Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 100–104
FSI GENETICS
www.elsevier.com/locate/fsig

Review

Finding genes that underlie physical traits of forensic interest using genetic tools

Hannah Pulker [a], María Victoria Lareu [a], Christopher Phillips [b], Angel Carracedo [a,b,*]

[a] *Institute of Legal Medicine, University of Santiago de Compostela, Spain*
[b] *National Genotyping Centre-USC Node, Hospital Clinico Universitario, Santiago de Compostela, Spain*

Received 3 February 2007; accepted 6 February 2007

Abstract

Association studies using SNPs provides one of the best tools that we have at the moment for looking for genes involved in physical traits. However the studies should be carefully designed from the very beginning in all the steps of the procedure: pre-genotyping, genotyping and the mathematical analysis of the results. If the actual knowledge is correctly applied in the design of the study the probability of being successful in finding an association can be considerably increased. Improved statistical analysis techniques are helping in the robustness of the findings. The current consensus from the literature indicates that this would be a good time to investigate complex or quantitative traits via dense SNP genotyping, and a number of studies have been published, providing potential models. The state of the art of candidate genes for pigmentation, stature and facial morphology is described.

Keywords: Complex traits; Physical traits; Association studies; SNP typing; Forensics

Contents

1. Introduction

From the genetic point of view physical traits are generally speaking complex traits, this is to say multigenic and multifactorial traits where different genes interacting both between themselves and with the environment define the phenotype. The propensity of the genetic background to modify the phenotypic expression of most, if not all, Mendelian traits suggests that few traits are truly monogenetic and most are genetically complex [1].

Despite the characterisation in the last 20 years of many of the genes known to control simple Mendelian traits, relatively few genes underlying complex traits have been identified, but this situation is changing quickly. Genes that contribute to complex traits pose special challenges such as allelic heterogeneity, locus heterogeneity, phenocopies, phenotypic variability, variable expressivity and gene–gene or gene–environment interactions, making gene discovery difficult. The prospects for success have improved markedly with the recent development of an array of genomic and proteomics technologies and resources. Among them genetic association studies has proved to be an excellent tool to assess correlations between genetic variants and differences in traits on a population scale.

* Corresponding author. Tel.: +34 600942293; fax: +34 981580336.
E-mail address: apimlang@usc.es (A. Carracedo).

1872-4973/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2007.02.009

There are a variety of genetic tools to analyze the genetic component of a disease, each one having its domain of applicability. Thus, classical linkage analysis of families, although powerful for detecting loci involved in single gene disorders (such as BRCA genes), is less effective for complex traits where association studies have demonstrated more power to detect genes with small effects [2]. Association studies were until recently hampered by the low density of available markers. The great jump in the field was the discovery of millions of SNP markers in the human genome when DNA from multiple donors was sequenced and compared for the genome sequencing projects. Now more than 11 million SNPs have been gathered into the publicly accessible dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) with a large proportion of these listed with validated allele frequencies.

Genotyping technologies have also experienced a rapid evolution and there are now a range of high-throughput SNP typing approaches available, with a variety of platforms and chemistries allowing researchers to use the most appropriate one for each specific purpose. In some countries national genotyping facilities have been set up. This is the case in Spain where the Spanish National Genotyping Center (CeGen: www.cegen.org) offers Spanish researchers a complete range of technologies for genotyping plus pre-genotyping (SNP selection) and post-genotyping (association study analysis) services. These centres provides a straightforward and inexpensive facility for researchers to perform association studies of any size. In addition, experts can help with SNP assay designs and selection of the platform best suited for the characteristics of each project.

However progress in genotyping technology would not have been sufficient without the parallel advance of the HapMap project in mapping SNPs and their correlation as groups in haplotypes. The question is that if we had to perform a whole genome scan with 10 million SNPs in 1000 samples, a medium size for an association study, this would represent 10 billion genotypes, an impossible task in terms of workload and cost. The discovery [3] that clustering is observed in all the autosomes and that the human genome contains haplotype blocks, that is to say regions with little evidence of recombination, separated by recombination hot spots, gave another perspective to association studies. The ability to identify the blocks and the tagSNPs defining all the variation in the block, reduces the SNPs required to examine the entire genome for association with a phenotype from 10 million to 500,000 tagSNPs making genome scan approaches more efficient and comprehensive. For this reason the HapMap project (www.hapmap.org) was launched and the first phase recently finished [4]. Since linkage disequilibrium can be affected by a number of factors affecting any given population, the HapMap project initially examined the three main population groups (Asians, Europeans and Africans). Using HapMap software such as Haploview, researchers can use HapMap information to view patterns of haplotype distribution and select tagSNPs to help design the most efficient association studies.

Designing an association study is not easy and for traits such as cancer requires a large number of samples (so the establishment of networks is usually a pre-requisite), the definition of the phenotype, the definition of study populations and the decision to use a candidate gene approach or whole genome scans (WGS). WGS have the advantage of being free of bias towards specific genes but the disadvantage of being the most costly. If a candidate gene approach is chosen, appropriate candidates can be selected by looking at pathways, using comparative genomics, gene-expression profiling or reviewing markers informative for ancestry since selection signatures can provide clues for genes involved in complex traits.

A common approach used in selecting SNPs at candidate genes is a two stage strategy looking for possible causative SNPs initially (mis-sense, non-sense, splicing sites, transcription factor sites, AIMs) followed by the addition of SNPs obtained from regulating regions at frequencies $\geq$5% for haplotype analysis. Since collection of samples, meeting ethical requirements, definition of the phenotype and collection of clinical data is a substantial effort requiring networks it is always a good strategy to collect epidemiological data (with an appropriate protocol) with a view to the long-term research basis for the study of gene–environment interactions. Despite involving more effort this ultimately adds considerably more value to the research.

Inability to replicate results in association studies has led to increasing scepticism about the value of this approach to genetic analysis. It is true that many thousands of association studies have been performed with massive investment of research funding with relatively limited success, but the situation is changing and well designed studies now performed have higher probability of success since much has been learnt over the last few years. Now we know that without replication or functional studies, we cannot rule out the possibility of false positive results. Replication is key to the reliability of a study and it is a requirement for the publication of an association observation in any important journal. A correct design is essential and the population from which the samples are collected matters and although we can take advantages of specific populations for specific designs of association studies (*i.e.* isolated populations, populations that have experienced bottlenecks and expansions or populations with recent admixture) they can also represent a potential source of problems. Notably stratification is one of the most common causes of false association and checking for the influence of potential stratification in the population used for the study is also required. Equally important are the trait and sample size: the trait matters since the definition of the phenotype is far from easy, and as we have mentioned, the sample size matters as it is impossible to find weak associations without an adequate number of study subjects. Finally there are several genetic phenomena that can add to the complexity of the results, for example pleiotropy, when a single gene influences multiple phenotypic traits.

Despite all these problems and the complexity of physical traits, a fairly extensive amount of information has accumulated in the last few years. Some examples of particular relevance to forensic analysis are described below, but many other physical traits related with diseases (especially common traits such as myopia) are targets of potential forensic interest.