



**HAL**  
open science

## **Biobanks and the importance of detailed phenotyping: a case study – the European Glaucoma Society GlaucoGENE project**

Panayiota Founti, Fotis Topouzis, Leonieke van Koolwijk, Carlo Enrico Traverso, Norbert Pfeiffer, Ananth C Viswanathan

### ► **To cite this version:**

Panayiota Founti, Fotis Topouzis, Leonieke van Koolwijk, Carlo Enrico Traverso, Norbert Pfeiffer, et al.. Biobanks and the importance of detailed phenotyping: a case study – the European Glaucoma Society GlaucoGENE project. *British Journal of Ophthalmology*, 2009, 93 (5), pp.577-n/a. 10.1136/bjo.2008.156273 . hal-00477843

**HAL Id: hal-00477843**

**<https://hal.science/hal-00477843>**

Submitted on 30 Apr 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Biobanks and the importance of detailed phenotyping: a case study – the European Glaucoma Society GlaucoGENE project**

Panayiota Founti<sup>1</sup>, Fotis Topouzis<sup>1</sup>, Leonieke van Koolwijk<sup>2,3</sup>, Carlo Enrico Traverso<sup>4</sup>, Norbert Pfeiffer<sup>5</sup>, Ananth C. Viswanathan<sup>2,6</sup>

1. A' Department of Ophthalmology, School of Medicine, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece
2. Glaucoma Research Unit, Moorfields Eye Hospital, London, United Kingdom
3. Glaucoma Service, The Rotterdam Eye Hospital, Rotterdam, The Netherlands
4. Centro di Ricerca Clinica e Laboratorio per il Glaucoma e la Cornea, Clinica Oculistica, Di.N.O.G., University of Genoa, Genoa, Italy
5. Department of Ophthalmology, University Eye Hospital, Mainz, Germany
6. Department of Visual Science, Institute of Ophthalmology, London, United Kingdom

**Corresponding Author:** Ananth C. Viswanathan, FRCOphth MD

**Mailing Address:**

Glaucoma Research Unit

Moorfields Eye Hospital

City Road

London, EC1V 2PD

**Tel. Number:** +442075662625

**Fax Number:** +442075662972

**e-mail:** vis@ucl.ac.uk

**Keywords:** biobank, DNA databank, phenotyping, glaucoma

**Word count:** Abstract 188 words, main text 2996 words

## **Abstract**

Dissecting complex diseases has become an attainable goal through large-scale collaborative projects under the term “biobanks”. However, large sample size alone is no guarantee of a reliable genetic association study and the genetic epidemiology of complex diseases has still many challenges to face. Among these, issues such as genotyping errors and population stratification have been previously highlighted. However, comparatively little attention has been given to accurate phenotyping. Study procedures of existing large-scale biobanks are usually restricted to very basic physical measurements and non-standardised phenotyping, based on routine medical records and health registry systems. Considering that the objective of an association study is to establish genotype-phenotype correlations, it is doubtful how easily this could be achieved in the absence of accurate and reliable phenotype description. The use of non-specific or poorly defined phenotypes may partly explain the limited progress so far in glaucoma complex genetics. In this report we examine the European Glaucoma Society GlaucoGENE project, which is the only large multicentre glaucoma-specific biobank. Unlike previous biorepositories, this initiative focuses on detailed and standardised phenotyping and is expected to become a major resource for future studies on glaucoma.

