IJE vol.34 no.1 © International Epidemiological Association 2004; all rights reserved.
Advance Access publication 23 November 2004

International Journal of Epidemiology 2005;**34**:21–27
doi:10.1093/ije/dyh327

MORGAM (an international pooling of cardiovascular cohorts)

Alun Evans,[1]* Veikko Salomaa,[2] Sangita Kulathinal,[2] Kjell Asplund,[3] François Cambien,[4] Marco Ferrario,[5] Markus Perola,[6] Leena Peltonen,[6] Denis Shields,[7] Hugh Tunstall-Pedoe,[8] and Kari Kuulasmaa,[2] for the MORGAM Project

Accepted 5 August 2004

Key words GenomEUtwin, MORGAM, case–cohort study, cardiovascular risk, genetic epidemiology, prospective studies

How did the study come about?

Dissatisfaction has been voiced over recent decades concerning the lack of a relevant cardiovascular disease (CVD) scoring system for European populations. Recently this deficiency has been repaired with the publication of SCORE,[1] although nonfatal events are still not catered for. In addition, the entire sequence of the human genome has recently been published.[2] Common chronic diseases, such as coronary heart disease (CHD) and stroke, may have a strong genetic component. They are, however, caused not by a single genetic defect but by the interactions of many genetic and environmental factors. Hence, they are often called complex, multifactorial diseases. Moreover, the biological effects of common genetic variants are likely to be small in magnitude; indeed, variants with large biological effects tend to be rare, for example familial hypercholesterolaemia. Investigators examining the genetic background of complex, multifactorial diseases should, therefore, realize that they are looking for interactions between genetic variants with small, or at most moderate, effects. It is obvious that the reliable detection of these effects requires large sample sizes and abundant statistical power, which can be achieved only in a large collaborative study using high-throughput genotyping. It should be emphasized, however, that moderate and even small effects can carry considerable public health significance if the genetic variants in question are common in the population. The remarkable success of the Human Genome Project has been possible only through the multinational collaboration of several research laboratories and the open exchange of information through the Internet. Developments in genetics open up new possibilities for the prevention and treatment of chronic diseases, but to capitalize on this potential, a better understanding of the significance of genetic variation and the interactions of genetic variants with environmental factors is needed.

Towards the end of the WHO MONICA Project[3] it was realized that a follow-up of the cohorts recruited by the project would be ideal for exploring both issues mentioned above. This follow-up project was established under the name MORGAM (MONICA, Risk, Genetics, Archiving, and Monograph; www.ktl.fi/morgam) and now also includes cohorts from non-MONICA centres. It was initially funded under the Fourth Framework Programme of the European Union. Two of its components, archiving and monograph, have been completed. This profile describes the remaining two, risk and genetics, both of which are based on the pooling of prospective CVD cohorts.

For a subset of these cohorts DNA is available and central collation, preparation and genotyping in a case–cohort setting are well under way. Since 2002, these activities have become a component of GenomEUtwin (www.genomeutwin.org), a Network of Excellence for Genomics in Europe, funded under the Fifth Framework Programme. Centres have recruited their cohorts and organized the follow-up locally using their own funding. MORGAM is pooling these cohorts, and the funding is devoted to co-ordination, pooling of samples and data, quality assessment and control, central preparation of DNA, and laboratory analysis. Support for the participating centres is through access to their own results, the return of surplus prepared DNA, a modest subvention to support data preparation and sample handling, and attendance at an annual workshop. MORGAM has a Coordinator (A.E.) who also chairs the MORGAM Management Group, on which the participating laboratories, the MORGAM Data Centre at the Finnish National Public Health Institute (KTL) in Helsinki, and the GenomEUtwin Coordinator are represented.

[1] Department of Epidemiology and Public Health, Mulhouse Building, Queen's University Belfast, Belfast BT12 6BJ, UK.

[2] KTL—National Public Health Institute, Department of Epidemiology and Health Promotion, Mannerheimintie 166, 00300 Helsinki, Finland.

[3] National Board of Health and Welfare, SE-10630 Stockholm, Sweden.

[4] INSERM U525, Faculté de Médecine Pitié-Salpêtrière, 91 boulevard de l'Hôpital, 75634 Paris Cedex 13, France.

