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.

2. Pigmentation

A great deal of work has focused on human pigmentation, most notably skin colour.

Pigmentation traits overall show high heritability score (60–90%) with skin colour genes showing additive effects. The inheritance of human skin colour follows quasi-Mendelian patterns indicative of a polygenetic trait with a few major genes of strong effect coupled with modifier genes [5].

The first candidate genes and mutations were identified in mouse affecting coat colour, most of which have human homologues in which null mutations cause albinism [6]. For the same body region, light- and dark-skinned individuals have similar numbers of melanocytes (with considerable variation between different body regions), but the melansomes are larger, more numerous and more pigmented in dark compared to intermediate and light skin, the enzymes implicated in melanin synthesis are thus good contenders for genetic variation causing diversity in pigmentation. The different forms of albinism all have in common a normal number of melanocytes, but confer varying degrees of impairment in the production of melanin.

Modification of the skin pigmentation phenotype is caused by confounding variables and this is a concern for any study. Such factors affecting human pigmentation include:

- Exposure to UVR causes an increase in facultative melanin production, with a subsequent darkening of the skin.
- Different body sites are pre-programmed to have differing numbers of melanocytes and constitutive melanin production.
- The amount and type of melanin production varies with age and site.
- Children are paler than adults and females are generally paler than males but some studies contradict these observations.

In contrast, Shriver et al. [7] maintain that constitutive pigmentation in adults is a stable trait, and largely independent of environmental influences when measured in unexposed areas. To examine skin pigmentation as a quantitative trait it will be important to obtain quantitative and accurate measures of phenotype, in an attempt to reduce the ''noise'' that could mask any association. This will probably require new samples, not archive material. Skin colour can be measured via broadband spectrophotometry (reflectance/remittance)—but clearly it is necessary to consider variability by site, age, sex and changes due to ambient UVR. The quantity and type of melanin in skin can also be assayed via HPLC analysis of specific degradation products but this approach requires a biopsy. HPLC has been used in studies of hair colour, but an established method has yet to be published. Asking study subjects about their phototype (reaction to UVR) could be a useful, ordinal addition to reflectance data, as the two phenotypes are different variables.

The biochemical pathways of human pigmentation are relatively well characterized and a number of gene candidates for possible influence on the normal range of skin colour have been identified. The main candidate genes are shown in Table 1 and several comprehensive recent reviews and research on the topic can be found in [8–10]. To identify the loci that likely contribute to among-population human skin pigmentation differences Myles et al. [9] measured allele frequency differentiation among Europeans, Chinese and Africans for 24 human pigmentation genes from two large-scale SNP data sets. This revealed patterns of widely contrasting frequency differences between populations coupled with signals of strong recent selection. Notably the DCT gene was found to be strongly associated with pigmentation control in Asians. In addition interesting recent work on independent loss of skin

Table 1
Principal skin pigmentation candidate genes

Locus	Chromosome	Protein	Mut phenotype	Function
Melanosome proteins				
TYR	11q14-11q21	Tyrosinase	OCA1	Oxidation of tyrosine
TYRP1	9p23	Gp75, TRYP1	OCA3	DHICA-oxidase, TYR stabilisation
DCT	13q32	DCT, TRYP2		Dopachrome tautomerase
OCA2	15q11.2-15q12	P-protein	OCA2 (eye)	pH of melanosome
SLC45A2	5p14.3-5q12.3	MATP, AIM-1	OCA4 (skin)	Melansome maturation
SLC24A5	15q21.1	Cation exchanger		Melanosome precursor
Signal proteins				
ASIP	20q11.2-20q12	Agouti signal protein		MC1R antagonist
MC1R	16q24.3	MSH receptor	Red hair (skin)	G-protein coupled receptor
POMC	16q24.3	MSH receptor	Red hair	MC1R antagonist
OA1	Xp22.3	OA1 protein	OA1	G-protein coupled receptor
MITF	3p12.3-3p14.1	MITF	Waardenburg	Transcription factor
Proteins involved in melanosome transport or uptake by keratinocytes				
MYO5A	15q21	MyosinVa	Griscelli	Motor protein
RAB27A	15q15-15q21.1	Rab27a	Griscelli	RAS family protein
HPS1	10q23.1-10q23.3	HPS1	Hermansky-Pudlak	Organelle biogenesis and size
HPS6	10q24.32	HPS6	Hermansky-Pudlak	Organelle biogenesis

ACTH: adrenocorticotrophin hormone; DCT: dopachrome tautomerase; DHICA: 5,6-dihydroxyindole-2-carboxylic acid; MATP: membrane-associated transporter protein; MC1R: melanocortin-1 receptor; MITF: microphthalmia-associated transcription factor; MSH: melanocyte stimulating hormone; OCA: oculocutaneous albinism; POMC: pro-opiomelanocortin; TYRP1: tyrosinase-related protein 1.

pigmentation in Europeans and Asians [10] suggests that ASIP and OCA2 play important roles in global pigmentation patterns with SLC24A5, SLC45A2 and TYR affecting only pigmentation in Europeans. Clearly many of the pigmentation genes showing the largest frequency differences between populations also have considerable potential as components of ancestry informative panels for forensic use.