## Introduction

The major progress in identifying the genetic basis of Mendelian disorders has not been followed by similar achievements in mapping complex diseases, defined as diseases that not exhibit classic Mendelian inheritance attributable to a single gene but are determined by a number of genetic and environmental factors.<sup>1</sup> Specifically, there has been a failure of genetic association studies to discover susceptibility loci or replicate initial positive genotype-phenotype correlations in complex diseases.<sup>2-12</sup> Inadequate statistical power to detect small and moderate effects was recognised as one of the major limitations.<sup>2,3,13-15</sup> The need for large sample sizes has led to numerous large-scale collaborative projects that systematically store biological material linked to clinical and other information. These so-called “biobanks” are designed to create unprecedented opportunities for understanding the pathogenic basis of common diseases and ultimately for implementing genetic findings in clinical practice and public health.<sup>9,16,17</sup> On the other hand, they have raised profound ethical issues<sup>18-21</sup> and scepticism on whether benefits will outweigh costs.<sup>22-24</sup> What remains unquestionable is that the genetic epidemiology of complex diseases has still many challenges to face, mainly in terms of study design and methodology.<sup>7-11,22,23,25-30</sup> Among these, we emphasize the importance of detailed and standardised phenotyping, which has not been given the attention it deserves<sup>12</sup> and does not seem to have been employed in some large biobanks. Since complex diseases are characterised by large phenotypic variability<sup>1</sup>, this raises concerns such as how genetic findings derived from such initiatives could be correctly related to the different clinical aspects of a complex disease.

With regards to ophthalmic complex diseases, breakthroughs have been already made in mapping age-related macular degeneration (AMD).<sup>31-34</sup> The strong effect of a complement factor H variant in AMD (odds ratio >2.45 and population attributable risk up to 50%) was possibly behind the success of these studies, where

well-defined criteria for diagnosis were used, although no detailed phenotyping was considered. However, the identification of genetic variants with smaller effects and associations with specific aspects of the phenotype would have possibly required a more detailed phenotypic assessment. This seems to be the case in glaucoma, where genetic findings mostly refer to the minor fraction of cases that follow mendelian rules of inheritance, meaning that the genetic background of the common, non-mendelian forms of glaucoma remains largely unknown.<sup>35-37</sup>

In 2010 there will be 60.5 million people with glaucoma worldwide, among which 8.4 million will be bilaterally blind; by 2020, these numbers will increase to 79.6 million and 11.2 million respectively.<sup>38</sup> On the basis of new opportunities presented in the postgenomic era, the European Glaucoma Society (EGS) GlaucoGENE project has been designed to provide a reliable, extensive and stable resource to enhance research studies in glaucoma genetics. With detailed and standardised phenotyping as one of its basic principles, this project is not only one of the very few phenotype-genotype databases in the field of ophthalmology, but may also be regarded as a pioneer biobank.

### **Biobanks – definitions and classification**

A “biobank” generally refers to a repository of biological material. In genetic research the term is typically used to describe a biological sample collection from which genetic information can be extracted, matched with clinical and other information. However, several definitions can be found and no international consensus has been reached. “DNA databank”, “DNA bank” and “genetic dataset” are commonly used synonyms of “biobank”.

The American National Bioethics Advisory Commission uses the term “DNA bank” to describe a facility that stores extracted DNA or other biological materials for future DNA analysis, which are usually stored with some form of individual

identification for later retrieval.<sup>39</sup> The Public Population Project in Genomics (P3G), which is a principal international body for the harmonisation of biobanks, sets the additional criterion of large number of samples collected.<sup>40</sup> On the other hand, the Swedish Act on Biobanks (2002:297) focuses on the potential of data re-identification rather than the number of samples<sup>41</sup>. It has been suggested that what differentiates a genetic study from a biobank is that the former focuses on specific genetic hypotheses, while the latter is oriented toward future hypotheses that may not be framed at the outset.<sup>42</sup>

Based on overall methodology, biobanks may be either disease-specific or non disease-specific<sup>43</sup>. The term “population-based biobanks” is commonly used to describe the latter category, although subjects are not always randomly selected from the population of reference. Disease-specific biobanks are usually case-control studies recruiting subjects who have developed the disease of interest, as well as healthy individuals. Non disease-specific biobanks are typically cohort studies, where subjects are recruited from the general population to be followed-up over time; depending on study methodology, the recruitment process may involve only healthy individuals or not. However, this is a very crude classification and several study designs have been employed in biobanks so far.