[5] Università degli studi dell'Insubria, Medicina del Lavoro e Preventiva, Viale Borri 57, 21100 Varese, Italy.

[6] KTL—National Public Health Institute, Department of Molecular Medicine, Biomedicum, Haartmaninkatu 8, PO Box 104, 00251 Helsinki, Finland.

[7] Clinical Pharmacology, Royal College of Surgeons in Ireland, 123 St Stephen's Green, Dublin 2, Ireland.

[8] Cardiovascular Epidemiology Unit, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland.

* Correspondence: Department of Epidemiology and Public Health, Mulhouse Building, Queen's University Belfast, Grosvenor Road, Belfast BT12 6BJ, UK. E-mail: a.evans@qub.ac.uk

What does MORGAM cover, and how has this changed?

MORGAM was designed with the overall aim of eventually studying a limited number of well-defined phenotypes and several hundred genetic factors simultaneously by pooling the risk-factor data collected in the MONICA cross-sectional surveys, adding follow-up and genotyping, and including other relevant cohorts.

The main objective of the risk-factor component of the study is to assess the similarity of risk coefficients for the classic CVD risk factors in different parts of Europe, between men and women, and between age groups using large cohorts with standardized baseline measurements and carefully validated fatal and nonfatal CHD and stroke end-points. Furthermore, the data will be used to derive European risk scores and to assess the impact of socioeconomic and other factors for which data are available in the cohorts.

The MORGAM genetic component, which employs a case–cohort design, will provide a framework for the analysis of the associations among genetic variants, risk-factor phenotypes, and disease events. This could also allow an assessment of gene–gene and gene–environment interactions. The identification of meaningful combinations of genotypes and environmental factors will rely heavily on the use of appropriate mathematical models and statistical techniques. The main objective of the genetic component is to determine statistically significant combinations of single nucleotide polymorphisms (SNPs) from the multitude of genotypic data which, in combination with environmental factors and possible intermediate quantitative phenotypes, are predictors of incident CHD and stroke events and total mortality. We are particularly interested in examining the interactions of polymorphisms which are located along the same biological pathway. The pathways of interest will be genotyped systematically so that each known SNP with a frequency >1% will be considered. Ultimately, on average ~6 SNPs of each gene in the pathway will be chosen for genotyping on the basis of comparative genomics, linkage disequilibrium relationships, and the literature.

Additional objectives of the study for the participating populations are to (i) estimate the genotype and allele frequencies; (ii) assess the linkage disequilibrium relationships and haplotype frequencies; (iii) determine the population-attributable risk associated with the most common forms of adverse genotypes; and (iv) examine the relationships between different genotypes and risk-factor phenotypes. The main hypothesis of the genetic component is that the variation in genes—for example, regulating blood coagulation and/or inflammatory reaction and/or lipid metabolism—is a determinant of CVD risk.

Statistical methods

Analysis of the follow-up data to assess the effect of classic risk-factor phenotypes is being carried out using Cox's proportional hazards model. The adoption of a case–cohort design means that multiple end-points can be studied and the prevalence of gene polymorphism in different parts of Europe can be established.[4] Statistical methods for the analysis of case–cohort data with multiple end-points are being developed in MORGAM.

Moreover, new statistical approaches for analysing the resultant wealth of genetic markers will also be developed. Preliminary analysis of genotypic data will be carried out as follows: estimation of the allele frequencies, estimation of haplotypes, and application of data-mining methods to look for patterns which are associated with the incidence of CVD. The haplotype frequencies can be estimated using any of the existing methods.[5] Owing to the incompleteness of the data on haplotypes, simultaneous estimation of haplotype frequencies and the effect of haplotypes using an appropriate survival model is needed, and tools for such analysis are being developed as part of MORGAM.