3. Stature

Adult height is a complex trait that may be amenable to linkage analyses, in that heritability has been found to be high, and large groups of study subjects can be phenotyped by easy, reliable and accurate measurement. Thought to be the result of additive effects of a large number of genes; as a normally distributed quantitative trait, height is probably purely polygenic and influenced by multiple, potentially undetectable genes.

Studies of stature as a quantitative and complex trait appear to be less advanced than that of skin pigmentation with fewer groups engaged in studies. To date fewer gene candidates have been suggested and characterized, but options do exist for high-density linkage analysis. Furthermore the phenotype is more easily recorded and environmental impact within a population may be more easily controlled than for UV induced pigmentation.

In 2003 Cole [11] published an overview of human physical growth showing the recorded secular trend in human stature is spread over generations and from the current data 150 years of optimal growth conditions are required for a generation to reach their genetic potential for height. Environment has a large impact on growth; poor growth and short stature have been recognised as trade marks of deprivation, with factors such as social class, income, education, family size, housing and urban location all being implicated in the secular trend in height. Other authors assert that adult height has a heritability of between 76 and 90% [12], as determined from family and twin studies. In these studies correlation coefficients for sib pairs and parent–offspring pairs were essentially identical, suggesting that the high heritability in the study population is not due to shared environment (as greater similarity would be expected for sib pairs)—although it must be considered that culture and environment are also shared in families, for example the culture of the parent will affect both their mate-choice and the environment of the offspring.

Important summary points include:

- Mean length at birth has not changed—the secular trend in adult height occurs during the first 2 years of life and it is restricted to this period. The same author reports that the trend to increasing height is due almost entirely to an increase in leg length, *i.e.* secular height trend arises from increased leg length growth during the first 2 years of life [11].
- Height is normally distributed within the general population [12].
- There is evidence of assortative mating [13].
- Statistical analysis of association indicate that there is a considerable polygenic effect on height [13].
- Regions with strongest evidence of linkage in a given study show little evidence for linkage in most or all of the other studies [11].
- Sex difference in adult height has increased with time, because the female trend has been less than the male.

Linkage studies so far have yielded some new, but partially conflicting, data about the key factors that influence growth and final adult height. In contrast, the examination of 'candidate genes' has been very fruitful in identifying those genes that are responsible for some well-defined hormonal deficiencies in patients with severe short stature. As the systematic examination of such genes in short children with and without hormonal deficiencies has become feasible, it appears that the phenotypic appearance of some of these disorders is variable to an extent that makes it difficult to differentiate them from a short normal child. Both the candidate gene and whole genome scan approaches [14] have produced results that have already furthered our understanding of the complex mechanisms that influence growth. Several short stature-QTLs (STQTL) have been identified through genome-wide linkage analysis: STQTL1 on chromosome 6q24, STQTL2 (OMIM 606256) on chromosome 7q31-q36, STQTL3 (OMIM 606257) on 12p11-q14, STQTL4 (OMIM 606258) on 13q32-q33, STQTL5 (OMIM 608982) on 3p26, STQTL6 (OMIM 300591) on Xq24, STQTL7 (OMIM 609822) on 1p21, and STQTL8 (OMIM 610114) on 9q22.

Many of the genes found related with short stature are associated with mutations in growth hormone-related genes. X-linked short stature (OMIM 300582) is associated with mutations in the SHOX gene (OMIM 312865) (see OMIM 604271 for a discussion of autosomal short stature). Finally, polymorphisms in the Fibrillin I gene are associated with tall stature in normal individuals [15].