### **Existing biobanks – how are phenotypes assessed?**

Based on the catalogue of the P3G observatory, over 100 biobanks with a sample size of more than 10,000 subjects have been completed or are currently being conducted.<sup>44</sup> In addition there are several collaborative projects or networks, each one involving a number of biobanks, such as the GenomEUtwin<sup>45</sup> and the European Prospective Investigation into Cancer and Nutrition (EPIC).<sup>46</sup> Previous articles have extensively discussed study design with regards to cohort versus case-

control studies as the optimum approach for studying complex diseases.<sup>9,10,22,23,26</sup> In this report we focus on the phenotyping, which we describe as the methodology used to collect information on phenotypes. Following the well-known example of deCODE genetics<sup>47</sup>, several national cohort studies have been designed to provide DNA databases. The UK Biobank<sup>48</sup>, the CARTaGENE in Canada<sup>49</sup>, the Estonian Genome Project (EGP)<sup>50</sup> and the Kadoorie Study of Chronic Disease in China<sup>51</sup> are only some examples of large-scale cohort biobanks aiming to investigate the genetic basis of multiple important chronic diseases. In all of them, baseline assessment is limited to questionnaires and very basic anthropometric or physical measures. With regards to follow-up, no information is available for the EGP, while in all other projects outcomes will be assessed by focusing on end-points, that is whether a disease is present or not, through routine medical or other health-related records and national registry systems. Therefore, no standardised phenotyping is to be performed. Considering that case-control<sup>48,51</sup> and case-cohort<sup>48</sup> studies will be nested within these projects, it is very unclear how cases and controls may be reliably selected, and moreover, how the variety of each disease phenotype will be ascertained. The UK Biobank investigators acknowledge the need for intensive phenotyping in the future, however they recognize that this would not be feasible for the whole cohort, nor has there been detailed discussion on what these measures should be.<sup>48</sup>

In disease-specific biobanks, although a wide variety of clinical information is usually available, standardisation may not be included. For example, in the Inflammatory Breast Cancer Research Foundation Biobank<sup>52</sup> and the National Psoriasis Victor Henschel BioBank<sup>53</sup> clinical information is provided through medical records. Standardisation of phenotypes has been a concern even in projects involving standardised baseline assessment, such as the MORGAM, which is a multinational collaboration of cardiovascular cohorts and a component of the

GenomEUtwin.<sup>54</sup> Due to the nature of the study, a limited number of phenotypes could be standardised with precision.<sup>54</sup> Moreover, harmonisation in data management including quality, completeness and consistency is of particular importance in projects involving a large number of biobanks and such efforts have been already conducted by the GenomEUtwin and the EPIC investigators.<sup>55,56</sup>

### **The importance of detailed and standardised phenotyping in complex diseases**

As opposed to mendelian disorders, causal variants in complex diseases are expected to have rather small effects<sup>2,15</sup>, explaining why sample size is a key determinant in association studies<sup>2,3,13-15</sup>. However, due to the small effect size, the credibility of an association, meaning the likelihood that an association exists after some evidence has been accumulated, may largely depend on the ability to control for errors and bias.<sup>25</sup> This is a serious consideration for studies nested within biobanks where potential sources of errors and bias have not been properly addressed. To date, most reports on potential confounding focus on genotyping errors and population stratification<sup>10,25-27,57-59</sup>, while little attention has been given to phenotyping<sup>12</sup>. However, even modest levels of error in either the genotyping or the phenotyping may result in significantly diminished power of a study.<sup>11</sup>

Issues related to phenotypic assessment, such as establishing diagnostic criteria for a disease, determining what measurements to perform, using validated techniques for data collection and distinguishing cases from healthy individuals are not new to medical research. However, when investigating the genetic component of a complex disease, there are additional reasons why they become so important. The essence of an association in genetic epidemiology is to investigate how a genotype is correlated to a phenotype.<sup>60</sup> Complex diseases are typically characterised by phenotypic heterogeneity, which refers to the large variability of clinical