The relationships between different genotypes, risk-factor phenotypes, and disease end-points are being investigated and will be analysed in the case–cohort setting. We also propose to perform a systematic, structured analysis of the data using, for example, the Bayesian approach to statistical inference, which has recently been popularized in genetics. Lately, some of the MORGAM team, as part of the ECTIM study, have developed a method taking into account not only the raw association between a polymorphism and a phenotype but also the effects of all polymorphisms which are in linkage disequilibrium with it. They have elegantly demonstrated this for nine polymorphisms of the P-selectin gene by employing a new maximum likelihood method.[6] Detailed haplotype analysis confirmed the protective effect of the P715 allele and revealed that two asparagine codons were consistently associated with a higher risk of myocardial infarction, but only when they shared the same haplotype. Another statistical tool developed by the team which appears promising in this respect is DICE (detection of informative combined effects).[7] This employs automated data-mining to explore the effects of several polymorphisms or other nongenetic covariates. It combines the advantages of regressive approaches with data exploration tools, assesses interactions between polymorphisms, and is efficient at evaluating the spectrum of polymorphisms within a given biological system. This is relevant because MORGAM is mainly concentrating its efforts on biological systems.

Population stratification

Population stratification is theoretically an important problem as it implies confounding due to population structure. This could explain some of the inconsistencies observed in association studies. The problem arises because an 'admixture' of groups of people with differing genotypes within a population could lead to misleading results if they are unevenly distributed between case and control groups.[8] Obviously this is of importance to MORGAM in view of the large range of populations which are being included, and it has been suggested that unless random samples are selected from one homogeneous population, this effect is always a legitimate cause for concern over positive findings in association studies, except those which deliberately control for it. There are essentially two population-based approaches for controlling for the effect: Genomic control[9] and structure assessment.[10] The problem is being actively addressed in MORGAM.

Ethical issues

MORGAM has developed a system for dealing with the complexities of undertaking a multicentre study based on the genotyping of cohorts in several European countries. Added to these complexities was the fact that, when the project was

originally funded, the European Commission had no specific guidelines on the conduct of such studies. MORGAM has developed a system wherein ethical approval must be obtained from the local ethics committee and evidence of the participants' informed consent is required from each participating centre. The only exception to this is when a national ethics committee grants consent for fully anonymized samples to be analysed outside the country in question. All samples and data are processed anonymously. Because each centre can have full access to its genotypic data on request, it is left up to those centres to consult with their local ethics committees before the data are fed back to participants, with the caveat that such data have been generated for research purposes only, and as such are liable to error. Therefore, all results must be repeated on a fresh sample of the participant's DNA before any advice is given. In addition, a series of material and data transfer agreements have been devised to cover such exchanges (www.ktl.fi/morgam).

Further phenotypic study

MORGAM will study a large number of polymorphisms but only a few well-standardized phenotypes. There are plans in hand to measure a large number of phenotypes, which will be employed to address the inflammatory hypothesis of CHD.

Who is included in the sample?

Cohorts of adequate size and the quality of the measurements are critical both for the risk-factor component of the study and for addressing the genetic hypotheses involving multiple interactions. We are conducting a large multicentre study and we are giving a high priority to the quality of the data and to common polymorphisms in the participating populations. The data come from the centres participating in MORGAM, which are able to follow up their cohorts for CHD events, stroke, and total mortality and most of which are also able to provide a DNA sample. The cohorts are from Australia, Denmark, Finland, France, Italy, Lithuania, Northern Ireland, Poland, Russia, Scotland, Sweden, and Wales (Figure 1). Most of these are representative samples of populations from geographically defined areas (Table 1). There are several other potential centres. To date, MORGAM has identified a total of 12 564 deaths from all causes, and 8916 CHD and stroke (fatal and nonfatal) cases in a total cohort of 128 874 subjects. In 52 446 of these, for whom DNA is available, 2631 cases have been validated. For cohorts in the genetic component, DNA from all deaths and CVD cases and from a random sample of the full cohort will be extracted and stored at the Department of Molecular Medicine at KTL in Helsinki. The design of MORGAM allows any positive findings to be confirmed in multiple, independent populations.

How often have study participants been followed up?

All the cohorts were examined once at baseline. The length of the follow-up period varies between centres (Table 1), and many centres will extend their follow-up of end-points in future. The follow-up procedures vary between centres and are summarized in Table 2.

