND METHODS IN SCIENTIFIC RESEARCH

Anna **GRZYWACZ** [A-F]

Faculty of Physical Education, University of Rzeszów, Rzeszów, Poland

Keywords:
- Genetics
- Polymorphisms
- Association study

Abstract:

The genetic studies with regard to sport trace their beginnings to 1998, when, for the first time, the relationship between the insertion-deletion polymorphism of the ACE gene and an aptitude for sport was described. Since then our knowledge of genetics in sports, a range of studies performed with this regard, but also potential possibilities of applying this kind of knowledge have changed significantly. In the latest reports on the genetic determinants of sports predisposition, there are descriptions of genetic research. This article describes the most important research strategies for genetic research. As in the clinical sciences, the relationship between the gene or polymorphism and the disease is examined, so in the studies of athletes we combine certain phenotypic traits with the genotype. Knowledge of basic strategies and research methods may facilitate interpretation of results, understanding of research or designing of a research project.

Polymorphism

The major genetic factors in complex disease models are not single mutations that dramatically change a gene or its product, but rather, those that involve more subtle genetic changes that may slightly alter the expression or function of a gene product. Because these gene variants (alleles) alter susceptibility to disease, they are referred to as functional variants. Many functional variants occur with a relatively high frequency in the general population. When a specific allele occurs in at least 1% of the population, it is said to be a genetic polymorphism. More than 10 million single nucleotide polymorphisms (SNPs) have been identified in the human genome [Hart 2003]. Polymorphism arises as a result of mutation. The different types of polymorphisms are typically referred to by the type of mutation that created them. The simplest type of polymorphism results from a single base mutation which substitutes one nucleotide for another, and has recently been termed as a single nucleotide polymorphism (SNP). Other types of polymorphism are restriction fragment length polymorphism (RFLP) and simple tandem repeats (STRs), consisting of relevant allele or nucleotide repetition. [Kinane 2003, Vijayalakshmi 2010].

Genetic polymorphisms are very useful in genetic studies of the population. Frequencies of genotypes and alleles may differ between a diseased group and a healthy group. Subsequently, when a given allele is identified as associated with a disease, functional studies may be conducted to investigate the possible role of that gene in the etiology and pathogenesis of the disease [Schafer 2011, Vijayalakshmi 2010] . A number of SNPs are

likely to be important determinants in disease susceptibility for the more common, genetically complex diseases [Pontes 2004, Trevilatto 2011]. In this model, a single functional genetic polymorphism associated with disease (at a population level) is not sufficient to cause disease, and is therefore not deterministic of disease in itself. Consequently, such a functional polymorphism may be found in individuals who have no evidence of disease and who may not be at great risk for disease. A fundamental characteristic of this genetic model is that such genetic polymorphisms are more frequent in the population than are mutations, and the correlation between genetic polymorphisms and disease is generally much weaker than the relationship between a functional mutation and a disease phenotype [Backdahl 2009].

Epigenetics

Epigenetics is described as the study of changes in patterns of gene expression, which do not involve changes in the DNA sequence. Epigenetic events act through chemical modifications of DNA and its associated proteins by blocking the binding of transcription factors through histone modifications (considered more transient) or DNA methylation (stable form of gene regulation) [Backdahl 2009].

Genetic association studies

Until the middle of the last decade,candidate-gene association studies were the most important strategy used for the identification of diseasesusceptibility genes. In such study designs, an a-priori hypothesis is made on the involvement of a selected gene in the disease risk and on the presence of a functional variant within this particular gene [Wilkening 2009]. The hypotheses for the selection of these risk genes are based on the current knowledge of the molecular biological mechanisms of the disease. Candidate loci can be selected on the basis of a hypothesis on the relevance of a specific gene in the disease etiology and will give an answer on the presence or absence of disease-associated variants within this gene. The limitation of the candidate gene approach is that hundreds of genes, which may have an influence on the disease, will not be selected because their functions are unknown or they lie within pathways that have not yet been implicated with the disease. In contrast to the candidate gene approach, GWAS are advocated to allow 'hypothesis-free' and unbiased analysis of the genome for identifying disease-associated genetic variants. The current standard for declaring statistical significance at a genomewide level is a combined p-value (initial exploratory and replication cohorts) [Manolio 2010, Thompson 2011]. To achieve such significance thresholds, a arge number of cases and controls are required. Accordingly, meaningful genome-wide testing of polymorphisms is difficult, if not impossible, in small (i.e. underpowered) populations, which include only several hundred cases and several hundred controls. Nonetheless, the allele frequencies of disease-associated variants may be enriched by choosing a case sample of patients with early-onset disease or of patients with very severe phenotypes, for which it is believed that genetic factors largely contribute to disease development [Schaefer 2011]. By doing so, the statistical power of a case–control sample, of limited numbers, can be enriched. However, the statistical power is also dependent on the minor allele frequency of the variants of interest in the study population [Conrad 2006, Marchini 2010, Schaefer 2011a, Schoof 2011]. As a consequence, rare variants, which may contribute strongly to the genetic risk of a disease, will not reach statistical significance at a genome- wide level. Efforts in reducing phenotypic heterogeneity and also focussing on severe phenotypes will improve the power to detect causal variants, which will be especially helpful; the strong contrast between cases and controls should provide researchers with greater differences in the carriage rate within the case–control study population and therefore could improve the power of a study [Schaefer 2011, McCarthy 2008].