manifestations within the same disease.<sup>1</sup> Phenotypes belonging to a complex disease are composed of a constellation of clinical signs, only some of which may be present in an individual. Elevated intraocular pressure may or may not be present in a patient with glaucoma. Alternatively, clinical signs belonging to several 'pure' phenotypes may be present in the same individual. When examining an individual with pseudoexfoliative glaucoma, clinical signs of the optic disc do not exclusively belong to the pseudoexfoliative glaucoma phenotype. Both these situations preclude meaningful phenotypic classification into discrete disease states. Moreover, phenotypes may vary with respect to age of onset of clinical symptoms. Chronic late-onset disorders are typically the result of decades-long processes, developing slowly along a continuum from health to pathology. Therefore, clinical signs may be present at below the threshold for definite classification and early cases may be misclassified. For the same reason, it is often difficult to characterise an individual as unaffected. Accordingly, for gene polymorphisms and mutations to be correctly related to the variable aspects of a complex disease it is important to ensure that phenotypic variation is captured with the same precision as genetic variation.<sup>12,61</sup>

Balancing measurement precision and feasibility is a difficult task in research, especially when aiming to recruit thousands of participants. Wong MY et al presented the formula for calculating the sample size required to study the interaction between a continuous exposure and a genetic factor. According to their calculations, smaller studies with better measurements would be as powerful as studies even 20 times bigger, which employ fewer and less accurate measurements.<sup>23</sup>

However, a large number of measurements alone cannot guarantee a high quality phenotypic dataset. It is imperative that the fundamental principles of "traditional" epidemiology, including use of standardised and reproducible measurements, strict criteria for training, certification and quality control are adopted in genetic association studies.<sup>28</sup> Standardisation is of particular importance to ensure

that a uniform set of data is collected across the study and to avoid data unreliability and inconsistency. Also, because biobanks usually rely on multicentre collaborations, standardisation should be the goal both within and between centres. Based on the consensus meeting of the Human Genome Epidemiology Network (HuGENet) Working Group on the Assessment of Cumulative Evidence, any bias due to phenotypic measurements could affect not only the magnitude, but even the presence or absence of an association.<sup>25</sup> Also, prospective standardisation of phenotypes is the only way to ensure that there is low to no likelihood of bias to invalidate an observed association, even in small effect sizes (odds ratio<1.15).<sup>25</sup>

### **The European Glaucoma Society (EGS) GlaucoGENE Project**

In the context of detailed and standardised phenotyping, we present the basic principles and overall methodology of the EGS GlaucoGENE project, which is a large scale pan-European genetic epidemiology research network. This initiative has been developed by GlaucoGENE, a Special Interest Group of the EGS. Its objective is to create a central database consisting of genetic and standardised phenotypic information from people throughout Europe. With the additional component of proteomics, the database is expected to become a major resource for future studies on glaucoma genetics.

The combined genotype-phenotype approach of glaucoma should inform the strategy for future advances in glaucoma risk stratification and therapy. In addition, because recent studies suggest that glaucoma patients reveal specific patterns of protein and peptides<sup>62-66</sup>, the identification of potential protein biomarkers, and furthermore the correlation between protein expression and genotype is likely to lead to a better understanding of disease mechanisms.

The EGS GlaucoGENE project focuses on several subtypes of open-angle glaucoma and angle-closure glaucoma. With systematic phenotyping and ascertainment of probands, family members and controls, genetic analysis will be possible at a number of different levels. These range from the estimation of heritability to quantify the relative importance of genetic and environmental factors, commingling and segregation analyses to identify genes of major effect, to genetic mapping by linkage and association. In addition, the relationships between various glaucoma-related phenotypes and possible gene-environmental interactions may be examined at all these levels.

Standard operating procedures for a most detailed clinical examination, special training and certification have been incorporated to ensure standardisation within and between centres. Also, discrete levels of the phenotypic dataset have been defined to surmount anticipated differences in equipment and infrastructure among centres. The complete dataset involves, among others, imaging of the optic nerve structure with laser imaging technologies, diurnal intraocular pressure curves and laboratory diagnostics. Similarly, standard operating procedures have been employed for biological samples handling to minimize genotyping errors and to ensure high quality of serum samples. A web-based system with limited access to authorized personnel allows reliable, high quality and accessible data management and ensures data integrity and safety. Applications have also been developed to facilitate data completeness and data flow control, as well as automated perimetry and imaging quality control. All data are anonymised and held in a central database. Guidelines address the circumstances, in which data and samples will be re-identified, the designation of the personnel who will approve and perform the re-identification and the procedures to be used for this purpose.