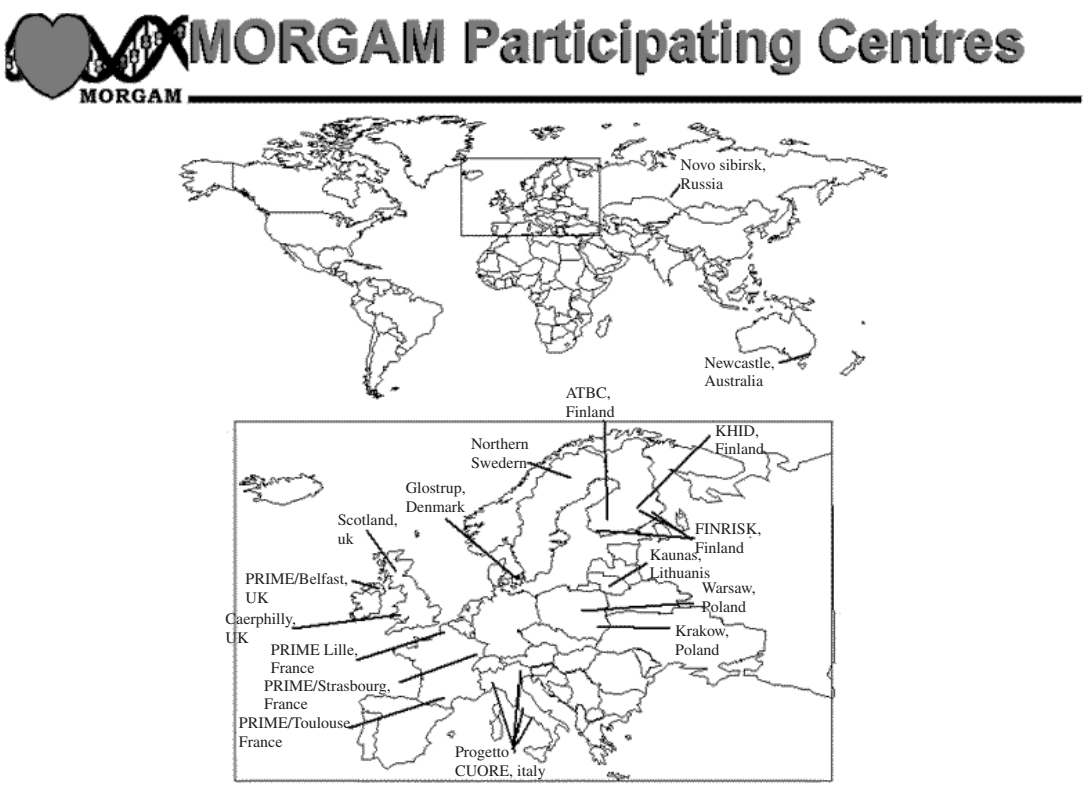

Figure 1 MORGAM participating centres

Table 1 Summary of MORGAM cohorts (as of June 30, 2004)

Country	Population	Cohorts[a]	Cohorts with DNA	Baseline survey period	Age group at baseline	Size of cohorts	Size of cohorts with DNA	End of follow-up
Australia	Newcastle	3	0	1983–94	24–70	5 873	0	1998
Denmark	Glostrup	4	2	1982–94	25–74	6 914	4 281	1998
Finland	ATBC[b]	1	1	1992–93	54–77	5 073	5073	1999
	FINRISK/East	3	1	1982–92	24–65	13 192	3 013	2001
	KIHD[b]	1	1	1991–93	46–65	1 038	1 038	2000
	FINRISK/West	4	2	1982–92	24–65	7 647	2 986	2001
France	PRIME/Lille[b]	1	1	1991–93	49–64	2 633	2 633	
	PRIME/Strasbourg[b]	1	1	1991–93	49–64	2 612	2 612	5 years
	PRIME/Toulouse[b]	1	1	1991–93	49–64	2 610	2 610	
Italy	CUORE/Brianza	4	4	1986–94	25–75	6 976	6 976	2002
	CUORE/Friuli	4	2	1986–96	25–64	5 910	2 186	1998
	CUORE/Rome	5	2	1982–96	18–81	1 0235	4 489	1998
	CUORE/Naples[c]	1	1	1993–97	29–72	5 062	5 062	1998
Lithuania	Kaunas	3	0	1983–93	33–65	4 485	0	1998
Poland	Krakow	3	0	1983–93	34–65	5 362	0	1998
	Warsaw	3	0	1983–93	35–64	5 577	0	1998
Russia	Novosibirsk	4	1	1983–95	23–65	11 438	3 273	1998
Sweden	Northern Sweden	3	2	1986–94	24–74	5 094	3 469	1999
UK	PRIME/Belfast[b]	1	1	1991–94	49–60	2745	2 745	5 years
	Caerphilly[b]	1	0	1984–88	47–67	2 398	0	2000
	Edinburgh	1	0	1986	25–65	1 299	0	2000
	Glasgow	4	0	1986–95	25–75	5 536	0	2000
	Scotland	1	0	1984–87	38–61	9165	0	2000
Total		57	23			128 874	52 446	