Linkage disequilibrium

If the same alleles of several single nucleotide polymorphisms (SNPs) in a given locus tend to be inherited together, they form a haplotype. This co-occurrence of specific DNA markers at a higher frequency than predicted by random chance is termed linkage disequilibrium (LD). The degree of LD is commonly expressed by the r2 value. If the r2 value of two SNPs of similar frequency is > 80% they are considered to be in high LD, meaning that the same alleles of these SNPs are usually inherited together. If, for any two alleles, the LD is lower, it may be less likely that they are inherited together. It would be difficult, time-consuming and expensive to genotype each existing SNP and to determine whether it plays a role in human disease. Using haplotypes, researchers can sample one SNP of each haplotype instead of studying each single one [definitions from the US National Library of Medicine, NIH (http://www.nlm.nih.gov/)]. [Manolio 2009].

Genome-wide association studies

GWAS designs are matched case–control cohort studies [Kruglyak 2008]. The samples are specifically matched, in terms of genetic ancestry (ethnicity), with a similar distribution of gender, whilst age is matched based on the study design and the research question. This matching is crucial to avoid populationpopulation stratification. For example, genetic heterogeneity occurs if different ethnic populations have been mixed. This may prevent their detection and/or replication. The identification and mapping of polymorphisms through the International HapMap Project [Frazer 2007, Altshuler 2010, HapMap 2003]. It is currently estimated that there are more than 30–40 million SNPs within the human genome, most of which have neutral effects [Abecasis 2010]. Imputation of GWAS data, with HapMap and 1000 Genomes Project genotypes as references, allows the researcher to localize the diseaseassociated region more accurately. Many large genome-wide association meta-analyses have recently been performed for various disease consortia, for example coronary artery disease [Fehringer 2012] , type 2 diabetes [Zeggin 2008], inflammatory bowel disease and rheumatoid arthritis [Stahl 2010]. Through meta-analyses, investigators were able to combine their genotype data and to substantially increase the statistical power. Thus, many new risk variants were identified and previously published associations were validated. In a GWAS, it is generally difficult to select a true SNP association from random SNP associations, which are caused by chance fluctuations of allele frequencies between case and control samples. Using a commonly preassigned significance threshold of 0.05, one of every 20 markers tested will pass this significance threshold by chance alone. If 500,000 or 1 million SNPs are tested, a considerable correction for multiple hypothesis testing must be applied, and, as a consequence, GWAS generally suffer from a lack of statistical power to overcome the large multiple hypothesis correction threshold. Thus, GWAS are forced to ignore weaker association signals, which necessarily will result in true associations being undetected [Manolio 2009]. One way to ovfunctionally related and/or interact with one another in biological pathways that are assumed to have the potential to contribute to the development of the disease [Marchini 2005, Peng 2010]. It is likely that a limited number of biological pathways have the potential to contribute to the etiology of a specific disease [Carlborg 2004]. A pathway-based geneset analysis reduces the number of variants tested, and, as a consequence, the threshold for multiple testing. This will allow the identification of some of the variation that otherwise remains undetected in a GWAS analysis. Pathway analyses within GWAS data sets have been used to investigate genetic-susceptibility gene sets in various diseases such as melanoma, lung cancer and skin cancer [Schoof 2011, Zhang 2011, Fehringer 2012].