A feasibility study for the EGS GlaucoGENE project began in May 2007 and is expected to be completed by the end of 2008, with the participation of four centres:

Moorfields Eye Hospital, London, UK, Aristotle University of Thessaloniki, Greece, University of Genoa, Italy and University of Mainz, Germany. The Institute of Ophthalmology, University College London (UCL), UK, and the University of Mainz are responsible for handling and storage biological samples. Prospective standardisation of phenotypes, validation of the information system for digital data entry and storage and evaluation of the web service supporting digital data transfer are among the goals of the feasibility study. During the main study a large number of European centres are expected to participate.

## **Discussion**

Major advances in genetics<sup>67-69</sup>, coupled with progress in bioinformatics and statistics have revolutionized genetic studies of complex diseases. In addition, large sample sizes have become feasible through national and international collaborative initiatives, such as biobanks and consortia. The recent findings of the Wellcome Trust Case Control Consortium (WTCCC) in 7 complex diseases denote the effectiveness of the genome wide association approach.<sup>70</sup> On the other hand, since there are major problems in dissecting the molecular basis of even simple monogenic diseases, this challenge is far greater in complex diseases.<sup>29</sup> Considering the amount of human and financial resources invested in biobanking, issues related to study design become of critical importance. Among these, phenotyping requires special attention in terms of both adequacy and standardisation, but has not been properly addressed in several large-scale biobanks. This is partly due to the trade-off between sample size and measurement precision. However, employing better measurements may be a more appropriate strategy than attempting to deal with error by increasing sample size.<sup>23</sup> In order ultimately to implement genetic findings in clinical practice, more refined

questions should be addressed and this may not be possible through broad phenotypes.<sup>60</sup>

The concept of multifactorial genetics holds the promise for future advances in glaucoma management. A personalized approach involving effective screening to identify individuals at risk, establishing a precise diagnosis and predicting rate of progression and response to treatment is clearly a long-term but not an unrealistic goal.<sup>71</sup> The progress achieved so far in glaucoma complex genetics, involves only the reported association of LOXL1 gene with exfoliation glaucoma<sup>72</sup>, which has already been replicated in independent studies.<sup>73-76</sup> However, genetic association studies on primary open-angle glaucoma have had conflicting results or have not been replicated.<sup>37</sup> Poor specificity in the currently used phenotype parameters is a possible explanation<sup>37</sup>, indicating that glaucoma genetics should focus on quantitative trait locus (QTL) studies, using variables such as intraocular pressure and cup-to-disc ratio.<sup>77,78</sup>

To this end, a glaucoma-specific biobank would be of great scientific value. Although we agree that cohort studies, case-control studies and family studies will all be needed for optimum progress<sup>9</sup>, the case-control design holds two major advantages: far greater statistical power to detect associations can be achieved, because a larger number of cases can be studied and a more detailed and disease-specific ascertainment of the phenotype is feasible than in a cohort design.<sup>22</sup> To date, there are very few glaucoma-specific biobanks<sup>79,80</sup>, while the EGS GlaucoGENE project is the only large multicentre biobank covering this need. Under the umbrella of the WTCCC2, another promising initiative is currently under construction, where data are available from 3 well-designed population-based studies in glaucoma.<sup>81</sup> On the other hand, the eyeGENE, which is a biobank involving specifically ophthalmic diseases, focuses on mendelian disorders<sup>82</sup> and therefore may not be of great value for glaucoma complex genetics.