[a] Number of cohorts examined at different times.

[b] Men only.

[c] Women only.

Table 2 Broad categories of data items collected in MORGAM[a]

Phase	Data	Measurements
Baseline	Questionnaire	Smoking, alcohol use, socioeconomic indicators, history of CHD, stroke, diabetes, family history of myocardial infarction and stroke, questions specific to women, ethnicity, availability of DNA, Rose questionnaire, Minnesota code
	Measured	Anthropometric measurements, blood pressure, cholesterol, triglycerides, fibrinogen, and SNP genotypes[b]
Follow-up	Mortality	Mostly by linkage to the national death register covering the whole country, in some centres by periodic follow-up by letter or linkage to health service register
	Coronary and stroke events	MONICA coronary and stroke event register, hospital discharge register, clinical event questionnaire, regional health information system

[a] Details of the data collected are available at www.ktl.fi/morgam.

[b] SNP genotypes are determined using a case–cohort design, for a random sample of the study cohort and for cases who experienced coronary heart disease, stroke, death, with a history of cardiovascular disease at baseline.

What has been measured?

Table 2 categorizes the data items collected or measured in MORGAM. The details of the measurements, other than genotypes, are described in the MORGAM Manual (www.ktl.fi/morgam).

The DNA will be genotyped using up-to-date high-throughput methods at KTL (mass spectrometry and DNA array-based chips), at INSERM U525, Paris, and at the Royal College of Surgeons, Dublin (mass spectrometry). The list of genes considered for genotyping in the first phase is given in Table 3. In Dublin, genes affecting platelets and thrombosis, which have not been tested in other studies, and genes for antioxidants will be studied. The Paris laboratory will initially concentrate on the integrin system, employing advanced laboratory techniques. The use of high-throughput centralized laboratories will facilitate cost-effective analysis which will be very sparing of DNA resources. In the near future several hundred polymorphisms of candidate genes will be typed in MORGAM. Thus, the genotyping will permit very powerful and very challenging analyses. Where the amount of DNA is very limited, whole genome amplification will be performed. In the future, whole genome scans may be practicable in MORGAM. So far a total of 109 000 genotypes have been processed.

Over the life of MORGAM the DNA requirements have become increasingly frugal, so that at present a mere 10 μg of

Table 3 Primary analysis set of genes to be genotyped (as of June 30, 2004)