4. Facial morphology

A search of the literature has not revealed any published work on normal variation of continuous features contributing to human facial morphology. Some characteristics have been anticipated to be monogenetic, following Mendelian inheritance such as Chin dimple (Y-shaped fissure of chin) (OMIM 119000), facial dimples (OMIM 126100), hairy ears (helix of pinnae) (OMIM 139500), earlobe attachment (OMIM 128900), widow's peak (pointed frontal hairline) (OMIM 194000) and freckles [16]

The possibility of objectively measuring facial morphology to define phenotype gives the potential for a "first view" of variation, whether to be defined as continuous or discrete. The accepted and established facial similarity between zygotic twins when compared with dizygotic twins, indicates high heritability for these features. However a concern must be the correct targeting of real and meaningful variables in the dimensions of the human face.

5. Final remarks

Association studies using SNPs is 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. Only in this way can we ultimately benefit from the impressive advances in genomics.

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. Statistics is a challenge in this area with many opportunities for innovation [17]. Better tools for proper clustering (the process of subdividing data into homogenous groups) and new tests for checking for interactions are required.

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 for this approach.

Acknowledgements

This work has been supported by the grants BIO2006-06178 (Ministerio de Educacion) and PGIDT06PXIB228195PR (Xunta de Galicia/PGIDTIT).

References

[1] A.M. Glazier, J.H. Nadeau, T.J. Altman, Finding genes that underlie complex traits, Science 298 (2002) 2345–2348.
[2] N. Risch, K. Merikangas, The future of genetic studies of complex human diseases, Science 273 (1996) 1516–1517.
[3] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, D. Altshuler, The structure of haplotype blocks in the human genome, Science 296 (2002) 2225–2229.
[4] International HapMap Consortium, A haplotype map of the human genome, Nature 437 (2005) 1299–1320.
[5] G.S. Barsh, What controls variation in Human skin colour? PloS Biology 1 (2003) 19–22.
[6] D.C. Bennett, M.L. Lamoreux, The colour loci of mice—a genetic century, Pigment Cell Res. 16 (2003) 333–344.
[7] M.D. Shriver, E.J. Parra, S. Dios, C. Bonilla, H. Norton, C. Jovel, C. Pfaff, C. Jones, A. Massac, N. Cameron, A. Baron, T. Jackson, G. Argyropoulos, L. Jin, C.J. Hoggart, P.M. McKeigue, R.A. Kittles, Skin pigmentation, biogeographical ancestry and admixture mapping, Hum. Genet. 112 (2003) 387–399.
[8] B. McEvoy, S. Beleza, M.D. Shriver, The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model, Hum. Mol. Genet. 15 (2006) 176–181, Spec No. 2.
[9] S. Myles, M. Somel, K. Tang, J. Kelso, M. Stoneking, Identifying genes underlying skin pigmentation differences among human populations, Hum. Genet. 120 (2007) 613–621.
[10] H.L. Norton, R.A. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, V.A. Canfield, D.G. Bradley, B. McEvoy, M.D. Shriver, Genetic evidence for the convergent evolution of light skin in Europeans and East Asians, Mol. Biol. Evol. 24 (2007) 710–722.
[11] T.J. Cole, The secular trend in human physical growth: a biological view, Econ. Hum. Biol. 1 (2003) 161–168.
[12] M.R. Palmert, J.N. Hirschhorn, Genetic approaches to stature, pubertal timing, and other complex traits, Mol. Genet. Metab. 80 (2003) 1–10.
[13] J. Xu, E.R. Bleecker, H. Jongepier, T.D. Howard, G.H. Koppelman, D.S. Postma, D.A. Myers, Major recessive gene(s) with considerable residual polygenic effect regulating adult height: confirmation of genomewide scan results for chromosomes 6, 9 and 12, Am. J. Hum. Genet. 71 (2002) 646–650.
[14] J.M. Hirschhorn, C.M. Lindgren, M.J. Daly, A. Kirby, S.F. Schaffner, N.P. Burtt, D. Altshuler, A. Parker, J.D. Rioux, J. Platko, D. Gaudet, T.J. Hudson, L.C. Groop, E.S. Lander, Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence if linkage to adult height, Am. J. Hum. Genet. 69 (2001) 106–116.
[15] M. Mamada, T. Yorifuji, J. Yorifuji, K. Kurokawa, M. Kawai, T. Momoi, T. Nakahata, Fibrillin I gene polymorphism is associated with tall stature of normal individuals, Hum. Genet. 120 (2007) 733–735.
[16] X.J. Zhang, P.P. He, Y.H. Liang, S. Yang, W.T. Yuan, S.J. Xu, W. Huang, A gene for freckles maps to chromosome 4q32-q34, J. Invest. Dermatol. 122 (2004) 286–290.
[17] T. Thornton-Wells, J. Moore, J.L. Hainees, Genetics, statistics and human diseases: analytical rettoling for complexity, Trends Genet. 20 (2004) 640–647.