Unlike previous biobanks, the EGS GlaucoGENE project focuses on both detailed and standardised phenotyping and therefore may be regarded as an innovative effort in genetic epidemiology overall . Based on its multicentre structure, a large number of well-characterised glaucoma cases and controls will be achieved, which are also expected to be representative of the European population. Special training, periodic control of data completeness and data quality control by certified centres are also among the strengths of the study. In addition, the feasibility study is almost completed, providing prospective standardisation of procedures, which will increase the likelihood to identify associations even in small effect sizes. For all these reasons the EGS GlaucoGENE project should provide a broad and comprehensive framework for future studies in glaucoma genetics.

**Funding:** The EGS GlaucoGENE Project is supported by a research grant from the European Glaucoma Society Foundation

**Licence statement:** The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in British Journal of Ophthalmology and any other BMJ PGL products to exploit all subsidiary rights, as set out in BJO licence (<http://bjo.bmj.com/ifora/licence.pdf>).

## REFERENCES

1. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994;265:2037-48.
2. Colhoun HM, McKeigue PM, Davey SG. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361:865–72.
3. Lohmueller KE, Pearce CL, Pike M, *et al*. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genet*. 2003;33:177–82
4. Hirschhorn JN, Lohmueller K, Byrne E, *et al*. A comprehensive review of genetic association studies. *Genet. Med*. 2002;4:45–61.
5. Ioannidis JP, Ntzani EE, Trikalinos TA, *et al*. Replication validity of genetic association studies. *Nature Genet*. 2001;29:306–9
6. Ioannidis JP. Common genetic variants for breast cancer: 32 largely refuted candidates and larger prospects. *J. Natl Cancer Inst*. 2006;98:1350–3
7. NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, Boehnke M, *et al*. Replicating genotype-phenotype associations. *Nature*. 2007;447:655-60
8. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1-22.
9. Davey Smith G, Ebrahim S, Lewis S, *et al*. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet*. 2005;366(9495):1484-98.
10. Cordell HJ, Clayton DG. Genetic association studies. *Lancet*. 2005;366:1121-31.
11. Page GP, George V, Go RC, *et al*. "Are we there yet?": Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet*. 2003;73:711-9.
12. Healy DG. Case-control studies in the genomic era: a clinician's guide. *Lancet Neurol*. 2006;5:701-7.
13. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet*. 2005;366:941-51.
14. Ioannidis JP. Genetic associations: false or true? *Trends Mol Med*. 2003;9:135-8.
15. Ioannidis JP, Trikalinos TA, Ntzani EE, *et al*. Genetic associations in large versus small studies: an empirical assessment. *Lancet*. 2003;361:567-71.
16. Collins FS, Green ED, Guttmacher AE, *et al*; US National Human Genome Research Institute. A vision for the future of genomics research. *Nature*. 2003;422:835-47.

17. Khoury MJ, Millikan R, Little J, *et al.* The emergence of epidemiology in the genomics age. *Int J Epidemiol.* 2004;33:936-44.
18. Haga SB, Beskow LM. Ethical, legal, and social implications of biobanks for genetics research. *Adv Genet.* 2008;60:505-44.
19. Lowrance WW, Collins FS. Ethics. Identifiability in genomic research. *Science.* 2007;317:600-2
20. Blatt RJ. Banking biological collections: data warehousing, data mining, and data dilemmas in genomics and global health policy. *Community Genet* 2000;3:204-11
21. Cassa CA, Schmidt BW, Kohane IS, *et al.* My sister's keeper?: genomic research and the identifiability of siblings. *BMC Med Genomics.* 2008;1:32.
22. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet.* 2001;358:1356-60.
23. Wong MY, Day NE, Luan JA, *et al.* The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *Int J Epidemiol.* 2003;32:51-7.
24. Barbour V. UK Biobank: a project in search of a protocol? *Lancet.* 2003;361:1734-8
25. Ioannidis JP, Boffetta P, Little J, *et al.* Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol.* 2008;37:120-32.
26. Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet.* 2005;366:1315-23.
27. Clayton DG, Walker NM, Smyth DJ, *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet.* 2005;37:1243-6
28. Ellsworth DL, Manolio TA. The Emerging Importance of Genetics in Epidemiologic Research III. Bioinformatics and statistical genetic methods. *Ann Epidemiol.* 1999;9:207-24.
29. Peltonen L, McKusick VA. Genomics and medicine. Dissecting human disease in the postgenomic era. *Science.* 2001;291:1224-9.
30. Burton PR, Hansell AL, Fortier I, *et al.* Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol.* 2008 Aug 25. [Epub ahead of print]
31. Edwards AO, Ritter R 3rd, Abel KJ, *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* 2005;308:421–44.
32. Klein RJ, Zeiss C, Chew EY, *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385–9.