Gene	Locuslink name	Chromosome
Candidate genes related to lipids, inflammation, and coagulation		
ATP-binding cassette, subfamily A (ABC1), member 1	ABCA1	9q31.1
Beta2-adrenoceptor	ADRB2	5q31–q32
Beta3-adrenoceptor	ADRB3	8p12–p11.2
Angiotensin II receptor type 1	AGTR1	3q24
Apolipoprotein A-1	APOA1	11q23–24
Apolipoprotein A-2	APOA2	1q21–23
Apolipoprotein A-4	APOA4	11q23
Apolipoprotein A-5	APOA5	11q23
Apolipoprotein E	APOE	19q13.2
Monocyte differentiation antigen CD14	CD14	5q31.1
Cholesteryl ester transfer protein	CETP	16q21
Thrombin-activable fibrinolysis inhibitor	CPB2	13q14.11
Cholesterol 7alpha-hydroxylase gene	CYP7A1	8q11–q12
Coagulation factor X	F10	13q34
Coagulation factor XII	F12	5q33–qter
Coagulation factor XIII, alpha subunit	F13A	6p25.3–p24.3
Prothrombin (coagulation factor 2)	F2	11p11–q12
Thrombin receptor (F2R)	F2R	5q13
Coagulation factor V	F5	1q23
Coagulation factor VII	F7	13q34
Fibrinogen, alpha chain	FGA	4q28
Fibrinogen, beta chain	FGB	4q28
Fibrinogen, gamma chain	FGG	4q28
Forkhead box C2 (MFH-1, mesenchyme forkhead 1)	FOXC2	16q22–q24
Fucosyltransferase 1,3	FUT3	19p13.3
Familial combined hyperlipidemia 1	HYPLIP1	1q21–q23
Intercellular adhesion molecule 1	ICAM1	19p13.3–p13.2
Interleukin 10	IL10	1q31–q32
Interleukin 1 alpha	IL1A	2q14
Interleukin 1 beta	IL1B	2q14
Interleukin 6	IL6	7p21
Integrin alpha 2	ITGA2	5q23–q21
Integrin beta 3	ITGB3	17q21.32
Klotho	KL	13q12
Lecithin-cholesterol acyltransferase	LCAT	16q22.1
Lactase	LCT	2q21
Low-density lipoprotein receptor	LDLR	19p13.3
Hepatic lipase	LIPC	15q21–q23
Mannose binding lectin	LMAN1	18q21.3–q22
Melanocortin 4 receptor	MC4R	18q22
5,10-Methylenetetrahydrofolate reductase	MTHFR	1p36.3
Microsomal triglyceride transfer protein	MTP	4q24
Preproneuropeptide Y	NPY	7p15.1
Phosphodiesterase 4D	PDE4D	5q11.2
Tissue plasminogen activator	PLAT	8p12
Phospholipid transfer protein	PLTP	20q12–q13.1
Peroxisome proliferative activated receptor, alpha	PPARA	22q13.31
Peroxisome proliferator activator receptor, delta	PPARD	6p21.2–p21.1
Peroxisome proliferative activated receptor, gamma	PPARG	3p25
Protein C	PROC	2q13–q14
RXR beta	RXRB	6p21.3
Selectin E	SELE	1q22–q25
Selectin L	SELL	1q23–q25
Selectin P	SELP	1q22–q25
Plasminogen activator inhibitor 1	SERPINE1	7q21.3–q22
Tissue factor pathway inhibitor	TFPI	2q31–q32.1
Thrombomodulin	THBD	20p12–c
Tumor necrosis factor alpha	TNFA	6p21.3
Upstream transcription factor 1	USF1	1q22–23
Vascular cell adhesion molecule 1	VCAM1	1p32–p31
Pathogen response modulating genes		
Chemokine (C-C) receptor 5	CCR5	3p21
Fucosyltransferase 2	FUT2	19q13.3
Natural resistance-associated macrophage protein 1	SLC11A1	2q35
Vitamin D (1,25 dihydroxyvitamin D3) receptor	VDR	12q12–12q14
Fc gamma, low affinity IIa receptor	FCGR2A	1q21–q23
Interleukin 13	IL13	5q31
Toll-like receptor 4	TLR4	9q32–q33
IL12 receptor beta1	IL12RB1	19p13.1
Chemokine SDF1 (CXCL12)	CXCL12	10q11.1
Chemokine receptor CCR2	CCR2	3p21
Candidate genes related to adhesion and inflammation systems		
Interleukin-18	IL18	11q22.2–q22.3
Interleukin-18 binding protein	IL18BP	
Interleukin-18 receptor 1	IL18R1	2q12
Interleukin 1-receptor antagonist	IL1RN	2q14.2
Tumor necrosis factor receptor superfamily, member 1A	TNFRSF1A	12p13.2
Tumor necrosis factor receptor superfamily, member 1B	TNFRSF1B	1p36.3–36.2
Tumor necrosis factor-alpha-induced protein 6	TNFAIP6	2
Tumor necrosis factor ligand superfamily, member 5 (CD40 ligand)	TNFSF5	Xq26
Glutathione peroxidase 1	GPX1	3p21.3