REFERENCES

1. Abecasis G.R., Altshuler D., Auton A. et all. (2010), A map of human genome variation from population-scale sequencing. "Nature" vol. 467, pp. 1061–1073.
2. Altshuler D.M., Gibbs R.A ., Peltonen L. (2010), Integrating common and rare genetic variation in diverse human populations. "Nature " vol. 467, pp 52–58.
3. Backdahl L., Bushell A., Beck S. (2009), Inflammatory signalling as mediator of epigenetic modulation in tissue-specific chronić inflammation. "Int J Biochem Cell Biol", vol.41(1), pp. 176-84.
4. Carlborg O., Haley C.S. (2004), Epistasis: too often neglected in complex trait studies? "Nat Rev Genet" vol. 5, pp. 618–625.
5. Conrad D.F., Jakobsson M., Coop G. et al. (2006), A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. "Nat Genet", vol. 38, pp. 1251–1260.
6. Hart T.C. Genes and gene polymorphisms associated with periodontal disease. (2003), "Crit Rev Oral Biol Med", vol. 14(6), pp. 430-449.
7. Fehringer G., Liu G., Briollais L. et al. (2012), Comparison of pathway analysis approaches using lung cancer GWAS data sets. "PLoS ONE", vol. 7, pp. 3181-3186.
8. Frazer K.A., Ballinger D.G., Cox D.R. et al. (2007), A second generation human haplotype map of over 3.1 million SNPs. "Nature", vol. 449, pp. 851–861.
9. Kruglyak L. The road to genome-wide association studies.(2008), "Nat Rev Genet" vol. 9, pp. 314–318.
10. Manolio T.A. Genomewide association studies and assessment of the risk of disease. (2010), "N Engl J Med", vol. 363, pp. 166–176.
11. Manolio T.A., Collins F.S., Cox N.J. et al. (2009), Finding the missing heritability of complex diseases, "Nature", vol. 461, pp. 747–753.
12. Marchini J., Donnelly P., Cardon L.R. (2005), Genome-wide strategies for detecting multiple loci that influence complex diseases, "Nat Genet", vol. 37, pp. 413–417.
13. Marchini J., Howie B. (2010), Genotype imputation for genome-wide association studies, "Nat Rev Genet" vol. 11, pp. 499–511.
14. McCarthy M.I., Abecasis G.R., Cardon L.R. et al. (2008), Genome-wide association studies for complex traits: consensus, uncertainty and challenges, "Nat Rev Genet", vol. 9, pp. 356–369.
15. Peng G., Luo L., Siu H. et al. (2010), Gene and pathway-based second-wave analysis of genome-wide association studies, "Eur J Hum Genet", vol. 18, pp. 111–117.
16. Pontes C.C., Gonzales J.R., Novaes J., Taba Junior M., Grisi M.F., Michel J., et al. (2004), Interleukin-4 gene polymorphism and its relation to periodontal disease in a Brazilian population of African heritage, "J Dent", vol. 32(3), pp. 241-246.
17. Schaefer A.S., Jepsen S., Loos B.G. (2011a), Periodontal genetics: a decade of genetic association studies mandates better study designs, "J Clin Periodontol",vol. 38, pp. 103–107.
18. Schafer A.S., Jepsen S., Loos B.G. (2011), Periodontal genetics: a decade of genetic association studies mandates better study designs, "J Clin Periodontol", vol. 38, pp. 103-107.
19. Schoof N., Iles M.M., Bishop D.T., Newton- Bishop J.A., Barrett J.H. (2011), Pathwaybased analysis of a melanoma genomewide association study: analysis of genes related to tumour-immunosuppression, "PLoS ONE", vol. 6, pp. 29451.
20. Stahl E.A., Raychaudhuri S., Remmers E.F. et al. (2010), Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. "Nat Genet", vol. 42, pp. 508–514.

21. The International HapMap Consortium. The International HapMap Project. (2003), "Nature", vol. 426, pp. 789–796.
22. Thompson J.R., Attia J., Minelli C. (2011), The meta-analysis of genome-wide association studies. "Brief Bioinform", vol. 12, pp. 259–269.
23. Trevilatto P.C., de Souza Pardo A.P., Scarel-Caminaga R.M., de Brito R.B., Jr., Alvim-Pereira F., Alvim-Pereira C.C., et al. (2011), Association of IL1 gene polymorphisms with chronic periodontitis in Brazilians, "Arch Oral Biol", vol.56(1), pp. 54-62.
24. Vijayalakshmi R., Geetha A., Ramakrishnan T., Emmadi P. (2010), Genetic polymorphisms in periodontal diseases: an overview, "Indian J Dent Res",vol. 21(4), pp. 568-574.
25. Wilkening S., Chen B., Bermejo J.L., Canzian F. (2009), Is there still a need for candidate gene approaches in the era of genomewide association studies?, "Genomics", vol. 93, pp. 415–419.
26. Zeggini E., Scott L.J., Saxena R. et al. (2008), Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for Type 2 diabetes, "Nat Genet", vol. 40, pp. 638–645.
27. Zhang M., Liang L., Xu M., Qureshi A.A., Han J. (2011), Pathway analysis for genomewide association study of basal cell carcinoma of the skin, "PLoS ONE", vol.6, pp. 22760.