33. Haines JL, Hauser MA, Schmidt S, *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005;308:419–21.
34. Hageman GS, Anderson DH, Johnson LV, *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* 2005;102:7227–32.
35. Wiggs JL. Genetic etiologies of glaucoma. *Arch Ophthalmol.* 2007;125:30-7
36. Iyengar SK. The quest for genes causing complex traits in ocular medicine: successes, interpretations, and challenges. *Arch Ophthalmol.* 2007;125:11-8.
37. Hewitt AW, Craig JE, Mackey DA. Complex genetics of complex traits: the case of primary open-angle glaucoma. *Clin Experiment Ophthalmol.* 2006;34:472-84.
38. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol.* 2006;90:262-7.
39. National Bioethics Advisory Commission:  
<http://bioethics.georgetown.edu/nbac/hbm.pdf>
40. P3G Consortium: <http://www.p3gconsortium.org>
41. Swedish Act on Biobanks 2002:297:  
<http://www.sweden.gov.se/content/1/c6/02/31/26/f69e36fd.pdf>
42. Lavori PW, Krause-Steinrauf H, Brophy M, *et al.* Principles, organization, and operation of a DNA bank for clinical trials: a Department of Veterans Affairs cooperative study. *Control Clin Trials.* 2002;23:222-39
43. Yuille M, van Ommen GJ, Bréchet C, *et al.* Biobanking for Europe. *Brief Bioinform.* 2008;9:14-24.
44. P3G Observatory: <http://www.p3gobservatory.org>
45. Peltonen L; GenomEUtwin. GenomEUtwin: a strategy to identify genetic influences on health and disease. *Twin Res.* 2003;6:354-60.
46. Riboli E, Hunt K, Slimani N, *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): Study populations and data collection. *Public Health Nutr.* 2002;5:1113-24.
47. deCODE Genetics: <http://www.decode.com>
48. UK Biobank protocol: <http://www.ukbiobank.ac.uk/docs/UKBProtocolfinal.pdf>
49. CARTaGENE: <http://www.cartagene.qc.ca>
50. Estonian Genome Project: <http://www.geenivaramu.ee>

51. Chen Z, Lee L, Chen J, *et al.* Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). *Int J Epidemiol.* 2005;34:1243-9.
52. Inflammatory Breast Cancer Research Foundation Biobank:  
<http://www.ibcresearch.org/diagnosed/biobank>
53. National Psoriasis Victor Henschel BioBank:  
<http://www.psoriasis.org/research/biobank>
54. Evans A, Salomaa V, Kulathinal S, *et al.*; MORGAM Project. MORGAM (an international pooling of cardiovascular cohorts). *Int J Epidemiol.* 2005;34:21-7
55. Muilu J, Peltonen L, Litton JE. The federated database--a basis for biobank-based post-genome studies, integrating phenome and genome data from 600,000 twin pairs in Europe. *Eur J Hum Genet.* 2007;15:718-23.
56. Slimani N, Deharveng G, Unwin I, *et al.* The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr.* 2007;61:1037-56.
57. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet.* 2003;361:598-604
58. Hoggart CJ, Parra EJ, Shriver MD, *et al.* Control of confounding of genetic associations in stratified populations. *Am J Hum Genet.* 2003;72:1492-1504
59. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev.* 2002;11:505-512
60. Donahue MP, Kraus WE. Genetic association studies; the good, the bad, and the ugly. *Am Heart J.* 2007;154:610-2.
61. Schulze TG, McMahon FJ. Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Hum Hered.* 2004;58:131-8.
62. Zhao X, Ramsey KE, Stephan DA, *et al.* Gene and protein expression changes in human trabecular meshwork cells treated with transforming growth factor-beta. *Invest Ophthalmol Vis Sci.* 2004;45:4023-34.
63. Bhattacharya SK, Rockwood EJ, Smith SD, *et al.* Proteomics reveal Cochlin deposits associated with glaucomatous trabecular meshwork. *J Biol Chem.* 2005;280:6080-4.
64. Bhattacharya SK, Crabb JS, Bonilha VL, *et al.* Proteomics implicates peptidyl arginine deiminase 2 and optic nerve citrullination in glaucoma pathogenesis. *Invest Ophthalmol Vis Sci.* 2006;47:2508-14.
65. Steely HT, Dillow GW, Bian L, *et al.* Protein expression in a transformed trabecular meshwork cell line: proteome analysis. *Mol Vis.* 2006;12:372-83.