Table 3 continued

Gene	Locuslink name	Chromosome
Interleukin 18 receptor accessory protein	IL18RAP	2q12
Caspase 1	CASP1	11q22.2–q22.3
Interleukin-1 receptor, type 1	IL1R1	2q12
Interleukin-1 receptor, type 2	IL1R2	2q12-q22
Interleukin-1 receptor accessory protein	IL1RAP	3q28
Interleukin 12A	IL12A	3p12–q13.2
Interleukin 12B	IL12B	5q31.1–q33.1
Interleukin 15	IL15	4q31
Lymphotoxin-alpha	LTA	6p21.3
Selectin P ligand	SELPLG	12q24
Integrin alpha-4	ITGA4	2q31–q32
Integrin alpha-V	ITGAV	2q31
Integrin alpha-L	ITGAL	16p11.2
Integrin alpha-X	ITGAX	16p11.2
Integrin beta-1	ITGB1	10p11.2
Integrin beta-2	ITGB2	21q22.3
Intracellular adhesion molecule 2	ICAM2	17q23–q25
Intracellular adhesion molecule 3	ICAM3	19p13.3–p13.2
Platelet-endothelial cell adhesion molecule 1	PECAM1	17q23
Cadherin-15	CDH15	16q24.30

DNA is required. This reflects the rapid advances which have been made in genotyping—requiring smaller and smaller amounts of DNA; it is planned to keep a small aliquot of DNA for analysis on a future high-throughput platform.

What is the attrition rate?

Particular attention has been paid to the coverage of the follow-ups. Most of the centres have used national death registers covering the whole country for the follow-up of mortality. However, the geographic coverage of the follow-up of nonfatal events varies between centres from the whole country to the study area. In most centres, information about loss-to-follow-up is available.

What has MORGAM found?

The study is at an advanced stage. The baseline and follow-up data have been centrally collected and genotyping is underway. The first publications on the effects of the classic risk factors and genetics on CHD and stroke are expected shortly.

What are the main strengths and weaknesses?

The strengths of the risk-factor cohort are manifestly its size, standardized baseline and end-point assessment, the inclusion of non-fatal cases, and the ability to compare different geographic areas.

There is currently considerable interest in large population-based genetic studies of complex diseases and the contribution of environmental factors to their manifestation. MORGAM now forms part of GenomEUtwin, and the rationale for this is that various putative genetic traits identified in the twin research approach which are relevant to CVD can be tested in the MORGAM cohorts.[11] Classic data from the Swedish Twin Registry testify to the importance of genetics in CHD: the relative hazard of one twin succumbing to the disease if the other has died from the condition before the age of 55 years is very significant, and this persists into the eighth decade.[12] There may also be scope for investigating the level of risk factors: in GenomEUtwin the heritability of both systolic and diastolic blood pressure in monozygotic twins has been found to be ~50%. Similarly, related twin studies have found that the heritability of lipoprotein(a) is substantial, and, in smokers, genetic factors determine 86% of the amount smoked.[13]

It is a central tenet in any discussion of the contribution of polymorphism to complex disease that the cumulative effects of variants carrying a slight excess of risk, particularly when they interact with environmental factors such as diet, contribute more than rare, serious mutations at the population level. The dominant condition of familial hypercholesterolaemia was, until recently, pretty disastrous for the individual, but its population-attributable risk was small.[13] The complexities of unravelling the genetic contribution of many polymorphisms to the development of CVD may be huge in view of the several hundreds which may be involved. The problem is that many of the single polymorphisms found to a be related to a disease in association do not pass the test of time, that is to say, further study. There has been a spate of papers bringing researchers face to face with this stark reality.[13] However, there is hope: in an excellent paper, Cardon and Bell state, 'Despite their recognized limitations, association studies represent an essential step in advancing the field to [*sic*] the definition of disease-mediating genetic variants.'[14] They go on to observe that 'Control ascertainment can be improved by using a PROSPECTIVE COHORT study. This requires a substantial collection of individuals to be selected before the onset of disease and to be followed under the same experimental protocol.' And they conclude that 'When properly applied and interpreted, it is likely that association will continue to provide an essential component of the expanding arsenal needed to dissect and characterize the genetic basis of common disease.'