66. Zhang Y, Gao Q, Duan S, *et al.* Upregulation of Copine1 in trabecular meshwork cells of POAG patients: a membrane proteomics approach. *Mol Vis.* 2008;14:1028-36.
67. Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860-921.
68. Venter JC, Adams MD, Myers EW, *et al.* The sequence of the human genome. *Science.* 2001;291:1304-51.
69. The International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426:789-96.
70. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007;447:661-78.
71. Wiggs JL. Genomic promise: personalized medicine for ophthalmology. *Arch Ophthalmol.* 2008;126:422-3.
72. Thorleifsson G, Magnusson KP, Sulem P, *et al.* Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* 2007;317:1397-400.
73. Fingert JH, Alward WL, Kwon YH, *et al.* LOXL1 mutations are associated with exfoliation syndrome in patients from the midwestern United States. *Am J Ophthalmol.* 2007;144:974-5.
74. Hewitt AW, Sharma S, Burdon KP, *et al.* Ancestral LOXL1 variants are associated with pseudoexfoliation in Caucasian Australians but with markedly lower penetrance than in Nordic people. *Hum Mol Genet.* 2008;17:710-6.
75. Pasutto F, Krumbiegel M, Mardin CY, *et al.* Association of LOXL1 common sequence variants in German and Italian patients with pseudoexfoliation syndrome and pseudoexfoliation glaucoma. *Invest Ophthalmol Vis Sci.* 2008;49:1459-63.
76. Ozaki M, Lee KY, Vithana EN, *et al.* Association of LOXL1 Gene Polymorphisms with Pseudoexfoliation in the Japanese. *Invest Ophthalmol Vis Sci.* 2008;49:3976-80.
77. Viswanathan AC, Hitchings RA, Indar A, *et al.* Commingling analysis of intraocular pressure and glaucoma in an older Australian population. *Ann Hum Genet.* 2004;68:489-97.

78. Charlesworth JC, Dyer TD, Stankovich JM, *et al*. Linkage to 10q22 for maximum intraocular pressure and 1p32 for maximum cup-to-disc ratio in an extended primary open-angle glaucoma pedigree. *Invest Ophthalmol Vis Sci*. 2005;46:3723-9.
79. Centre for Eye Research Australia (CERA):  
<http://cera.unimelb.edu.au/eyehealth/glaucoma.html>
80. Gift of Sight Eye Research Centre:  
[http://www.giftofsight.org.uk/about\\_gos/index.html](http://www.giftofsight.org.uk/about_gos/index.html)
81. Wellcome Trust Case Control Consortium 2:  
<http://www.parliament.uk/documents/upload/stGMWellcomeTrustCentreforHumanGenetics.pdf>
82. Brooks BP, Macdonald IM, Tumminia SJ, *et al*; National Ophthalmic Disease Genotyping Network (eyeGENE). Genomics in the era of molecular ophthalmology: reflections on the National Ophthalmic Disease Genotyping Network (eyeGENE). *Arch Ophthalmol*. 2008;126:424-5.