There are a few other large population genomics research projects currently running or at the planning stage: three of these have come together with GenomEUtwin to form an international consortium: Public Population Project in Genomics (www.p3gonsortium.org/). The other partners are CART@GENE, a Canadian venture which will initially recruit 1700 participants beginning in Autumn 2004; the Estonian Genome Project, which currently has 9000 participants; and UK Biobank, which aims to recruit 500 000 volunteers but is still at a planning stage.[15]

The aim is to develop MORGAM further as an open, collaborative research network. The risk scores for CHD and stroke are at an advanced stage of analysis and genotyping is well under way. As mentioned ('How did the study come about?'), participating centres receive a modest degree of financial support, but the real benefit they enjoy is that MORGAM genotypes their DNA samples for them. By its nature, MORGAM can only standardize a limited number of phenotypes with precision across the many cohorts; once a centre receives its genotyping results it is free to relate them to whatever phenotypes it may have measured locally. Moreover, MORGAM

does not have a fixed closing date and will remain open to new cohorts for which DNA is available and to considering innovative research proposals, provided that appropriate ethical clearances are in position. Frozen sera or plasma samples from subsets of MORGAM cohorts may be used for additional phenotyping, in particular to explore the inflammatory hypothesis for CHD.

It is possible to pool cohorts internationally, provided that adequate quality assurance procedures are in place. These have been meticulously developed within the MONICA Project over the past two decades. It is only through drawing on the unique resource developed by the MONICA Project that its individual participating centres could adequately pool their data and samples, which are potentially so important.

Can I get hold of the data? Where can I find out more?

The data at present remain the property of the participating centres, in conjunction with MORGAM. The primary way of gaining access to the data is through collaboration with the MORGAM Project. As mentioned ('Ethical issues'), transfer of data and samples is subject to a system of material and data transfer agreements. Readers who wish to find out more should visit the MORGAM website, where a list of MORGAM publications is being assembled (www.ktl.fi/morgam).

Acknowledgements

Some of these data originate from the GenomEUtwin Project, which is supported by the European Union (Contract No. QLG2 CT-2002-01254).

References

[1] Conroy RM, Pyorala K, Fitzgerald AP *et al*. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;**24:**987–1003.

[2] McPherson JD, Marra M, Hillier L *et al*. A physical map of the human genome. *Nature* 2001;**409:**934–41.

[3] Tunstall-Pedoe H (ed). *MONICA: Monograph and Multimedia Sourcebook. Geneva: WHO*, 2003.

[4] Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *J Clin Epidemiol* 1999;**52:**1165–72.

[5] Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;**12:**921–27.

[6] Tregouet D-A, Barbaux S Escolano S *et al*. Specific haplotypes of the P-selectin gene are associated with myocardial infarction. *Hum Mol Genet* 2002;**11:**2015–23.

[7] Tarhi-Daizadeh N, Tregouet D-A, Nicaud V, Manuel N, Cambien F, Tiret L. Automated detection of informative combined effects in genetic association studies of complex traits. *Genome Res* 2003;**13:**1952–60.

[8] Deng H-W, Chen W-M, Recker RB. Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. *Genetics* 2001;**157:**885–97.

[9] Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;**55:**997–1004.

[10] Pritchard JK, Stephens M, Rosenberg NA, *et al*. Association mapping in structured populations. *Am J Hum Genet* 2000;**67:**170–81.

[11] Peltonen L. GenomEUtwin: a strategy to identify genetic influences on health and disease. *Twin Res* 2003;**6:**354–58.

[12] Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death in a study of twins. *N Engl J Med* 1994;**330:**1041–46.

[13] Evans A, van Baal GC, McCarron P *et al*. The genetics of coronary heart disease: the contribution of twin studies. *Twin Res* 2003;**6:**432–41.

[14] Cardon LR, Bell JI. Association study designs for complex disease. *Nat Rev Gent* 2001;**2:**91–99.

[15] The Wellcome Trust, Medical Research Council, Department of Health. Protocol for Biobank UK: a study of genes, environment and health. London, February 2002.